

XML валидация чрез DTD

Цели на упражнението:

1. Валидация на XML документ с вътрешно и външно DTD
2. Използване на entity (общо, параметризирано, вложено, рекурсивно)
3. Използване на нотации
4. Упражняване на основните елементи на DTD - ID, IDREF, #REQUIRED, #IMPLIED и т.н.

Средства за XML валидация чрез DTD:

За реализация на това упражнения могат да бъдат използвани някои от следните инструменти:

- ✓ XML validator (с налична поддръжка на вътрешен и външен DTD документ)
- ✓ W3 XML validator (с налична поддръжка само на вътрешен DTD документ)
- ✓ XML/DTD валидатора наличен в Eclipse

Задача 1: Свържете и валидирайте дадения по-долу XML файл с дадения DTD документ по следните 2 начина:

- С вътрешен DTD
- С външен DTD

```
<?xml version="1.0"?>
<shiporder orderid="889923">
  <orderperson>John Smith</orderperson>
  <shipto>
    <name>Ola Nordmann</name>
    <address>Langgt 23</address>
    <city>4000 Stavanger </city>
    <country>Norway </country>
```

```
</shipto>
<item>
  <title>Empire Burlesque</title>
  <note>Special Edition</note>
  <quantity>1</quantity>
  <price>10.90</price>
</item>
<item>
  <title>Hide your heart</title>
  <quantity>1</quantity>
  <price>9.90</price>
</item>
</shiporder>
```

```
<?xml version="1.0"?>
<!DOCTYPE shiporder[
<ELEMENT shiporder (orderperson,shipto,item+)>
<ATTLIST shiporder orderid CDATA #REQUIRED>
<ELEMENT orderperson (#PCDATA)>
<ELEMENT shipto (name,address,city,country)>
<ELEMENT name (#PCDATA)>
<ELEMENT address (#PCDATA)>
<ELEMENT city (#PCDATA)>
<ELEMENT country (#PCDATA)>
<ELEMENT item (title,note?,quantity,price)>
<ELEMENT title (#PCDATA)>
<ELEMENT note (#PCDATA)>
<ELEMENT quantity (#PCDATA)>
<ELEMENT price (#PCDATA)>
]>
```

Упътване

Синтаксис за включване на външен DTD документ: <!DOCTYPE root-element SYSTEM "file-name">

Синтаксис за включване на вътрешен DTD документ: <!DOCTYPE root-element [element-declarations]>

Решение

Задача 2: За дадения по-долу XML файл създайте DTD документ и го валидирайте спрямо него.

```
<?xml version="1.0"?>
<games>
  <game score="1-1">
    <home-team>Roma</home-team>
    <ex-team>Lazio</ex-team>
    <scores>
      <score me="15">
        <player>Klose</player>
      </score>
      <score me="85" type="penalty">
        <player>Tox</player>
      </score>
    </scores>
    <yellows>
      <player>Tox</player>
      <player>Hernanes</player>
    </yellows>
    <reds>
      <player>Kjaer</player>
    </reds>
  </game>
</games>
```

Решение

Задача 3: За дадения по-долу DTD документ създайте XML файл и го валидирайте спрямо нея.

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT Chair (Professor)>
<!ELEMENT Title (#PCDATA)>
<!ELEMENT Course (Title, Description?, Instructors, Prerequisites?)>
<!ATTLIST Course
  Number (CS106A | CS106B | CS107 | CS109 | CS124 | CS143 | CS145 | CS221 | CS228 | CS229 | EE108A | EE108B | LING180) #REQUIRED
  Enrollment (1070 | 110 | 130 | 180 | 280 | 320 | 500 | 60 | 620 | 90) #IMPLIED
>
```

```

<!ELEMENT Prereq (#PCDATA)>
<!ELEMENT Lecturer (First_Name, Middle_Initial?, Last_Name)>
<!ELEMENT Last_Name (#PCDATA)>
<!ELEMENT Professor (First_Name, Middle_Initial?, Last_Name)>
<!ELEMENT Department (Title, Chair, Course+)>
<!ATTLIST Department Code (CS | EE | LING) #REQUIRED>
<!ELEMENT First_Name (#PCDATA)>
<!ELEMENT Description (#PCDATA)>
<!ELEMENT Instructors ((Lecturer, Professor*) | (Professor+, Lecturer?))>
<!ELEMENT Prerequisites (Prereq+)>
<!ELEMENT Course_Catalog (Department+)>
<!ELEMENT Middle_Initial (#PCDATA)>

```

Решение

Задача 4: За DTD документ, намираща се на адрес: <http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd>, създайте XML файл. Включете в XML файла дадения DTD документ като публичен и го валидирайте спрямо нея с W3 XML validator.

Упътване

Синтаксис за включване на публичен външен DTD документ: <!DOCTYPE rootname PUBLIC FPI URL>

Пример:

```

<?xml version="1.0" ">
<!DOCTYPE rss PUBLIC "-//RSS//DTD RSS 0.91//EN" "http://rss.com/publish/formats/rss-0.91.dtd">
<rss version="0.91">
<channel> ..... </channel>
</rss>

```

A Formal Public Identifier (FPI) is a short piece of specially formatted text that may be used to uniquely identify a product, specification or document.

An FPI is made up of 4 fields, each separated by double forward slashes (//): field 1//field 2//field 3//field 4

field 1 - Indicates whether the DTD is connected to a formal standard or not. If the DTD hasn't been approved (for example, you've defined the DTD yourself), use a hyphen (-). If the DTD has been approved by a nonstandards body, use a plus sign "+". If the DTD has been approved by a formal standards body this field should be a reference to the standard itself.

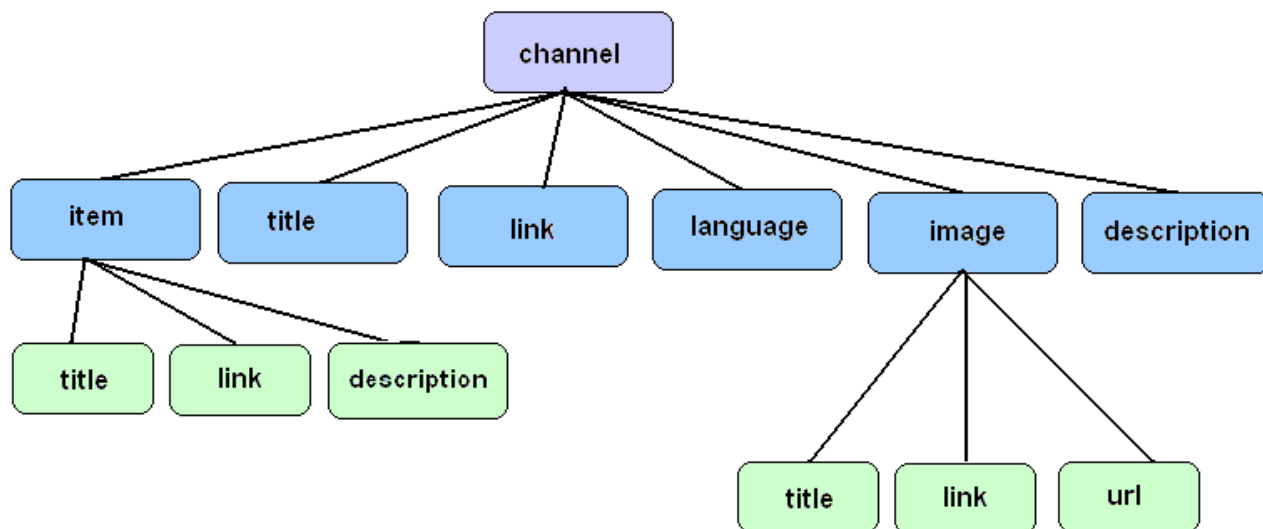
field 2 - Holds the name of the group (or person) responsible for the DTD. The above example is maintained by the W3C, so "W3C" appears in the second field.

field 3 - Indicates the type of document that is being described.

field 4 - Specifies the language that the DTD uses

Задача 5: За дадената по-долу схема създайте DTD документ и XML файл и го валидирайте спрямо нея. DTD документът трябва да изпълнява следните условия:

1. Елементът channel има атрибут с име version
2. Под-елементите на channel имат следния ред на подреждане: item, title, link, image, language и description
3. Под-елементите на channel - item, title, link и description са задължителни, а останалите елементи image и language - не
4. Елементите item и image могат да се срещат много пъти
5. Под-елементите на item (т.е. title, link, description) и image (т.е. title, link, url) нямат определена последователност
6. Под-елементът description на item е незадължителен



Задача 6: Съставете DTD документ и XML файл отразяващ резултатите от футболните мачове и включващ следната информация:

1. Футболните отбори участващи в един мач
2. Крайния резултат за всеки мач
3. Играчите отбелязали гол в мача
4. Времето, в което е отбелязан всеки гол
5. Играчите получили наказателни картони (жълти или червени)

Решение

Задача 7: Да се включи в DTD файла от задача 5 и да се използва в XML файл:

1. Една вътрешна (System) и една външна декларация (Public) на DTD нотация за някои от MIME типовете image/jpeg, image/png или image/gif. След това да се декларира entity използващо тези 2 нотации
2. Общо entity задаващо стойността на под-елемента link на елемента image
3. Параметризирано entity със стойност "title" и да се използва навсякъде където тази дума се среща в DTD файла
4. Вложено entity в entity
5. Рекурсивно entity

Упътване

Синтаксис за дефиниране на нотации: `<!NOTATION Name SYSTEM SystemLiteral>` или `<!NOTATION Name PUBLIC PublicID>` или `<!NOTATION Name PUBLIC PublicID SystemID>`. Пример: `<!NOTATION gif SYSTEM "image/gif">`, `<!NOTATION gif Public "GIF">`

Синтаксис за дефиниране на общо entity: `<!ENTITY name definition>`

Синтаксис за дефиниране на външно общо entity: `<!ENTITY name SYSTEM uri>` или `<!ENTITY name PUBLIC FPI uri>`

Синтаксис за дефиниране на общо entity, което използва нотации: `<!ENTITY name SYSTEM/PUBLIC uri NDATA notationName>`

Синтаксис за дефиниране на параметрично entity: `<!ENTITY % name definition>` или `<!ENTITY % name SYSTEM uri>` или `<!ENTITY % name PUBLIC FPI uri>`

Решение

Задача 8: За по-долу дадения XML файл и схема да се създаде DTD документ, в която са налични:

1. ID, IDREF и IDREFS, за атрибутите при които това е необходимо (напр. за InstrID, Code, Instructors)

2. #REQUIRED, за атрибутите при които това е необходимо (напр. за Number)

3. #IMPLIED, за атрибутите при които това е приложимо (напр. за Enrollment)

4. #FIXED, за атрибут по избор

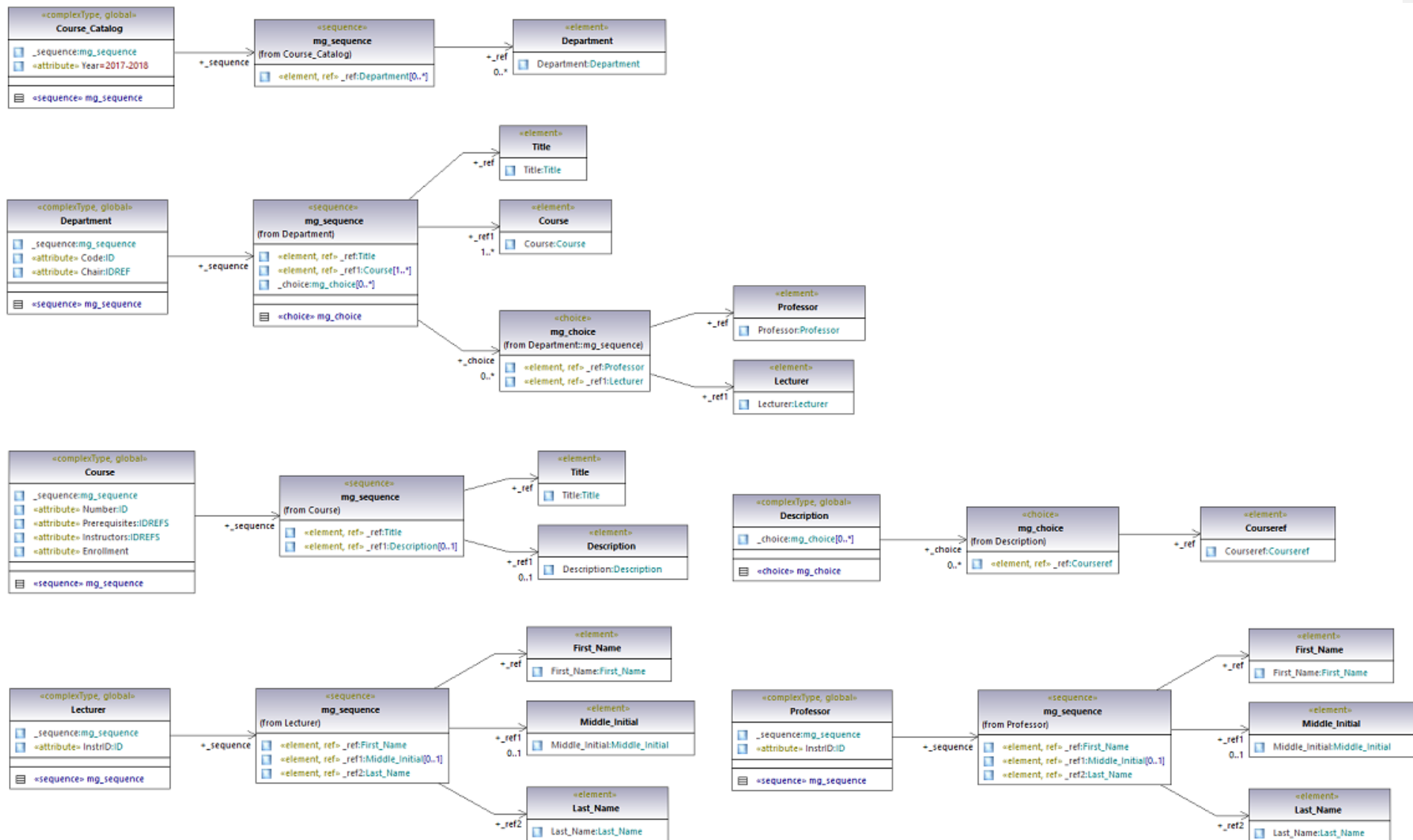
```
<?xml version="1.0" ?>
<Course_Catalog Year="2017-2018">
  <Department Code="CS" Chair="JW">
    <Title>Computer Science</Title>
    <Course Number="CS106A" Instructors="JC ER MS" Enrollment="1070">
      <Title>Programming Methodology</Title>
      <Description>Introduction to the engineering of computer applications emphasizing modern software engineering principles.</Description>
    </Course>
    <Course Number="CS106B" Prerequisites="CS106A" Instructors="JC ER" Enrollment="620">
      <Title>Programming Abstractions</Title>
      <Description>Abstraction and its relation to programming.</Description>
    </Course>
    <Course Number="CS107" Prerequisites="CS106B" Instructors="JZ" Enrollment="500">
      <Title>Computer Organization and Systems</Title>
      <Description>Introduction to the fundamental concepts of computer systems.</Description>
    </Course>
    <Course Number="CS109" Prerequisites="CS106B" Instructors="MS" Enrollment="280">
      <Title>Introduction to Probability for Computer Scientists</Title>
    </Course>
    <Course Number="CS124" Prerequisites="CS107 CS109" Instructors="DJ" Enrollment="60">
      <Title>From Languages to Information</Title>
      <Description>Natural language processing. Cross-listed as <Course Number="LING180"/>.</Description>
    </Course>
    <Course Number="CS143" Prerequisites="CS107" Instructors="AA" Enrollment="90">
      <Title>Compilers</Title>
      <Description>Principles and practices for design and implementation of compilers and interpreters.</Description>
    </Course>
    <Course Number="CS145" Prerequisites="CS107" Instructors="JW" Enrollment="130">
      <Title>Introduction to Databases</Title>
```

```
<Description>Database design and use of database management systems for applications.</Description>
</Course>
<Course Number="CS221" Prerequisites="CS107" Instructors="AN ST" Enrollment="180">
  <Title>Artificial Intelligence: Principles and Techniques</Title>
</Course>
<Course Number="CS228" Instructors="DK" Enrollment="110">
  <Title>Structured Probabilistic Models: Principles and Techniques</Title>
  <Description>Using probabilistic modeling languages to represent complex domains.</Description>
</Course>
<Course Number="CS229" Instructors="AN" Enrollment="320">
  <Title>Machine Learning</Title>
  <Description>A broad introduction to machine learning and statistical pattern recognition.</Description>
</Course>
<Professor InstrID="AA">
  <First_Name>Alex</First_Name>
  <Middle_Initial>S.</Middle_Initial>
  <Last_Name>Aiken</Last_Name>
</Professor>
<Lecturer InstrID="JC">
  <First_Name>Jerry</First_Name>
  <Middle_Initial>R.</Middle_Initial>
  <Last_Name>Cain</Last_Name>
</Lecturer>
<Professor InstrID="DK">
  <First_Name>Daphne</First_Name>
  <Last_Name>Koller</Last_Name>
</Professor>
<Professor InstrID="AN">
  <First_Name>Andrew</First_Name>
  <Last_Name>Ng</Last_Name>
</Professor>
<Professor InstrID="ER">
  <First_Name>Eric</First_Name>
  <Last_Name>Roberts</Last_Name>
</Professor>
<Professor InstrID="MS">
  <First_Name>Mehran</First_Name>
  <Last_Name>Sahami</Last_Name>
</Professor>
<Professor InstrID="ST">
```



```
<First_Name>Sebastian</First_Name>
<Last_Name>Thrun</Last_Name>
</Professor>
<Professor InstrID="JW">
  <First_Name>Jennifer</First_Name>
  <Last_Name>Widom</Last_Name>
</Professor>
<Lecturer InstrID="JZ">
  <First_Name>Julie</First_Name>
  <Last_Name>Zelenski</Last_Name>
</Lecturer>
</Department>
<Department Code="EE" Chair="MH">
  <Title>Electrical Engineering</Title>
  <Course Number="EE108A" Instructors="SM">
    <Title>Digital Systems I</Title>
    <Description>Digital circuit, logic, and system design.</Description>
  </Course>
  <Course Number="EE108B" Prerequisites="EE108A CS106B" Instructors="WD OO">
    <Title>Digital Systems II</Title>
    <Description>The design of processor-based digital systems.</Description>
  </Course>
  <Professor InstrID="WD">
    <First_Name>William</First_Name>
    <Middle_Initial>J.</Middle_Initial>
    <Last_Name>Dally</Last_Name>
  </Professor>
  <Professor InstrID="MH">
    <First_Name>Mark</First_Name>
    <Middle_Initial>A.</Middle_Initial>
    <Last_Name>Horowitz</Last_Name>
  </Professor>
  <Professor InstrID="SM">
    <First_Name>Subhasish</First_Name>
    <Last_Name>Mitra</Last_Name>
  </Professor>
  <Professor InstrID="OO">
    <First_Name>Oyekunle</First_Name>
    <Last_Name>Olukotun</Last_Name>
  </Professor>
```

```
</Department>
<Department Code="LING" Chair="BL">
  <Title>Linguistics</Title>
  <Course Number="LING180" Prerequisites="CS107 CS109" Instructors="DJ" Enrollment="60">
    <Title>From Languages to Information</Title>
    <Description>Natural language processing. Cross-listed as <Course ref Number="CS124"/>.</Description>
  </Course>
  <Professor InstrID="DJ">
    <First_Name>Dan</First_Name>
    <Last_Name>Jurafsky</Last_Name>
  </Professor>
  <Professor InstrID="BL">
    <First_Name>Beth</First_Name>
    <Last_Name>Levin</Last_Name>
  </Professor>
</Department>
</Course_Catalog>
```



Решение

Задача 9: Използвайки параметрични entity и условни секции, декларирайте по-долу описаните 2 варианта за елемента `channel` в DTD документ от задача 6. Включете/Изключете всеки един от двата варианта и направете съответните промени в XML файла

1. Вариант 1: Елементът channel включва само задължителните под-елементи дефинирани в задача 6 - т.е. item, title, link и description

2. Вариант 2: Елементът channel включва всички под-елементи дефинирани в задача 6 - т.е. item, title, link и description, image и language

Упътване

1. DTD документите дават възможност да бъдат включвани и изключвани дадени декларации от тях, чрез условните секции. Ключовата дума INCLUDE специфицира декларации, които са включени в DTD документа, а ключовата дума IGNORE обратното.
2. Синтаксис: `<![IGNORE [DTD Declarations]]>`; `<![INCLUDE [DTD Declarations]]>`
3. Обикновено условните секции се използват в комбинация с параметризирани entity по следния начин:
`<!ENTITY % ignoredPart "IGNORE">`
`<![%ignoredPart;[<!ELEMENT element (sub-element1", sub-element2, sub-element3)>]]>`
4. Условните секции могат да бъдат използвани само във външен DTD документ

Решение



1

2

3

4

5

6

7

8