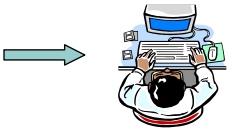
XML y Bases de Datos

Ana Belén Martínez Prieto
Universidad de Oviedo

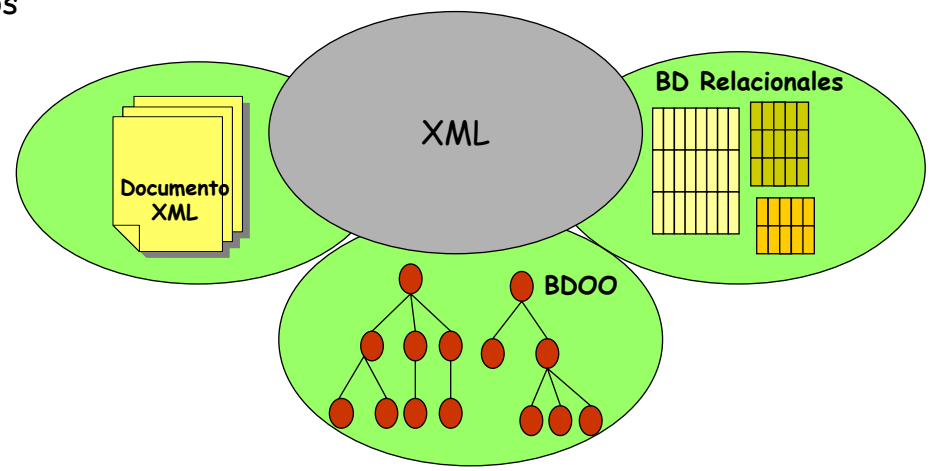
Contenidos

- Introducción
- XML y Bases de Datos Relacionales
- XML y Bases de Datos Orientadas a Objetos
- XML y Bases de Datos Nativas
- XQuery



XML – Introducción

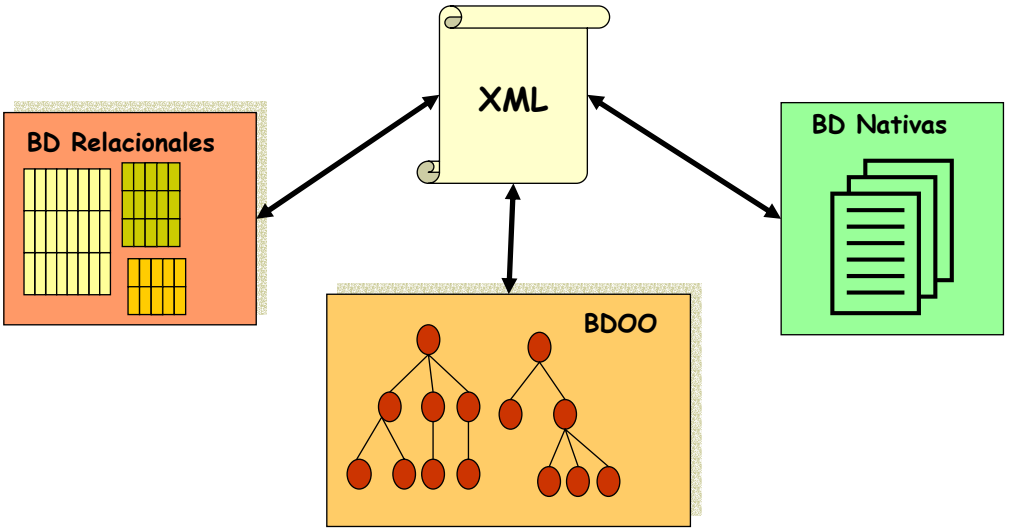
- Estándar para el intercambio de datos entre aplicaciones en Internet independiente del formato de almacenamiento de los mismos



- Es lógico que las consultas entre aplicaciones se expresen como consultas contra los datos en formato XML → **XQuery**

XML – Introducción (II)

- Muchas aplicaciones requieren el almacenamiento de datos XML
- Existen diferentes alternativas:

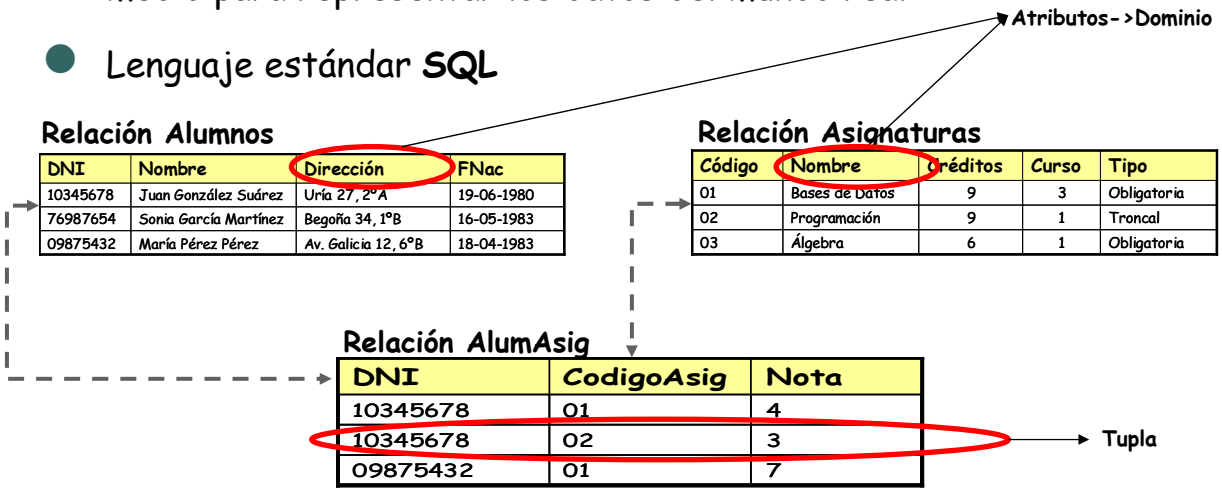


Contenidos

- Introducción
- **XML y Bases de Datos Relacionales**
- XML y Bases de Datos Orientadas a Objetos
- XML y Bases de Datos Nativas
- XQuery

Bases de Datos Relacionales

- Se basan en las relaciones (tablas bidimensionales) como único medio para representar los datos del mundo real
- Lenguaje estándar **SQL**



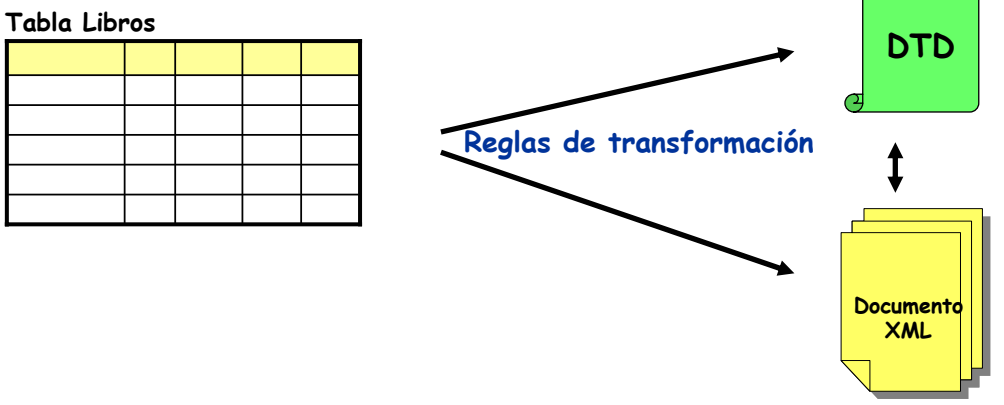
Bases de Datos Relacionales (II)

- Se han creado complejas teorías y patrones para encajar objetos o estructuras jerarquizadas en bases de datos relacionales
  - Existen **numerosos middlewares** encargados de la transferencia de información entre estructuras XML y bases de datos relacionales

Nombre	Fabricante	Licencia	XML->BD	BD->XML
ADO	Microsoft	Comercial	X	X
Allora	HiT Software	Comercial	X	X
ASP2XML	Stonebroom	Comercial	X	X
Data Junction	Data Junction Inc.	Open Source	X	X
DBX	Swift Inc.	Comercial	X	X
InterAccess	XML Soft. Corporation	Comercial	X	X
JaxMe	Jochen Wiedmann	Open Source	X	X
XML-DBMS	Ronald Bourret	Open Source	X	X

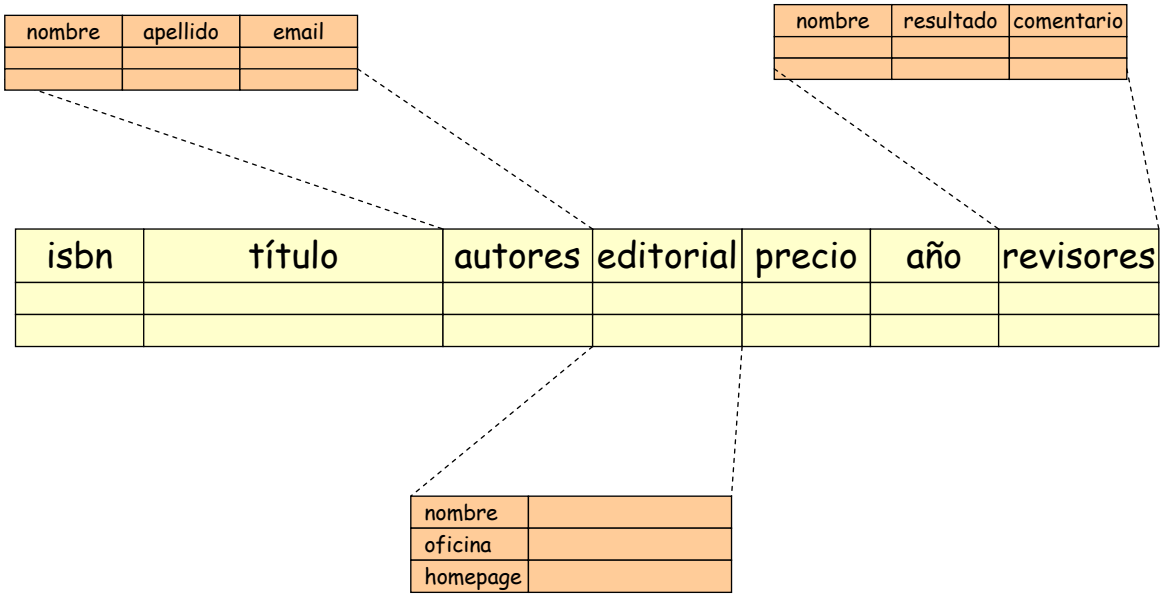
Bases de Datos Relacionales
Transformación a XML

- Veamos un ejemplo de transformación de una tabla a un documento XML



Bases de Datos Relacionales

Transformación a XML - Tabla Libros



Bases de Datos Relacionales

Reglas de Transformación Relacional - DTD

- Regla 1
  - **Para cada tabla** en el esquema de la base de datos hay que **crear un elemento** con el mismo **nombre de la tabla** y la cardinalidad apropiada.

```
<!DOCTYPE libros [
        <!ELEMENT libros (libro)*>

]>
```

Bases de Datos Relacionales

Reglas de Transformación Relacional - DTD

- Regla 2
  - Las columnas de la tabla son incluidas en otro elemento (sub-elemento del elemento creado en la regla anterior), que representa un **registro en la tabla**

```
<!ELEMENT libro (isbn, titulo, autores, editor, precio, año, revisores)>
```

Bases de Datos Relacionales

Reglas de Transformación Relacional - DTD

- Regla 3
  - Para cada columna en la tabla cuyo tipo de dato es simple (char, integer, etc.) crear un elemento, sub-elemento del elemento creado en el paso anterior, de tipo #PCDATA con el mismo nombre de la columna.

```
<!ELEMENT isbn (#PCDATA)>
<!ELEMENT titulo (#PCDATA)>
```

Bases de Datos Relacionales

Reglas de Transformación Relacional - DTD

- Regla 4
  - Para cada columna en la tabla cuyo tipo de dato es complejo (tipo objeto), crear un elemento complejo, sub-elemento del elemento creado en el paso 2, con el mismo nombre de la columna. Para cada propiedad del tipo objeto crear un elemento con el mismo nombre de la propiedad.

```
<!ELEMENT autores (autor)+>
        <!ELEMENT autor (nombre, apellidos, email)>
        <!ELEMENT nombre  (#PCDATA)>
        <!ELEMENT apellidos (#PCDATA)>
        <!ELEMENT email (#PCDATA)>
```

Bases de Datos Relacionales

Reglas de Transformación Relacional - DTD

- Regla 5
  - Para cada columna en la tabla que es una tabla anidada, crear un elemento con el mismo nombre de esa columna y la cardinalidad apropiada. Repetir todos los pasos desde el 2.

Bases de Datos Relacionales

DTD Resultante

```
<!DOCTYPE libros [
    <!ELEMENT libros (libro)*>
    <!ELEMENT libro (isbn, titulo, autores, editor, precio, año, revisores)>
    <!ELEMENT isbn (#PCDATA)>
    <!ELEMENT titulo (#PCDATA)>
    <!ELEMENT autores (autor)+>
    <!ELEMENT autor (nombre, apellidos, email)>
    <!ELEMENT editor (nombre,oficina,homepage) >
    <!ELEMENT nombre  (#PCDATA)>
    <!ELEMENT apellidos (#PCDATA)>
    <!ELEMENT email (#PCDATA)>
    <!ELEMENT oficina (#PCDATA)>
    <!ELEMENT homepage (#PCDATA)>
    <!ELEMENT precio (#PCDATA)>
    <!ELEMENT año (#PCDATA)>
    <!ELEMENT revisores (revisor)*>
    <!ELEMENT revisor (nombre,resultado,comentarios)>
    <!ELEMENT nombre (#PCDATA)>
    <!ELEMENT resultado (#PCDATA)>
    <!ELEMENT comentario (#PCDATA)>]>
```

Bases de Datos Relacionales

Posible Documento XML Resultante

```
<?xml version="1.0">
<libros>
  <libro>
      <isbn> 1-55655-767-6 </isbn>
      <titulo> Fundamentos de Bases de Datos </titulo>
      <autores>
          <autor>
              <nombre> Abraham</nombre>
              <apellido> Silberschatz</apellido>
              <email> silbers@hotmail.com </email>
          </autor>
          <autor>
              <nombre> Henry </nombre>
              <apellido> Korth </apellido>
              <email> korth@ hotmail.com </email>
          </autor>
      </autores>
      <editor>
          <nombre> McGraw-Hill </nombre>
          <oficina> Av. Santander s/n </oficina>
          <homepage> http://www.mcgrawhill.es </homepage>
      </editor>
      <precio> 40 </precio>
      <año> 2003 </año>
      <revisores>
          <revisor>
              <nombre> James  Smith </nombre>
              <resultado> 8 </resultado>
              <comentario> Es un libro de texto básico ...</comentario>
          </revisor>
      </revisores>
  </libro>
</libros>
```

Resumen Transformación
Esquema Relacional - DTD

```
<!DOCTYPE tabla [
    <!ELEMENT tabla (registro)*>
    <!ELEMENT registro (columna1, columna2,..., columnaN)>
    <!ELEMENT columna1 (#PCDATA)>
    <!ELEMENT columna2 (propiedad1, propiedad2,..., propiedadN) >
    . . .
    <!ELEMENT columnaN  (#PCDATA)>
    <!ELEMENT propiedad1 (#PCDATA)>
    <!ELEMENT propiedad2 (#PCDATA)>
   . . .
    <!ELEMENT propiedadN (#PCDATA)> ]>
```

Resumen - Transformación

Esquema Relacional - XML SCHEMA

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">
<xsd:element name="tabla">
      <xsd:complexType>
          <xsd:sequence>
              <xsd:element name="registro" minOccurs="0" maxOccurs="unbounded">
                <xsd:complexType>
                            <xsd:sequence >
                                    <xsd:element name="columna1"  type="xsd:positiveInteger"/>
                                    <xsd:element name="columna2"/>
                                        <xsd:complexType>
                                          <xsd:sequence >
                                            <xsd:element name="propiedad1" type="xsd:string"/>
                                            <xsd:element name="propiedad2" type="xsd:boolean"/>
                                             ....
                                            <xsd:element name="propiedadn" type="xsd:positiveInteger"/>
                                          </xsd:sequence>
                                        </xsd:complexType>
                                  </xsd:element >
                                  <xsd:element name="columnaN"  type="xsd:string"/>
                            </xsd:sequence >
                </xsd:complexType>
              </xsd:element >
            </xsd:sequence >
          </xsd:complexType>
        </xsd:element >
      </xsd:schema>
```

Bases de Datos Relacionales -

- Suponen una posibilidad para el almacenamiento de datos XML
- Sin embargo, no están bien preparadas para almacenar estructuras de tipo jerárquico como son los documentos XML:
  - BD relacionales tienen una estructura regular frente al carácter heterogéneo de los documentos XML
  - Documentos XML suelen contener muchos niveles de anidamiento mientras que los datos relacionales son "planos"
  - Documentos XML tienen un orden intrínseco mientras que los datos relacionales son no ordenados
  - Datos relacionales son generalmente "densos" (cada columna tiene un valor) mientras que los datos XML son "dispersos" pueden representar la carencia de información mediante la ausencia del elemento

Contenidos

- Introducción
- XML y Bases de Datos Relacionales
- **XML y Bases de Datos Orientadas a Objetos**
- XML y Bases de Datos Nativas
- XQuery

Bases de Datos Orientadas a Objetos

Introducción

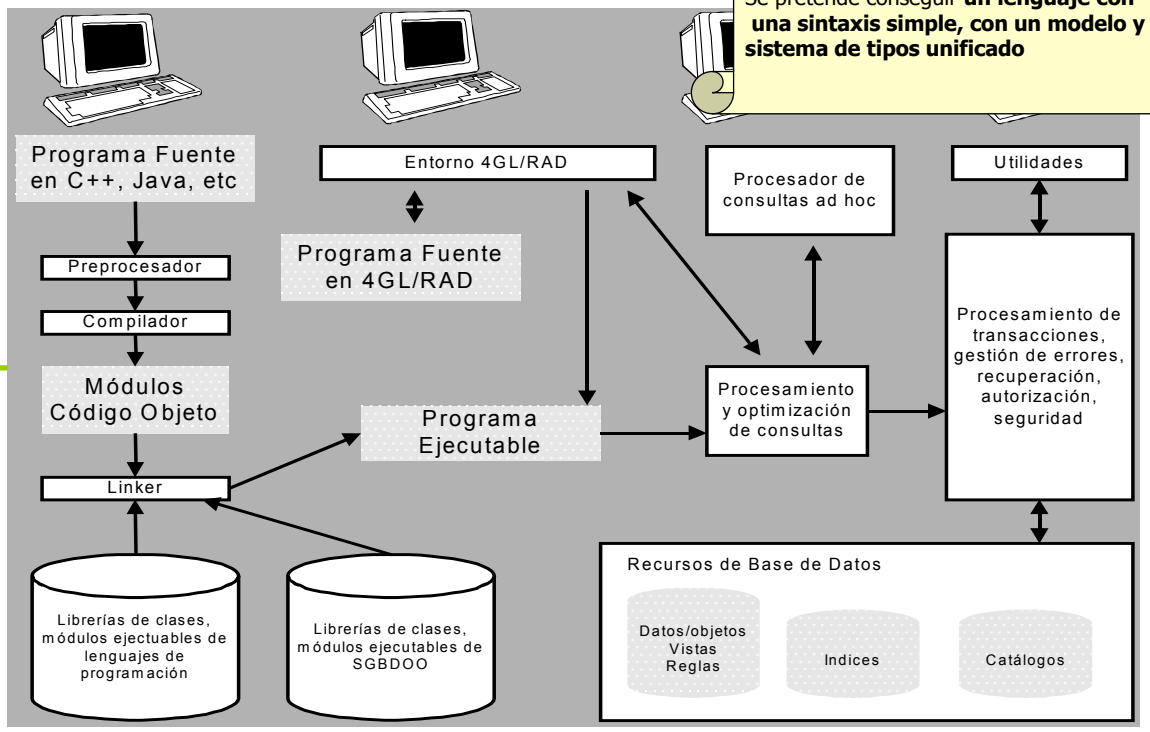
- Los Sistemas de Bases de Datos Orientadas a Objetos soportan un modelo de objetos puro, en la medida de que no están basados en extensiones de otros modelos más clásicos como el relacional:
  - Están fuertemente influenciados por los lenguajes de programación orientados a objetos
  - Pueden verse como un intento de añadir la funcionalidad de un SGBD a un lenguaje de programación
- Ejemplos de SGBDOO son
  - Poet (http://www.poet.com)
  - Jasmine (http://www.cai.com)
  - ObjectStore (http://www.odi.com)
  - GemStone (http://www.gemstone.com)
- Representan **una alternativa viable para el almacenamiento y gestión de documentos XML**

Bases de Datos Orientad

Estructura de Referencia

Extienden el lenguaje de programación y su implementación (compilador, preprocesador, ...) para incorporar funcionalidad de base de datos

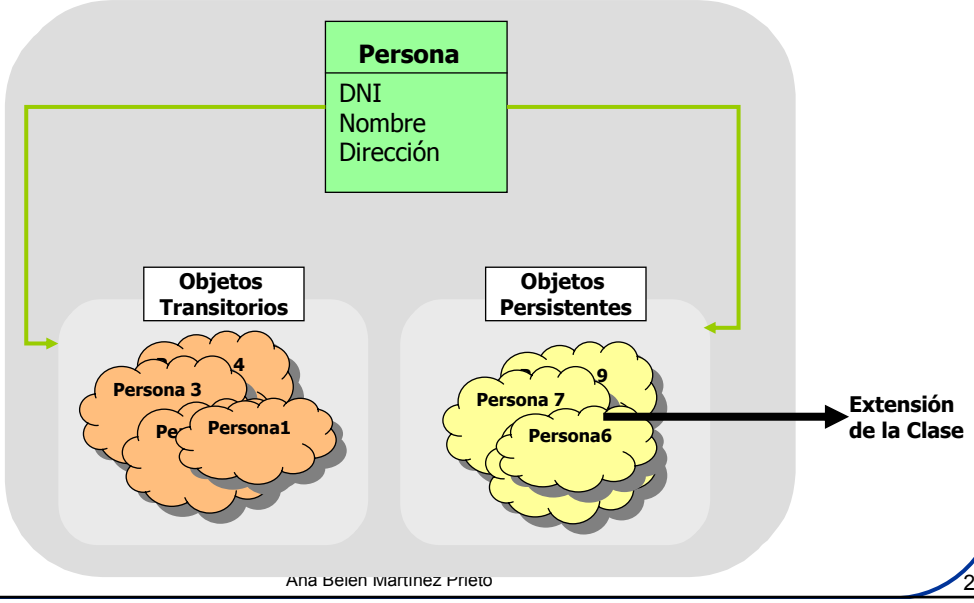
Se pretende conseguir **un lenguaje con una sintaxis simple, con un modelo y sistema de tipos unificado**



Bases de Datos Orientadas a Objetos

Características: Persistencia

- Cualidad de algunos objetos de mantener su identidad y relaciones con otros objetos con independencia del sistema o proceso que los creó.



Bases de Datos Orientadas a O

Características: Persistencia por nombramiento

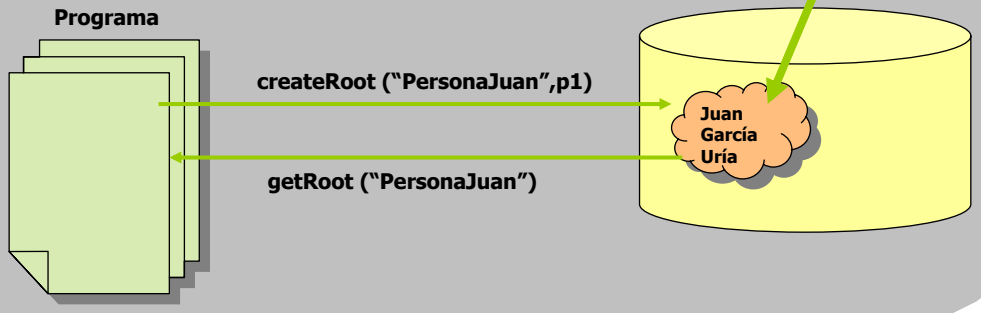
Inconveniente:

El nombrado de cada objeto a nivel individual no resulta práctico para una base de datos de tamaño considerable

- Por llamada explícita - **nombramiento**
  - Implica dar un nombre único al objeto persistente por el que podrá ser recuperado por otros programas
  - Los nombres han de ser únicos en cada base de datos

Ejemplo

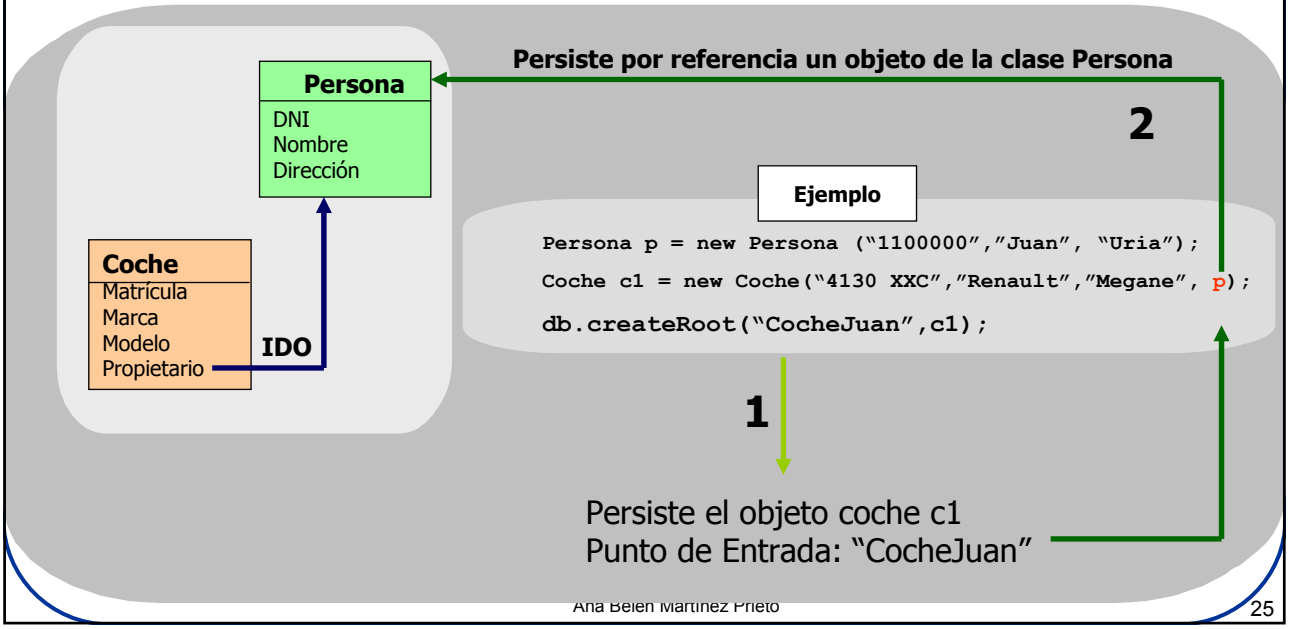
```
Persona p1 = new Persona("11111111","Juan","Uria");
db.createRoot("PersonaJuan",p1);
```



Bases de Datos Orientadas a Objetos

Características: Persistencia por alcance

- Por referencia o **alcance**
  - Un objeto se convierte en persistente si es alcanzado desde otro objeto persistente

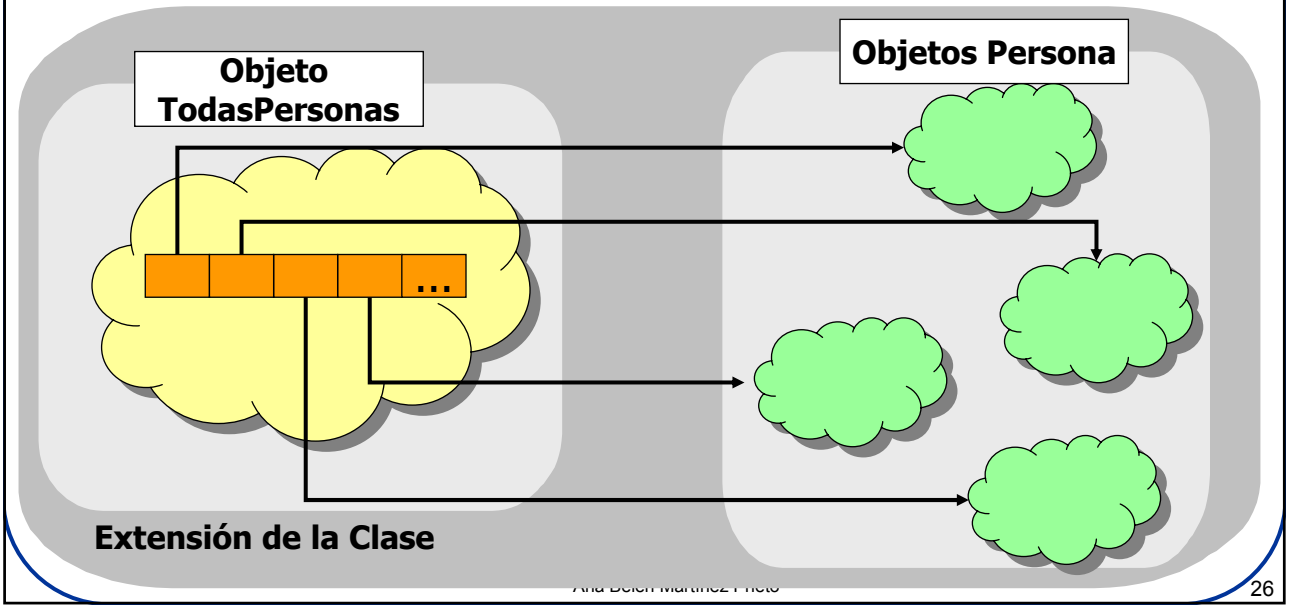


```
TodasPersonas = new OSTreeSet();
db.createRoot("TodasPersonas",TodasPersonas);
.....
Persona p = new Persona("44444444","Pedro","Avilés");
TodasPersonas.add(p);
.....
```

Bases de Datos O

Características: Persistencia por alcance + nombramiento

- Es el mecanismo más utilizado: **Nombramiento y Alcance combinados**
  - Crear un objeto persistente (con nombre) que representa un conjunto o lista de objetos de una determinada clase



Bases de Datos Orientadas a Obje

Estándar ODMG

ODMG 93 - Primera Versión
ODMG 2.0 - Segunda Versión
ODMG 3.0 - Versión actual

http://www.odmg.org

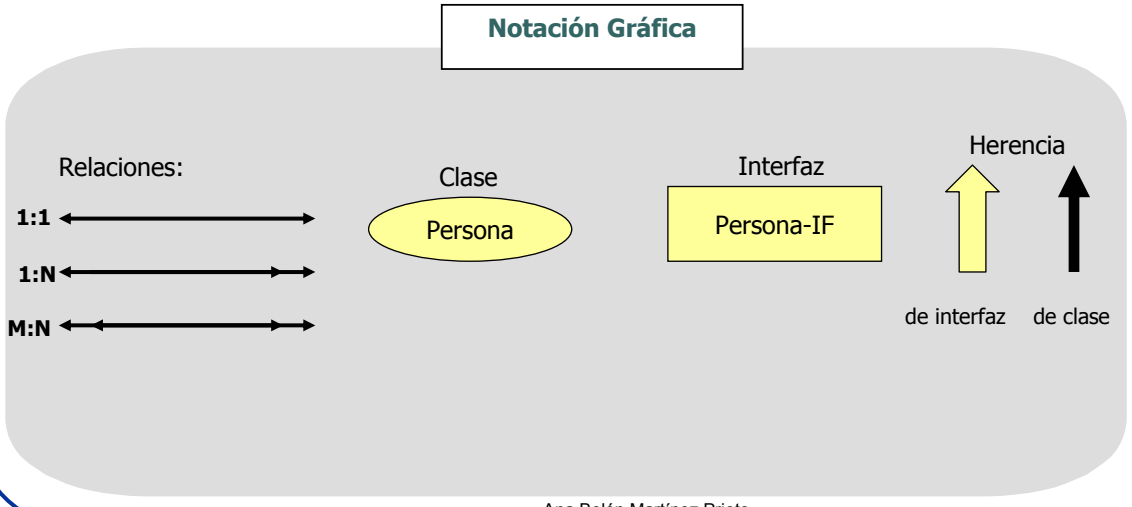
- ODMG (*Object Data Management Group*) representa el estándar para los SGBDOO confeccionado por un conjunto de vendedores de dichos sistemas
- **Componentes del Estándar:**
  - **Modelo de Objetos**
    - Está concebido para proporcionar un **modelo de objetos estándar** para las bases de datos orientadas a objetos
    - Es el modelo en el que se basan el lenguaje de definición de objetos (ODL) y el lenguaje de consultas (OQL)
  - **Lenguajes de Especificación de Objetos**
    - **Lenguaje de Definición de Objetos (ODL**, Object Definition Language)
  - **Lenguaje de Consulta**
    - Conocido como **OQL** (Object Query Language)
    - Lenguaje declarativo estilo SQL
  - **'Bindings' para C++, Java y Smalltalk**
    - Definen un lenguaje de manipulación de objetos (OML, Object Manipulation Language) que extiende el lenguaje de programación para soportar objetos persistentes. Además incluye soporte para OQL, navegación y transacciones

Bases de Datos Orientadas a Objetos
Estándar ODMG- Modelo de Objetos

- Basado en el modelo de objetos de OMG
- Principales características
  - Atributos y relaciones como propiedades de los objetos
  - Operaciones de los objetos y excepciones
  - Herencia múltiple
  - Extensiones y claves
  - Nombrado de objetos, tiempos de vida e identidad
  - Tipos atómicos, estructurados y literales colección
  - Clases colección *list, set, bag* y *array*
  - Control de concurrencia y bloqueo de objetos
  - Operaciones de base de datos

Bases de Datos Orientadas a Objetos
Estándar ODMG-Object Definition Language (ODL)

- Está diseñado para soportar constructores semánticos del modelo de objetos ODMG
- No es un lenguaje de programación completo



Bases de Datos Orientadas a Objetos

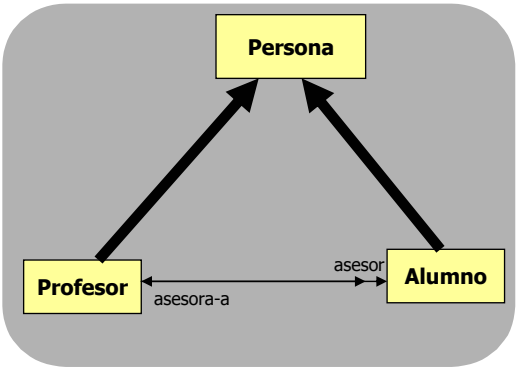
Estándar ODMG-Object Definition Language (ODL)-Ejemplo

Clase Profesor

```
class Profesor extends Persona
(extent profesores)
{
attribute string categoria;
attribute float salario;

relationship set <Alumno> asesora_a inverse Alumno::asesor;

void dar_aumento(in float aumento);
void promocionar (int string nueva_categoria);
}
```



Clase Alumno

```
class Alumno extends Persona
(extent alumnos
  key n-matricula)
{
Attribute string n-matricula;
attribute string clase;

relationship Profesor asesor inverse Profesor::asesora_a;
....
}
```

Bases de Datos Orientadas a Objetos
Estándar ODMG-Object Query Language (OQL)

- La sintaxis es similar a la de SQL con características adicionales para conceptos ODMG como:
  - identidad de objetos
  - objetos complejos
  - operaciones
  - herencia
  - polimorfismo y
  - relaciones
- Mantiene la integridad de los datos empleando las operaciones definidas sobre el objeto, en vez de sus propios operadores de actualización

Bases de Datos Orientadas a Objetos

Estándar ODMG-Object Query Language (OQL)- Ejemplos

Expresiones de Camino

```
Select e
From personas p
Where p.direccion.ciudad=Oviedo
```

Invocación de Métodos

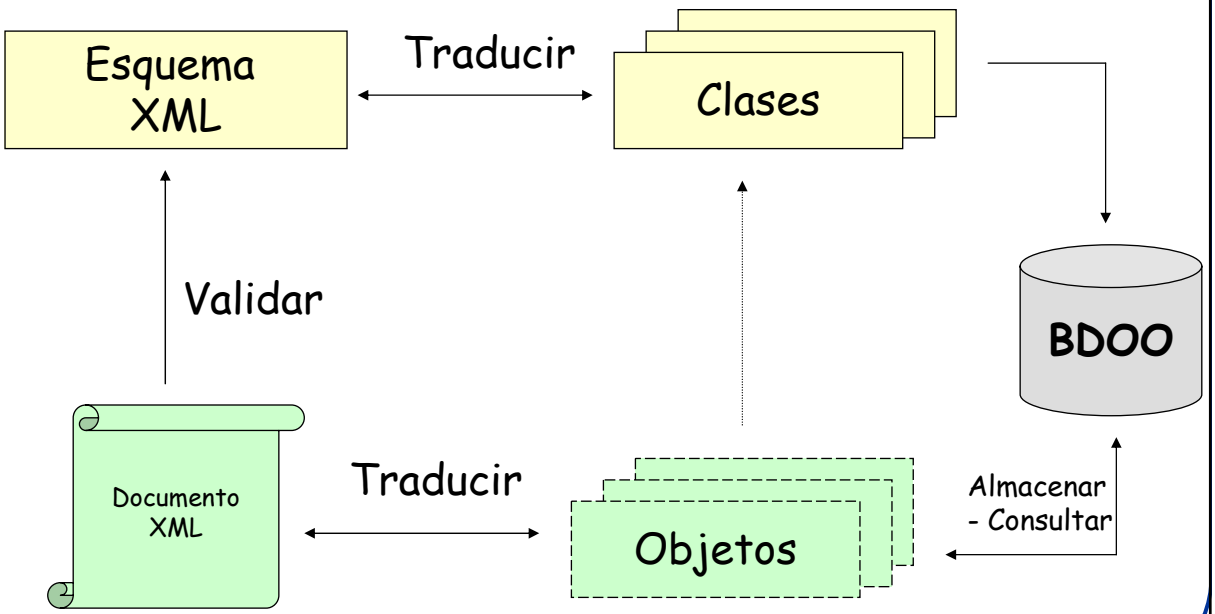
```
Select p.nombre
From personas p
Where p.vive_en("Oviedo")
```

Bases de Datos Orientadas a Objetos
Posibilidades ofrecidas para la gestión de XML: Específica

- **Específica**
  - Emplea una nueva clase para cada nuevo tipo de información. Se modelan los datos en el documento XML como un árbol de objetos pero que son específicos para los datos en el documento
  - Esta opción es especialmente apropiada para **documentos XML válidos** y para **aplicaciones centradas en los datos**
  - Existen muchos productos que implementan esta asociación de documentos XML a objetos ***(XML Data Binding)*** . Ej. JAXB

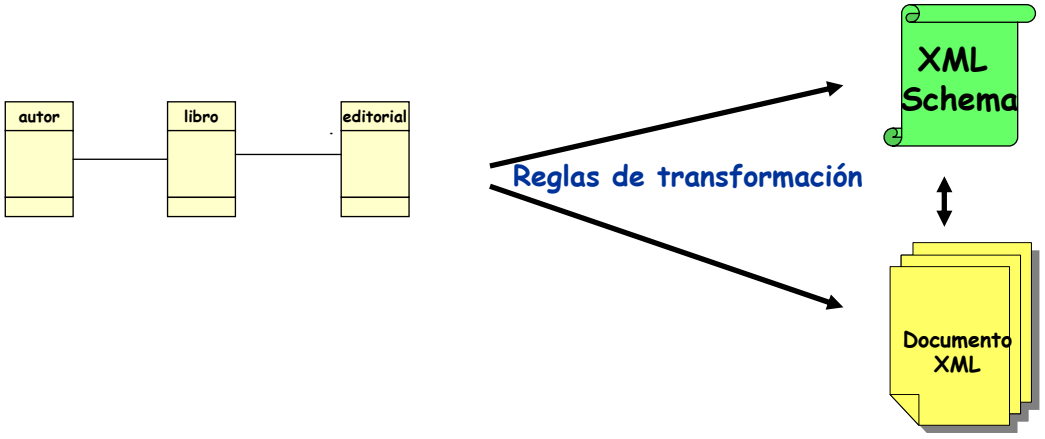
Bases de Datos Orientadas a Objetos

Posibilidades ofrecidas para la gestión de XML: Específica (II)



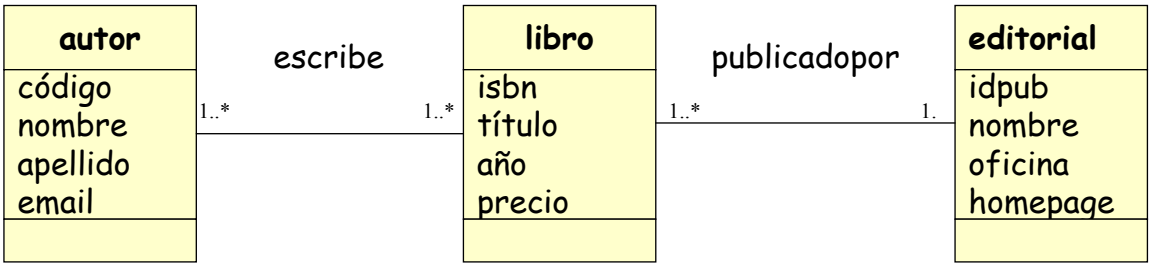
Bases de Datos Orientadas a Objetos
Transformación a XML

- Veamos un ejemplo de transformación de un esquema orientado a objetos a un documento XML



Bases de Datos Orientadas a Objetos

Ejemplo de tranformación de clases a XML Schema



Bases de Datos Orientadas a Objetos
Ejemplo de tranformación – Código ODL

```
Class libro
(extent libros key isbn) {
        attribute string isbn;
        attribute string titulo;
        attribute unsigned short año;
        attribute float precio;
        relationship set <autor> escritopor inverse autor::escribe;
        relationship editorial publicadopor inverse editorial::publica; };

Class autor
(extent autores key codigo) {
        attribute string codigo;
        attribute string nombre;
        attribute string apellido;
        attribute list<string> email;
        relationship set <libro> escribe inverse libro::escritopor; };

Class editorial
(extent editoriales key idpub) {
        attribute string idpub;
        attribute string nombre;
        attribute string oficina;
        attribute string homepage;
        relationship set <libro> publica inverse libro::publicadopor; };
```

Bases de Datos Orientadas a Objetos
Reglas de Transformación ODL – XML Schema

- Regla 1
  - Cada documento XML tiene un **elemento raíz** (puede ser el nombre de la base de datos) que debe tener asociado un **tipo complejo anónimo** que incluye un elemento especial *choice* con un atributo *maxOcurrs* con valor *unbounded*

```
<xsd:element name="libros">
   <xsd:complexType>
      <xsd:choice maxOccurs="unbounded">
```

Bases de Datos Orientadas a Objetos
Reglas de Transformación ODL – XML Schema

- Regla 2
  - **Cada clase** en el esquema ODL **se convierte en un elemento**, con el mismo nombre, que es incluido en el elemento *choice* creado en el paso anterior. A este elemento hay que asociarle un **tipo complejo con un elemento especial *sequence*** cuyo contenido se deriva de las siguientes reglas.

```
<xsd: element name="libro">
    <xsd:complexType>
        <xsd:sequence>
```

Bases de Datos Orientadas a Objetos
Reglas de Transformación ODL – XML Schema

- Regla 3
  - Los **atributos con tipos de datos básicos** (string, short, date, float, etc) se traducen en **elementos atómicos,** con el mismo nombre, que son incluidos en la secuencia, creada en el paso anterior.

```
<xsd: element name="libro">
  <xsd:complexType>
    <xsd:sequence>
      <xsd:element name="isbn" type="xsd:string"/>
      <xsd:element name="titulo" type="xsd:string"/>
      <xsd:element name="año" type="xsd:unsignedshort"/>
      <xsd:element name="precio" type="xsd:float"/>
```

Bases de Datos Orientadas a Objetos
Reglas de Transformación ODL – XML Schema

- Regla 4
  - Para los **atributos key** es necesario declarar un **elemento especial key** como un sub-elemento del elemento resultado de la regla 1. Este elemento se caracteriza por
    - un atributo *name,* cuyo valor será el nombre de la clase al que pertenece la clave, más la letra K.
    - *el atributo xpath* del sub-elemento ***selector*** debe contener una expresión como *.//nombre de la clase a la que pertenece el atributo clave.*
    - *El atributo xpath del sub-elemento* **field** debe contener una expresión Xpath como: keynombre-atributo.

```
<xsd:key name="librok">
    <xsd:selector xpath=".//libro"/>
    <xsd:field xpath="isbn"/>
</xsd:key>
```

Bases de Datos Orientadas a Objetos
Reglas de Transformación ODL – XML Schema

- Regla 5 A)
  - **Cada relación** que contiene las **palabras clave *set, list* o *bag*** se convierte en un elemento con el mismo nombre, que es incluido en la secuencia creada en la regla 2, y como la cardinalidad es mayor que 1 se le debe asociar explícitamente el atributo *maxOccurs* con un valor *unbounded*.
  - Este elemento incluirá un **tipo de dato complejo anónimo** con un elemento especial **complexContent**, que incluirá un elemento *restriction* que contiene un atributo con el nombre del atributo clave de la clase referenciada.

```
<xsd:element name="escritopor" maxOccurs="unbounded"/>
  <xsd:complexType>
     <xsd:complexContent>
        <xsd:restriction base="xsd:anyType">
             <xsd:attribute name="codigo" type=xsd:string">
        </xsd: restriction>
    </xsd:complextContent>
  </xsd:complexType>
</xsd:element>
```

Bases de Datos Orientadas a Objetos
Reglas de Transformación ODL – XML Schema

- Regla 5 B)
  - Se declarará un **elemento especial Keyref** como un sub-elemento del elemento que ha resultado de la regla 1. Este elemento tendrá
    - un atributo *name* con el nombre de la clase a la que pertenece la relación+Ref+nombre de la clase con la que se relaciona.
    - un atributo *refer* cuyo valor es el nombre del elemento clave de la clase referenciada
    - el atributo *xpath* del sub-elemento **selector** tendrá como valor .//nombre de la clase a la que pertenece la relación/nombre de la relación
    - el atributo *xpath* del sub-elemento **field** contendrá la expresión @nombre del atributo clave de la clase referenciada

```
<xsd:keyref name="libroRefAutor" refer="autorK">
        <xsd:selector xpath=".//libro/escritopor"/>
        <xsd:field xpath="@codigo"/>
</xsd:keyref>
```

Bases de Datos Orientadas a Objetos
Reglas de Transformación ODL – XML Schema

- Regla 6
  - Cada **relación que no contiene set, list o bag** se convierte en un elemento con el mismo nombre. La traducción es similar a la expresada en la regla anterior, con algunas excepciones:
    - el atributo *maxOccurs,* con valor *unbounded* no debería emplearse
    - en la declaración del **elemento keyref**
      - el atributo *xpath* del elemento **selector** contendrá el nombre de la clase a la que pertenece la relación
      - el atributo *xpath* del elemento **field** contendrá el nombre de la relación

```
<xsd:element name="publicadopor" type="xsd:string"/>
```

Bases de Datos Orientadas a Objetos
Reglas de Transformación ODL – XML Schema

- Regla 7
  - Cada **atributo de una clase** de tipo **list** se convierte en un **elemento**, con el mismo nombre, que es incluido en el elemento *sequence* que resulta de la regla 2. Este elemento incluye un **tipo de datos simple anónimo** que debe incluir un elemento *list con un atributo* itemType *cuyo valor puede ser definido por los usuarios o asignado por defecto.*

```
<xsd: element name="email">
   <xsd:simpleType>
          <xsd:list itemType="xsd:string"/>
   </xsd:simpleType>
</xsd:element>
```

Bases de Datos Orientadas a Objetos
XML Schema Resultante

```
<xsd:element name="bibdb">
        <xsd:complexType>
           <xsd:choice maxOccurs="unbounded">
             <xsd: element name="libro">
                    <xsd:complexType>
                       <xsd:sequence>
                          <xsd:element name="isbn" type="xsd:string"/>
                          <xsd:element name="titulo" type="xsd:string"/>
                           <xsd:element name="año" type="xsd:unsignedshort"/>
                           <xsd:element name="precio" type="xsd:float"/>
                           <xsd:element name="escritopor" maxOccurs="unbounded"/>
                               <xsd:complexType>
                                       <xsd:complexContent>
                                          <xsd:restriction base="xsd:anyType">
                                             <xsd:attribute name="codigo" type=xsd:string">
                                          </xsd: restriction>
                                       </xsd:complextContent>
                                </xsd:complexType>
                             </xsd:element>
                          <xsd:element name="publicadopor" type="xsd:string"/>
                        </xsd:sequence>
                      </xsd:complexType>
                </xsd:element>
. .
<xsd:key name="librok">
        <xsd:selector xpath=".//libro"/>
        <xsd:field xpath="isbn"/>
</xsd:key>

<xsd:key name="autork">
        <xsd:selector xpath=".//autor"/>
        <xsd:field xpath="codigo"/>
</xsd:key>
....
<xsd:keyref name="libroRefAutor" refer="autorK">
        <xsd:selector xpath=".//libro/escritopor"/>
        <xsd:field xpath="@codigo"/>
</xsd:keyref>
```

Bases de Datos Orientadas a Objetos

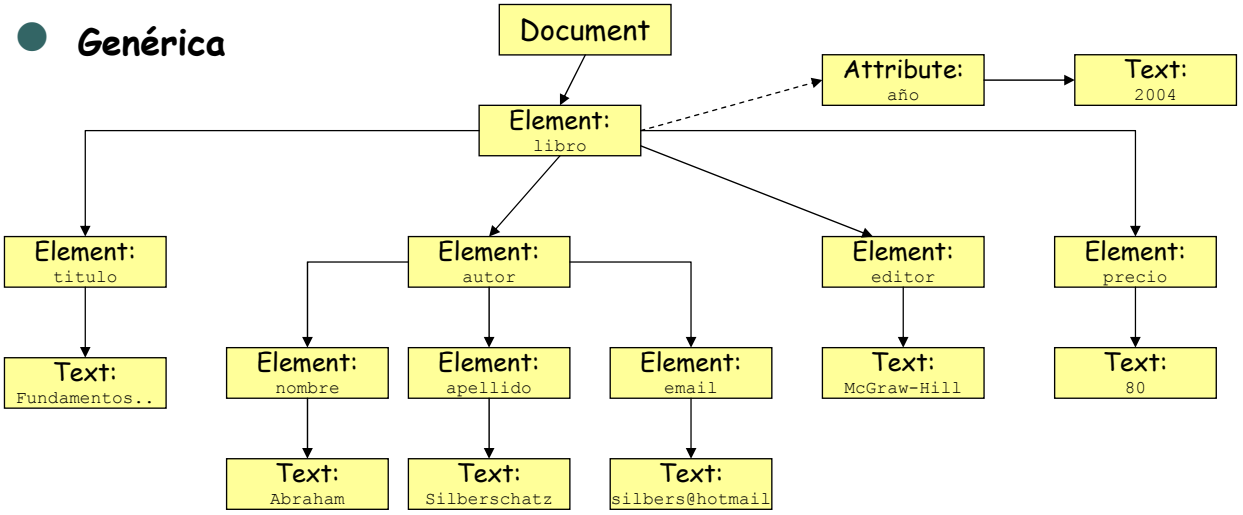
Posibilidades ofrecidas para la gestión de XML: Genérica

- **Genérica**
  - Considera toda la información del documento como objetos de clases predefinidas interconectadas mediante enlaces que preservan la estructura del documento XML
    - Ej. Únicamente hay una clase para cualquier tag XML de cualquier tipo de documento
  - DOM emplea esta aproximación y permite tratar documentos que están **bien-formados**
  - Esta opción es empleada por varias **bases de datos nativas**

Bases de Datos Orientadas a Objetos

Posibilidades ofrecidas para la gestión de XML: Genérica (II)

- **Genérica**



Bases de Datos Orientadas a Objetos

Consulta y Procesamiento

- Una vez transformado el documento XML en objetos, de forma específica o genérica, los objetos ya son gestionados directamente por el SGBDOO
- Dicha información se consulta acudiendo al **lenguaje de consulta OQL**
- Los **mecanismos de indexación, optimización, procesamiento de consultas**, etc. son las del propio SGBDOO, y por lo general, **no son específicos para el modelo XML**

Contenidos

- Introducción
- XML y Bases de Datos Relacionales
- XML y Bases de Datos Orientadas a Objetos
- **XML y Bases de Datos Nativas**
- XQuery

Bases de Datos Nativas

- **Son bases de datos**, y como tales soportan transacciones, acceso multi-usuario, lenguajes de consulta, etc., **diseñadas especialmente para almacenar documentos XML**
- Las BD nativas se caracterizan principalmente por
  - Emplear como **unidad lógica** de almacenamiento el **documento XML**
  - **Preservar el orden del documento**, las instrucciones de procesamiento, los comentarios, las secciones CDATA y las entidades
  - **Soportar lenguajes de consulta XML**
  - **No tener ningún modelo de almacenamiento físico subyacente concreto**. Pueden ser construidas sobre bases de datos relacionales, jerárquicas, orientadas a objetos o bien mediante formatos de almacenamiento propietarios

Bases de Datos Nativas
Tipos

- **Almacenamiento Basado en Texto**
  - Almacena el documento XML entero en forma de texto y proporciona alguna funcionalidad de base de datos para acceder a él.
    - Posibilidad sencilla: Almacenar el documento como un BLOB en una base de datos relacional, o mediante un fichero, y proporcionar algunos índices sobre el documento que aceleren el acceso a la información
    - Posibilidad sofisticada: Almacenar el documento en un almacén adecuado con índices, soporte para transacciones, etc.

Bases de Datos Nativas
Tipos

- **Almacenamiento Basado en el modelo**
  - Almacena un modelo binario del documento (p.e. DOM) en un almacén existente o bien específico.
    - Posibilidad 1: Traducir el DOM a tablas relacionales como Elementos, Atributos, Entidades, etc.
    - Posibilidad 2: Traducir el DOM a objetos en una BDOO
    - Posibilidad 3: Utilizar un almacén creado especialmente para esta finalidad

Bases de Datos Nativas

Ejemplos – Según sistema de almacenamiento

- **Sistema propietario**

Nombre	Licencia	API
Centor Interaction Server	Comercial	C++, Java
Coherty XML DB	Comercial	Java, JSP
dbXML	Comercial	Java, Javascript, JSP
eXtc	Comercial	
GoXML DB	Comercial	Java
Infonyte DB	Comercial	JSP, Java
Ipedo XML Database	Comercial	Java
Lucid XML DM	Comercial	Java
Tamino XML Server	Comercial	JSP
TeamXML	Comercial	Java
Virtuoso	Comercial	
XIndice	Open Source	Java
XDBM	Open Source	

Bases de Datos Nativas

Ejemplos – Según sistema de almacenamiento

- **Sistema relacional**

Nombre	Licencia	API
DBDOM	Open Source	Java
eXist	Open Source	Java
XDB	Open Soure	C++

- **Sistema orientados a objetos**

Nombre	Licencia	API
Birdstep RDM XML	Comercial	C, C++, Java
MindSuite XDB	Comercial	Java, C++
Ozone	Open Source	Java

Bases de Datos Nativas

Características Generales

- **Almacenamiento de documentos en colecciones**
  - Las colecciones juegan en las bases de datos nativas el papel de las tablas en las DB relacionales
  - Los documentos se suelen agrupar, en función de la información que contienen, en colecciones que a su vez pueden contener otras colecciones.
- **Validación de los documentos**
- **Consultas**
  - La mayoría de las BD XML soportan uno o más lenguajes de consulta
  - Uno de los más populares es Query

Bases de Datos Nativas

Características Generales (II)

- **Indexación XML**
  - Se ha de permitir la creación de índices que aceleren las consultas realizadas frecuentemente
- **Creación de identificadores únicos**
  - A cada documento XML se le asocia un identificador único por el que será reconocido dentro del repositorio
- **Actualizaciones y Borrados**
  - Las BD nativas tienen una gran variedad de estrategias para actualizar y borrar documentos
  - Muchos sistemas soportan el XUpdate para llevar a cabo esta funcionalidad

Contenidos

- Introducción
- XML y Bases de Datos Relacionales
- XML y Bases de Datos Orientadas a Objetos
- XML y Bases de Datos Nativas
- **XQuery** 

XQuery - Introducción

- *Query Working Group*, creado en octubre de 1999 por el W3C, es el grupo encargado de la elaboración del lenguaje de consulta XQuery para XML
- XQuery está todavía en proceso de elaboración.
- Se han publicado diferentes borradores de trabajo:
  - XQuery 1.0: An XML Query Language, especifica la sintaxis y una descripción informal del lenguaje
  - XML Query Requirements *(W3C Working Draft)*
  - XQuery 1.0 and XPath 2.0 Data Model *(W3C Working Draft)*
  - XQuery 1.0 Formal Semantics *(W3C Working Draft)*
  - XQuery 1.0 and XPath 2.0 Functions and Operators *(W3C Working Draft)*
  - XML Query Use Cases *(W3C Working Draft)*

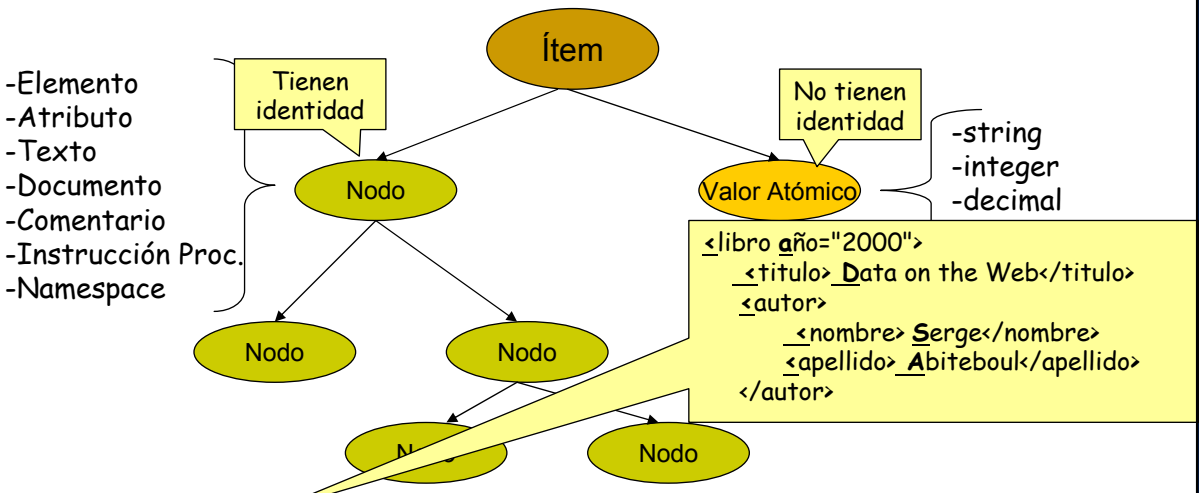
XQuery – Modelo de Datos

- La entrada y la salida de una consulta XQuery se define en términos de un modelo de datos
- El modelo de datos de la consulta proporciona una representación abstracta de uno o más documentos XML (o fragmentos de documentos)
- El modelo de datos contempla:
  - Secuencias
  - Valor especial llamado *"error value"* que es el resultado de evaluar una expresión que contiene un error

XQuery – Modelo de Datos – Características

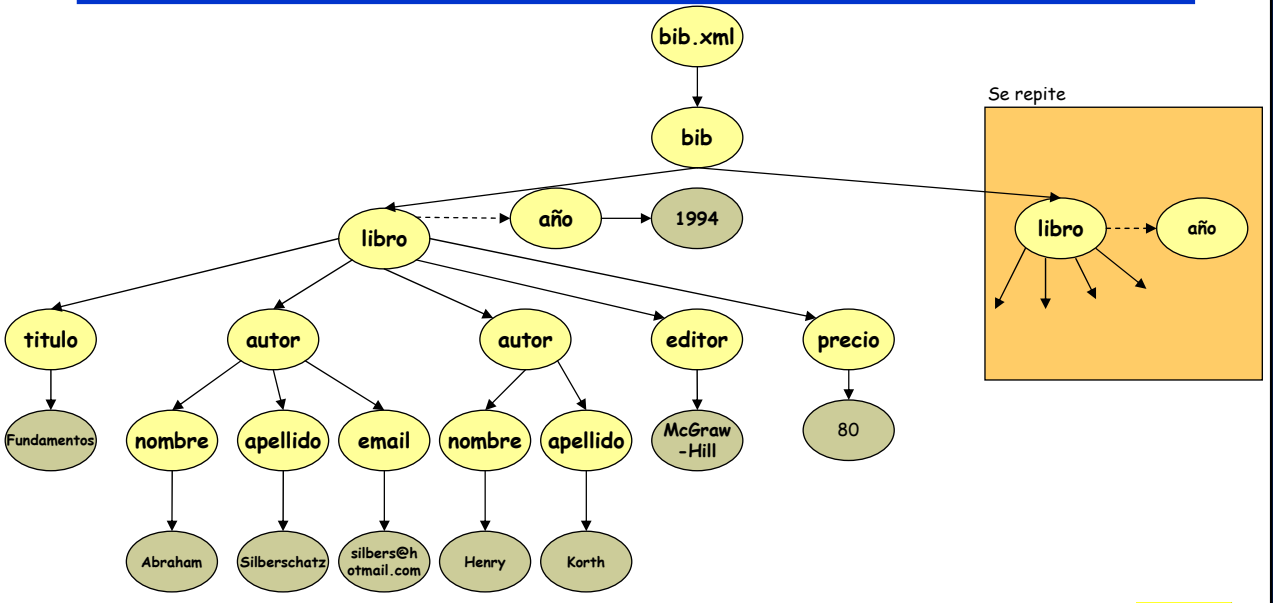
Secuencias pueden ser **heterogéneas**: pueden contener varios tipos de nodos y valores atómicos. Sin embargo, una secuencia nunca aparece como un ítem en otra secuencia

- Se basa en la noción de **secuencia**: Colección ordenada de cero o más ítems

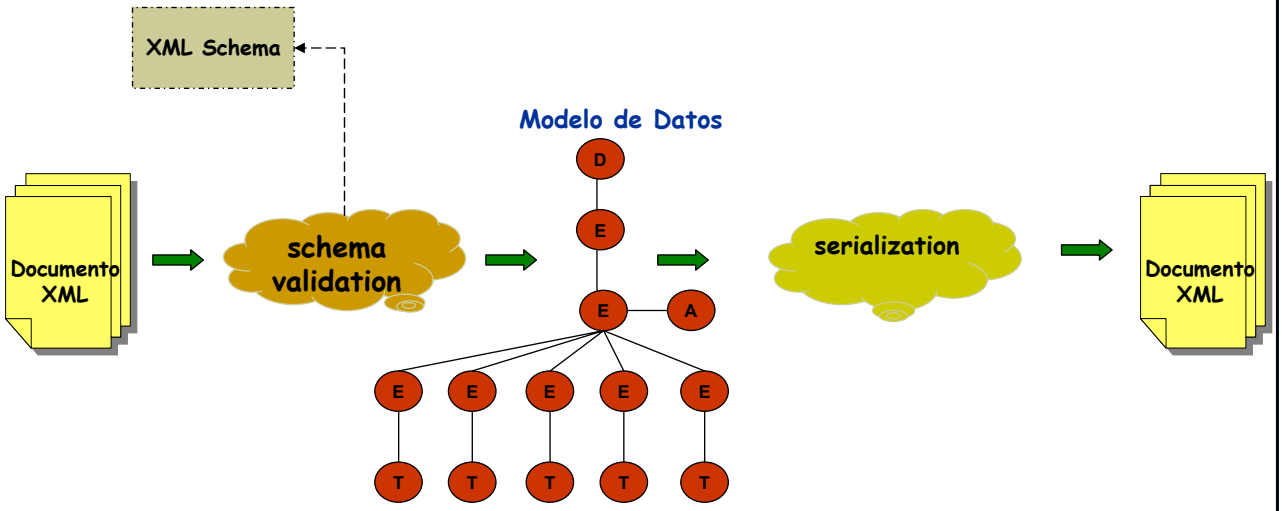


- **Orden del documento**: corresponde al orden en que los nodos aparecerían si la jerarquía de nodos fuese representada en formato XML (si el primer carácter de un nodo ocurre antes que el primer carácter de otro nodo, lo precederá también en el orden del documento)

XQuery – Modelo de Datos – Ejemplo



XQuery – Modelo de Datos – Transformación



XQuery – Expresiones

• XQuery presenta varios tipos de expresiones

• El valor de una expresión es una secuencia heterogénea de nodos y valores atómicos

•La mayoría de las expresiones son compuestas por expresiones de más bajo nivel combinadas mediante operadores y palabras reservadas

- Expresiones constituidas por Valores Atómicos
  - **53** -> literal integer
  - **5.3** -> literal decimal
  - **5.3E3** -> literal double
  - **"57"** -> literal string
  - **date ("2004-3-30")** -> invocación del constructor
  - **(2 + 4) * 5** -> los paréntesis expresan el orden de evaluación
  - **(1, 2, 3)** -> operador , concatena valores para formar una secuencia
  - **1 to 3** -> devuelve la secuencia 1, 2, 3
  - **$inicio** -> representa una variable que puede ser utilizada en una expresión
  - **Substring("Curso Extension", 1, 5)** -> invocación de funciones

• La instrucción **let**
let $inicio :=1 , $fin:=15
asocia un valor a una variable para ser empleada en una expresión

XQuery – Ejemplos (1)

- Ejemplo de invocación de funciones

document ("bib.xml")

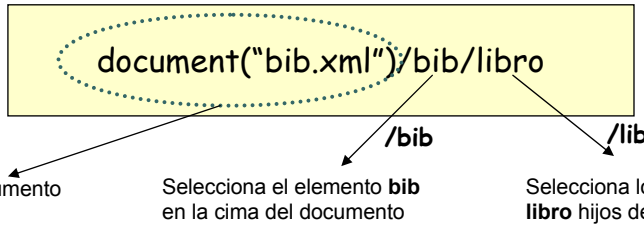
Devuelve el documento completo (el nodo documento)

XQuery – Expresiones

Obtener todos elementos ***libro*** *contenidos en el documento*

document ("bib.xml")//libro

- Expresiones constituidas por Path Expressions:
  - Path Expression en XQuery están basadas en la sintáxis de XPat
    - Está constituida por una serie de pasos separados por "/" or "//"
    - El resultado de cada paso es una secuencia de nodos
    - Su valor es la secuencia de nodos que resultan del último paso en el *path*
- Ejemplo:
  - *Obtener todos los elementos "libro" hijos del elemento "bib" en el documento "bib.xml"*



XQuery – Expresiones

- Los nodos seleccionados por cada paso sirven como **nodos de contexto** para los siguientes pasos
- Si un paso tiene varios nodos de contexto se evalua cada uno de ellos y el resultado es la **union** de los resultados dela evaluación de cada uno de ellos
- El resultado es siempre una secuencia de **nodos distintos** (no se admiten duplicados)

- Evaluación de las Path Expressions:
  - Cada paso es evaluado en el contexto de un nodo concreto (**nodo de contexto**)
  - Un paso puede ser cualquier expresión que devuelva una secuencia de nodos
  - Un **paso eje** comienza en el nodo de contexto y se desplaza a través de la jerarquía de nodos en una dirección concreta (la del eje)
  - Los criterios de selección para los nodos son:
    - **Nombre**
    - **Posición con respecto al nodo de contexto**
    - **Predicado basado en el valor de un nodo**
  - XQuery soporta 6 ejes: **child, descendant, parent, attribute, self y descendant-or-self**

- **Path Expressions** pueden ser escritas en su versión abreviada o **no abreviada**

XQuery – Expresiones - Predicados

- Un predicado es una expresión encerrada entre corchetes que se emplea para filtrar una secuencia de valores
  - Ej1. libro [titulo = " mi titulo"]
  - Ej2. libro [precio >10]
  - Ej3. libro[5] ->selecciona el quinto nodo hijo
  - Ej4. libro[precio]->selecciona los libros que tienen un nodo hijo precio
  - Ej.5 libro[@año>1997] -> selecciona los libros cuyo año es posterior al 97

XQuery – Ejemplos (2)

- Devolver los títulos de todos los libros publicados después de 1997

```
document ("bib.xml")/bib/libro[@año>1997]/titulo
```

- Devolver el primer autor de cada libro

```
document("bib.xml")/bib/libro/autor[1]
```

- Devolver los apellidos de los autores de los libros de bib.xml

```
document ("bib.xml")/bib/libro/autor/apellido
```

- Devolver los autores que tienen correo electrónico

```
document ("bib.xml")/bib/libro/autor[email]
```

XQuery – Creación de nodos

document{ } crea un nodo documento vacio

- Xquery proporciona un constructor de nodos documento:

```
document{
  <?xml-stylesheet type="text/xsl" href="c:\temp\ejemplo.xslt">
        <libro>
          <titulo> Fundamentos de Bases de Datos </titulo>
          . . .
        </libro>
}
```

- Los constructores pueden combinarse con otras expresiones XQuery para generar contenido dinámicamente
  - En un elemento constructor los { } delimitan las expresiones que son evaluadas para crear un documento abierto

XQuery – Creación de nodos - Ejemplo

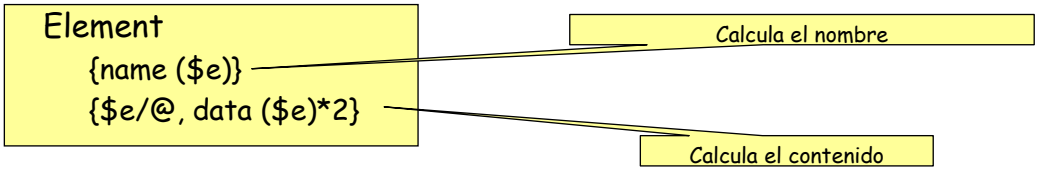
```
<ejemplo>
    <p> Esto es una consulta de ejemplo </p>
    <ej> document("bib.xml")//libro[1]/titulo </ej>
    <p> Este es el resultado de la consulta anterior </p>
    <ej> {document("bib.xml")//libro[1]/titulo } </ej>
</ejemplo>
```

Salida

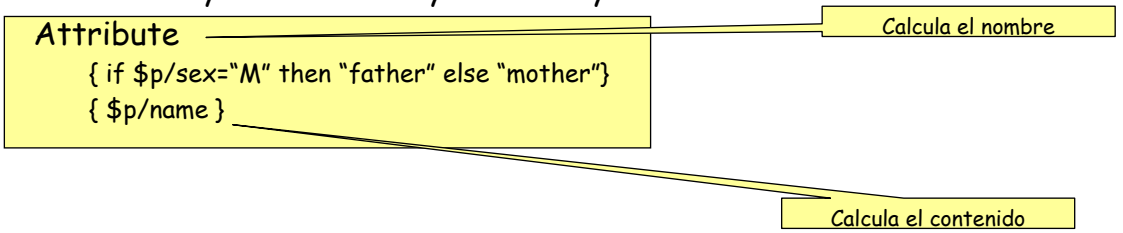
```
<ejemplo>
    <p> Esto es una consulta de ejemplo </p>
    <ej> document("bib.xml")//libro[1]/titulo </ej>
    <p> Este es el resultado de la consulta anterior </p>
    <ej> Fundamentos de Bases de Datos </ej>
</ejemplo>
```

XQuery – Expresiones - Constructores

- Constructor de Elementos Calculado
  - Construye elementos cuyo nombre y contenidos son calculados



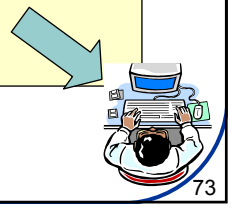
- Constructor de Atributos Calculado
  - Construye atributos cuyo nombre y contenido son calculados



XQuery – Expresiones FLWR

- FLWR (For Let Where Return) es una de las expresiones más potentes y típicas de XQuery
- Se basa en ligar valores a variables con las cláusulas for y let y emplear esas variables para crear nuevos resultados
- Ejemplo:
  - *Devolver el título de los libros que fueron publicados en el año 2000*

```
for $b in document("bib.xml")//libro
  where $b/@año="2000"
  return $b/titulo
```



XQuery – Expresiones FLWR (II)

- Una expresión FLWR
  - Comienza con una o más cláusulas *for* o *let* en cualquier orden (al menos uno de ellas)
  - Seguidas por una cláusula *where* opcional
  - Puede aparecer una cláusula *sortby* opcional
  - Finaliza con una cláusula *return* obligotoria

XQuery – Cláusula for

<fact> 2 veces 2 es 4 </fact>
<fact> 2 veces 5 es 10 </fact>
<fact> 3 veces 2 es 6 </fact>
<fact> 3 veces 5 es 15 </fact>

- **Cláusula for**
  - Proporciona una manera de iterar sobre una secuencia de valores, ligando una variable a cada uno de los valores y evaluando una expresión para cada valor de la variable
  - Ejemplos:

Secuencia de valores — Expresión a evaluar

1) for $n in (2,3) return $n+1

Resultado → (3,4)

2) for $m in (2,3), $n in (2,5)
return <fact> $m veces $n es ($m*$n) </fact>

XQuery – Cláusula let

- **Cláusula let**
  - Liga variables al resultado entero de una expresión y devuelve una única tupla
  - Ejemplo:

    for $i in (1 to 3) -> liga i a cada ítem en la secuencia 1 to 3

    let $j := (1 to $i) -> liga i a toda la secuencia (i to $i) Resultado

    $i =1, $j =1
    $i =2, $j = (1,2)
    $i =3, $j = (1,2,3)

XQuery – Cláusula where

- **Cláusula where**
  - Filtra las tuplas producidas por la cláusulas let y for.
  - Contiene una expresión que es evaluada para cada tupla. Si su evaluación es false esa tupla es descartada

XQuery – Cláusula return

- **Cláusula return**
  - Esta cláusula es ejecutada una vez para cada tupla retenida por la cláusula where
  - Los resultados de estas ejecuciones son concatenados en una secuencia que sirve como resultado de la expresión FLWR

XQuery – Cláusula sortby

- **Cláusula sortby**
  - Esta cláusula es ejecutada una vez para cada tupla retenida por la cláusula where
  - Esta cláusula es evaluada antes que la cláusula return
  - Los resultados de estas ejecuciones son concatenados en una secuencia que sirve como resultado de la expresión FLWR

```
for $i in document("bib.xml")//titulo
sortby ($i)
return $i
```

XQuery – Ejemplos FLWR

- Listar el título y precio de los libros cuyos precios sean inferiores a 100.0 euros. Sacar a continuación solo los que tengan precio

```
for $b in document("bib.xml")//libro
where $b/precio < 100.0
return <informe> {$b/titulo, $b/precio } </informe>
```

- Listar el título de cada libro junto con el número de autores

```
for $b in document ("bib.xml")//libro
let $c := $b/autor
return <libro> { $b/titulo, <numero> {count ($c)} </numero> }</libro>
```

XQuery – Ejemplos FLWR (II)

- Listar los títulos de los libros con más de dos autores

```
for $b in document("bib.xml")//libro
let $c := $b/autor
where count ($c) >2
return <autores> {$b/titulo, <numero>{count($c)}</numero>} </autores>
```

XQuery – Ejemplos FLWR (III)

- Combinar datos de dos fuentes distintas (dos ficheros xml)

- *Ejemplo:*
  - *Listar los títulos de los libros así como cualquier revisión que haya sobre ellos*

```
for $b in document("bib.xml")//titulo
for $i in document("revisores.xml")//revision
where $i/titulo = $b
return <revision>{$b, $i/comentarios} </revision>
```

XQuery – Expresiones

- Expresiones condicionales
  - if (expresion) then

    else (se ejecuta si la expresión devuelve false o una secuencia vacia)

- Ejemplo:

```
if (count($b/autor)>2)
 then <autor> et al. </autor>
else ()
```

XQuery – Expresiones

- Expresiones cuantificadas
  - Cuantificador existencial (*some*)
    - Comprueba si una condición es cierta para algún valor en la secuencia
    - Ejemplo:

```
some $n in (5, 7, 9, 11) satisfies $n > 10
```

Para cada valor se evalua la expresión de test

Devuelve true porque la expresión de test es cierta para algunos valores

  - Cuantificador universal (*every*)
    - Comprueba si una condición es cierta para cada valor en la secuencia
    - Ejemplo:

```
every $n in (5, 7, 9, 11) satisfies $n > 10
```

Devuelve false porque la expresión no es cierta para todos los valores

XQuery – Operadores

- **Comparación de valores:** **eq, ne, lt,le, gt, ge**
  - Comparan dos valores escalares y produce un error si alguno de los operandos es una secuencia de longitud mayor de 1
- **Comparación generales:** **=, !=, >, >=, <, <=**
  - Permiten comparar operandos que sean secuencias
- **Comparación de nodos:** **is e is not**
  - Comparan la identidad de dos nodos. Ej. $nodo1 is $nodo2 es true si ambas variables están ligadas al mismo nodo
- **Comparación de ordenes de los nodos:** **<<**
  - Compara la posición de dos nodos. $node1<<$node2 es true si el nodo ligado a $node1 ocurre primero en el orden del documento que el nodo ligado a $nodo2
- **Lógicos:** **and y or**
  - Se emplean para combinar condiciones lógicas dentro de un predicado. Ej. Item[seller="Smith" and precio]
- **Sobre secuencias de nodos:** **union, intersect y except**
  - Devuelven secuencias de nodos en el orden del documento y eliminan duplicados de las secuencias resultado
- **Aritméticos:** **+, -, *, div, mod**
  - Son definidos sobre valores numéricos
- **Negación:** **not**
  - Es una función más que un operador. Invierte un valor booleano.

XQuery – Funciones

- Funciones predefinidas como:
  - document()
  - **Funciones agregadas: sum, avg,count, max, min,** que operan sobre una secuencia de números y devuelven un resultado numérico
  - Substring(),upper-case(), lower-case(), concat(),…
  - Etc.
- Definidas por el usuario:

```
Define function libros-por-autor ($nombre,$apellido) as element()*
{
  for $b in document("bib.xml")/bib/libro
  where some $i in $b/autor satisfies ($i/apellido=$apellido and $i/nombre=$nombre)
  order by $b/titulo
  return $b/titulo
}
```

Direcciones de interés

XML + BD

- XML and Databases
  http://www.rpbourret.com/xml/XMLAndDatabases.htm
- XML Database Products:
  http://www.rpbourret.com/xml/XMLDatabaseProds.htm
- XML Query Engines
  - http://www.searchtools.com/info/xml-resources.html

Bibliografía

📖 Data on the Web. From relations to semistructured data and XML

Abiteboul S., Buneman P. Y Suciu Dan
Morgan Kaufmann, 2000

📖 The Object Database Standard: ODMG 3.0.

Rick Cattell et al.
Morgan Kaufmann, 1999.

📖 Succeeding with Object Databases: A practical look at today's implementations with Java and XML.

A. Chaudhri y R. Zicari et al.
John Wiley & Sons, 2000.

📖 XQuery 1.0: An XML Query Language

W3C Working Draft (November 2003)
http://www.w3.org/XML/Query