

10 XML Databases and XQuery

10.1 XML (Extensible Markup Language) and Xquery

- XML is becoming a major markup language for data interchange on the Internet.
- XML markup describes structures and semantics of data.
- Distributed databases are among the fields in which XML is especially helpful.
- XML is suitable for building heterogeneous systems:
 - XML can be used as the “common language” for data communication.
 - XML is nonproprietary.
 - XML is easy to read and write.
- The standard XML query language XQuery is being defined.
- With XQuery becomes available, XML will be used more and more in developing distributed databases.

10.2 XML Documents

- A document markuped with XML tags is called an *XML document*.
- An XML-based database is comprised of XML documents.
- The fundamental building blocks in an XML document are *elements*.
- An element may have zero or more *attributes*.
- An element may contain zero or more other elements.
- Elements and attributes in an XML document form a tree structure.

```

<?xml version="1.0">
<schema xmlns="http://www.opengis.net/gml">
<Polygon gid = 12345>
  <outerBoundaryIs>
    <LinearRing>
      <coordinates>0.0, 0.0 100.0, 0.0 100.0, 100.0
        0.0, 100.0 0.0, 0.0</coordinates>
    </LinearRing>
  </outerBoundaryIs>
  <innerBoundaryIs>
    <LinearRing>
      <coordinates>20.0, 20.0 40.0, 20.0 40.0, 80.0
        20.0, 80.0 20.0, 20.0</coordinates>
    </LinearRing>
  </innerBoundaryIs>
  <innerBoundaryIs>
    <LinearRing>
      <coordinates>60.0, 60.0 80.0, 60.0 70.0, 80.0,
        60.0, 60.0</coordinates>
    </LinearRing>
  </innerBoundaryIs>
</Polygon>

```

10.3 XML Schemas

- The schema of an XML document specifies the abstract data model of the document.
- The schema of a document **declares** classes of elements and **defines** data types.
- For each class of elements, the declaration specifies
 - their component elements and attributes, and
 - their relationships with other classes of elements.
- The declaration of a class of elements or attributes associates the name with a type.
- XML types:
 - simple types and complex types

– built-in types and user-derived types

- An important feature of XML schema is the use of `key` and `keyref` in element declarations.

```
<schema xmlns="http://www.w3.org/2000/10/XMLSchema"
  xmlns:gml="http://www.opengis.net/gml">
  ...
  <element name="Polygon" type="gml:PolygonType"/>
  <element name="LinearRing" type="gml:LinearRingType"/>
  <element name="coordinates" type="gml:CoordinateType"/>
  ...
  <complexType name="AbstractGeometryType" abstract="true">
    <attribute name="gid" type="ID" use="optional"/>
    <attribute name="srsName" type="uriReference"
      use="optional"/>
  </complexType>
  ...
  <complexType name="PolygonType">
    <complexContent>
      <extension base="gml:AbstractGeometryType">
        <sequence>
          <element name="outerBoundaryIs">
            <complexType>
              <sequence>
                <element ref="gml:LinearRing"/>
              </sequence>
            </complexType>
          </element>
          <element name="innerBoundaryIs" minOccurs="0"
            maxOccurs="unbounded">
            <complexType>
              <sequence>
                <element ref="gml:LinearRing"/>
              </sequence>
            </complexType>
          </element>
        </sequence>
      </extension>
    </complexContent>
  </complexType>
  ...
  <complexType name="LinearRingType">
  ...
```

```

</complexType>

<complexType name="CoordinateType">
    ...
</complexType>

```

10.4 The Standard XML Query Language XQuery

- Query module:
 - the *namespace declaration* section,
 - the *function definition* section, and
 - the *query expression*.
- The major query expressions:
 - XPath expressions
 - FLWR expressions
 - Element constructor expressions.
- An XPath expression is a notation of a “path” in the **schema** tree of an XML document.

```

document("polygon.xml")/Polygon/innerBoundaryIs/LinearRing
    [coordinator= "60.0, 60.0 80.0, 60.0 70.0, 80.0
                  60.0, 60.0"]

```

- A FLWR expression is constructed from FOR, LET, WHERE and RETURN clauses.

```

LET $y = document("polygon.xml")
        /Polygon/innerBoundaryIs/LinearRing
        [coordinator="60.0, 60.0 80.0, 60.0
                    70.0, 80.0 60.0, 60.0"]
FOR $x IN document("polygon.xml")
        /Polygon/innerBoundaryIs/LinearRing
WHERE area($x) > area ($y)
RETURN <boundary> { $x } </boundary>

```

- When the output of a query is required to be an XML document, the data can be organized as XML elements (and their attributes) by using the element constructor expressions.

10.5 XML and XQuery for Distributed Databases

- XQuery for data handling:
 - Functions can be defined for any applications.
 - Functions defined in a query module are used in the same module.
 - XQuery allows *function libraries*.
 - In XQuery, a function call is an expression. We can use a function in any places where expressions are allowed,
- A framework for XML-based distributed databases:
 - XML and XQuery are used as the “common languages”.
 - Each site is equipped with an XML engine.
 - The engine is the interface between the site and the rest of the distributed database.
 - The engine also serves as the interface between users and the distributed database.
 - The engine is responsible for translation.