

Technologie Obiektowe

Projekt

Temat: Generacja schematu dla plików xml
Paweł Kuształ, grupa: 1ID22A

1. Wstęp teoretyczny

XML to uniwersalny język znaczników przeznaczony do reprezentowania różnych danych w strukturalizowany sposób. Skrót XML czyli Extensible Markup Language oznacza rozszerzalny język znaczników. Został opracowany w 1998 roku przez World Wide Web Consortium (W3C). Jest językiem wywodzącym się z SGML służącym do opisu danych. Dane mają postać tekstową i przechowywane są w ściśle określonej strukturze. XML przechowuje jedynie dane, nie opisuje sposobu w jaki mają być prezentowane.

Dokument XML składa się z trzech sekcji: prologu, treści oraz epilogu z czego sekcją obowiązkową jest tylko treść.

Prolog jest pierwszą sekcją dokumentu i zawarta w nim jest deklaracja XML mówiąca o typie dokumentu i wersji XML, a także rodzaju kodowania znaków. Prolog może również zawierać deklarację DTD (Document Type Declaration).

Treść dokumentu musi się składać z jednego elementu głównego (ang. root element), wewnątrz którego może znajdować się dowolna liczba elementów. Każdy z elementów może zawierać dane oraz atrybuty.

Przykładowe elementy:

```
<ulica>Warszawska</ulica>

<miejscowosc rodzaj="miasto">Kielce</miejscowosc>

<adres>
  <miejscowosc rodzaj="miasto">Warszawa</miejscowosc>
  <ulica>Grafitowa</ulica>
  <numer>122</numer>
</adres>
```

Podstawowe zasady:

- dokument XML musi posiadać element główny,
- każdy element powinien być poprawnie zamknięty (wielkość liter ma znaczenie) np. <wiek> </wiek>,
- elementy, które zostały otwarte wewnątrz elementu rodzica muszą zostać zamknięte w tym samym elemencie tzn. muszą być poprawnie zagnieżdżone,
- atrybuty elementów muszą być ujęte w cudzysłowy.

Aby dokument XML był "poprawny" powinien zgadzać się z deklaracją DTD wewnątrz prologu lub w osobnym pliku. Innym sposobem na wykonanie walidacji dokumentu XML jest użycie pliku XSD, czyli definicji schematu XML (ang. XML Schema Definition).

XML Schema służy do definiowania struktury dokumentu XML. Pozwala zdefiniować elementy oraz ich atrybuty, zdefiniować wartości elementów, określić kolejność elementów oraz typy wartości elementów i atrybutów. Tak jak w przypadku XML, XSD posiada element główny <schema>. Element ten może posiadać dodatkowe atrybuty. Podstawowe typy elementów XML Schema znajdują się w przestrzeni nazw <http://www.w3.org/2001/XMLSchema>, najczęściej używa się dla nich prefiksów xs. W pliku użycie elementów z tej przestrzeni oznacza się w atrybucie elementu <schema>:

```
xmlns:xs="http://www.w3.org/2001/XMLSchema"
```

Najpopularniejsze typy to:

- xs:string,
- xs:decimal,
- xs:integer,
- xs:boolean,
- xs:date,
- xs:time,

oraz wiele więcej. W XML Schema można również definiować własne typy. XSD pozwala na definiowanie elementów prostych oraz złożonych. Element prosty posiada tylko tekst (np.: string, integer, date), nie może natomiast posiadać innych elementów i atrybutów. Element złożony zawiera elementy lub atrybuty lub oba. Jeśli złożony element nie zawiera innych elementów to jest elementem pustym. W zależności od potrzeb możemy zdecydować czy wszystkie podelementy muszą wystąpić w dokumencie czy nie. Używamy do tego:

- <all> - muszą wystąpić wszystkie podelementy, w dowolnej kolejności,
- <choice> - musi wystąpić tylko jeden z podanych podelementów, w ilości zgodnej z minOccurs oraz maxOccurs,
- <sequence> - muszą wystąpić wszystkie podelementy w podanej kolejności.

Przykład definicji elementu prostego:

```
<xs:element name="ulica" type="xs:string"/>
```

Przykłady definicji elementu złożonego:

```
<xs:element name="miejscowosc">
  <xs:complexType>
    <xs:simpleContent>
      <xs:extension base="xs:string">
        <xs:attribute name="rodzaj" type="xs:string" />
      </xs:extension>
    </xs:simpleContent>
  </xs:complexType>
</xs:element>
```

Atrybuty mogą być zdefiniowane tylko dla typów złożonych, jednak same są typu prostego. Atrybuty mogą mieć zdefiniowaną wartość podstawową lub stałą. Wartość podstawowa default zostaje przypisana do atrybutu automatycznie, gdy żadna wartość nie została podana. W przypadku wartości stałej fixed wartość przypisana do atrybutu automatycznie i nie można dla niej podać żadnej innej wartości. W przypadku atrybutów można również określić czy są one opcjonalne czy nie. Domyślnie każdy atrybut jest opcjonalny, aby to zmienić należy użyć `use="required"`.

W XSD można również nakładać ograniczenia na wartości elementów lub atrybutów.

Przykład elementu z ograniczeniem:

```
<xs:element name="cena">
  <xs:simpleType>
    <xs:restriction base="xs:integer">
      <xs:minInclusive value="0"/>
    </xs:restriction>
  </xs:simpleType>
</xs:element>
```

W podanym przykładzie cena nie może być mniejsza do 0.