

Solution Sheet 5

XML Data Models (InfoSet, PSVI, XDM)

Exercise 1: XDM

1.1.

An instance of the XDM is an ordered sequence of items. This way, the output of a query does not need to be a well-formed XML document. Furthermore, XML is ordered, mixing up the order can have bad results.

The XDM allows to decompose and reuse all expressions in XQuery

Items are either atomic values or nodes, so that an output of a query does not need to be an XML fragment either. This can be helpful in many situations.

1.2. In the XQuery Data Model, there are sequences, atomic types and complex types, unlike in the Information Set, where there are only DTD attribute types.

Exercise 2: Infoset

2.1. The Information Set looks like the following. Note that we did not detail the Character Information Items. For convenience, we gave names to the Items so as to refer to them, but these names are not part of the Information Set.

Document Information Item

[children] Element Information Item metadata
[document element] Element Information Item metadata
[notations] empty
[unparsed entities] empty
[base URI] .../info.xml
[character encoding scheme] UTF-8
[standalone] no value
[version] 1.0

Element Information Item metadata

[namespace name] http://www.systems.ethz.ch
[local name] metadata
[prefix] dc
[children] Element Information Items title, publisher

[attributes] empty
[namespace attributes] Attribute Information Item namespacesdc
[in-scope namespaces] Namespace Information Item namespacesdcsystems, Namespace Information Item namespacesxml
[base URI] .../info.xml
[parent] Document Information Item

Attribute Information Item namespacesdc

[namespace name] <http://www.w3.org/2000/xmlns>
[local name] dc
[prefix] xmlns
[normalized value] <http://www.systems.ethz.ch>
[specified] true
[attribute type] no value (no DTD!)
[references] unknown
[owner element] Element Information Item metadata

Namespace Information Item namespacesdcsystems

[prefix] dc
[namespace name] <http://www.systems.ethz.ch>

Namespace Information Item namespacesxml

[prefix] xml
[namespace name] <http://www.w3.org/XML/1998/namespace>

Element Information Item title

[namespace name] <http://www.systems.ethz.ch>
[local name] title
[prefix] dc
[children] Character Information Items, as many as characters in "Systems Group"
[attributes] Attribute Information Items lang, year
[namespace attributes] empty
[in-scope namespaces] Namespace Information Item namespacesdcsystems, Namespace Information Item namespacesxml
[base URI] .../info.xml
[parent] Element Information Item metadata

Attribute Information Item lang

[namespace name] <http://www.w3.org/XML/1998/namespace>
[local name] lang
[prefix] xml
[normalized value] "en"
[specified] true
[attribute type] no value (no DTD!)
[references] unknown
[owner element] Element Information Item title

Attribute Information Item year

[namespace name] empty
[local name] year

[prefix] empty
[normalized value] "2008"
[specified] true
[attribute type] no value (no DTD!)
[references] unknown
[owner element] Element Information Item title

Exercise 3: Validation and Post-Schema-Validation Infoset (PSVI)

3.1. Unlike in the Information Set, there is type information in the PSVI.

3.2. A possible schema is the following:

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  targetNamespace="http://www.systems.ethz.ch"
  xmlns:dc="http://www.systems.ethz.ch"
  elementFormDefault="qualified">
  <xs:import namespace="http://www.w3.org/XML/1998/namespace"
    schemaLocation="http://www.w3.org/2001/xml.xsd"/>
  <xs:element name="metadata" type="dc:metadata-type"/>
  <xs:complexType name="metadata-type">
    <xs:sequence>
      <xs:element name="title">
        <xs:complexType>
          <xs:simpleContent>
            <xs:extension base="xs:string">
              <xs:attribute ref="xml:lang"/>
              <xs:attribute name="year" type="xs:integer"/>
            </xs:extension>
          </xs:simpleContent>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
</xs:schema>
```

3.3.

For metadata:

[type definition type] complex
[type definition namespace] http://www.systems.ethz.ch
[type definition anonymous] false
[type definition name] metadata-type

For title:

[type definition type] complex
[type definition namespace] http://www.systems.ethz.ch
[type definition anonymous] true
[type definition name] <generated>

For xml:lang

[type definition type] simple
[type definition namespace] http://www.w3.org/XML/1998/namespace
[type definition anonymous] true
[type definition name] <generated>

For year:

[type definition type] simple

[type definition namespace] <http://www.w3.org/2001/XMLSchema>
[type definition anonymous] false
[type definition name] integer