DTD

## XML 문서의 검증

#### ❖제대로 만든 XML이 되려면

- Well-formed Document
  - ◆ XML의 기본 문법이 정확할 것
- Valid Document
  - ◆ Well-formed Document인가?
  - ◆ DTD 또는 Schema에 정한 형식대로 잘 작성되어 있는가
    - 문서의 구조와 적절한 요소, 속성들의 개수, 순서 등을 정의
  - ◆ 프로그램 방식으로 데이터 처리 시 문서가 valid 해야 안정성 제공 가능
- DTD: Document Type Definition
  - ◆ 하나의 시스템 내에서 사용할 XML 데이터의 구조 정의
- Schema:
  - ◆ 서로 다른 시스템 사이의 데이터를 주고 받아 사용할 때 유용

# **DTD** vs Schema

#### **❖DTD** vs Schema

	DTD	스키마
문법	XML과 유사한 문법	XML 1.0 표준에 만족
DOM 지원	DOM 기술을 지원하지 않음	DOM을 통한 조작 가능
컨텐츠 모델	순차, 선택 리스트만 제공	순차, 리스트 복합적으로 사용 가능
데이터 타입	문자열, 토큰, ID와 그 외에 한정된 타입 지원	문자열, 숫자, 날짜/시간 및 새로운 형식 정의 를 할 수 있음
Namespace	전역 이름만 사용	전역/ 로컬 이름 모두 사용 가능
상속성 -	제공하지 않음	제공
확장성	한계가 있음	제한 없음
기본 제약 조건	있음	없음
동적 스키마	불가능 – 읽기만 가능	가능 – 런타임에 상호작용으로 변경 가능

### ● Schema는 DTD를 대체하고 있음

#### DTD

#### ❖DTD를 사용하는 이유

- 응용 프로그램은 DTD를 사용해 XML 데이터가 유효한지 검증 가능
  - ◆ 독립적인 그룹의 사람들이 표준 DTD를 이용해 데이터 교환이 가능케 함

```
<?xml version="1.0" encoding="UTF-8"?>
                                                           <?xml version="1.0"?>
<!DOCTYPE note [
                                                           <!DOCTYPE note SYSTEM "note.dtd">
<!ELEMENT note (to, from, heading, body)>
                                                           <note>
<!ELEMENT to (#PCDATA)>
                                                             <to>Tove</to>
<!ELEMENT from (#PCDATA)>
                                                             <from>Jani
<!ELEMENT heading (#PCDATA)>
                                                             <heading>Reminder</heading>
<!ELEMENT body (#PCDATA)>
                                                             <body>Don't forget me this weekend!</body>
]>
                                                           </note>
                    이 형식대로 작성해야 valid
<note>
  <to>Tove</to>
  <from>Jani
                                                           <!ELEMENT note (to, from, heading, body)>
  <heading>Reminder</heading>
                                                           <!ELEMENT to (#PCDATA)>
  <body>Don't forget me this weekend</pody>
                                                           <!ELEMENT from (#PCDATA)>
</note>
                                                           <!ELEMENT heading (#PCDATA)>
                                                           <!ELEMENT body (#PCDATA)>
       내부 DTD의 예
                                                                       외부 DTD의 예
```

## DTD – XML building block

#### ❖DTD 관점에서 XML 구성 요소

● element: <>로 감싸진 태그

```
<body>some text</body>
<message>some text</message>
```

● attribute: element가 갖는 부가적인 정보 <img src="computer.gif" />

● Entity: xml 문서 내에서 특별한 의미를 갖는 문자들

 Entity References	Character	Entity References	Character	
<	<	>	>	
"	п	'	ı	
&	&			

- Pcdata(Parced Character DATA): 시작/종료 태그 사이의 텍스트
  - ◆ 파서에 의해 파싱되는 텍스트로 entity와 다른 markup으로 구성되었는지 점검됨
- cdata: 파서에 의해 파싱되지 않는 텍스트
  - ◆ 요소의 컨텐트에는 PCDATA만, 속성의 값에는 CDATA만 사용 가능

### **DTD** - Elements

```
❖DTD에서의 element 선언
<!ELEMENT element-name category>
<!ELEMENT element-name (element-content)>
● 비어있는 elements
<!ELEMENT element-name EMPTY>
Example:<!ELEMENT br EMPTY>
```

● pcdata를 갖는 elements

XML example: <br />

```
<!ELEMENT element-name (#PCDATA)>
Example: <!ELEMENT from (#PCDATA)>
```

● 어떤 contents라도 허용하는 elements

```
<!ELEMENT element-name ANY>
Example: <!ELEMENT note ANY>
```

#### **DTD** - Elements

```
❖DTD에서의 element 선언
```

● 순서대로 등장하는 자식을 갖는 elements

```
<!ELEMENT element-name (child1)>
or
```

<!ELEMENT element-name (child1,child2,...)>

```
Example: <!ELEMENT note (to,from,heading,body)>
```

- 선택적으로 자식을 갖는 element
  - <!ELEMENT note (to,from,header,(message|body))>
- 0개 이상의 여러 형태의 내용을 갖는 element
  - <!ELEMENT note (#PCDATA|to|from|header|message)\*>

### **DTD** - Elements

```
❖등장 회수의 지정
● 1회만 등장
  <!ELEMENT element-name (child-name)>
  Example:<!ELEMENT note (message)>
● 최소 1회 이상 등장
  <!ELEMENT element-name (child-name+)>
  Example:<!ELEMENT note (message+)>
● 0회 이상 등장
  <!ELEMENT element-name (child-name*)>
  Example: <!ELEMENT note (message*)>
● 0회 또는 1회 등장
  <!ELEMENT element-name (child-name?)>
  Example:<!ELEMENT note (message?)>
```

### **DTD** - Attributes

#### ❖기본 Attribute 선언 방식

<!ATTLIST element-name attribute-name attribute-type attribute-value>

DTD example: <!ATTLIST payment type CDATA "check">

XML example: <payment type="check" />

#### ● attribute 타입 항목

타입	설명	타입	설명
CDATA	값은 CDATA이다.	( <i>en1</i>   <i>en2</i>  )	값은 en1, en2, 중 하나이다.
ID	값은 유일한 id이다.		
IDREF	값은 다른 element의 id이다.	IDREFS	값은 다른 id들의 목록이다.
ENTITY	값은 entity이다.	Entity	값은 entity의 목록이다.

#### ● attribute value 항목

Value	설명	Value	설명
value	미 선언 시 사용할 attribute의 기본 값	#FIXED <i>value</i>	attribute의 값이 고정됨
#REQUIRED	attribute에 값이 필요함	#IMPLIED	attribute의 값이 옵션임

### **DTD** - Attributes

```
❖ Attribute 적용 예
            <!ATTLIST square width CDATA "0">
 DTD:
 Valid XML: <square width="100" />
            <!ATTLIST person number CDATA #REQUIRED>
 DTD:
  Valid XML: <person number="5677" />
  Invalid XML: <person />
 DTD:
              <!ATTLIST contact fax CDATA #IMPLIED>
 Valid XML:
              <contact fax="555-667788" />
 Valid XML:
              <contact />
 DTD:
            <!ATTLIST sender company CDATA #FIXED "Microsoft">
 Valid XML: <sender company="Microsoft" />
  Invalid XML: <sender company="W3Schools" />
 DTD:
              <!ATTLIST payment type (check cash) "cash">
  XML example: <payment type="check" />
              <payment type="cash" />
```

## **DTD** - Entity

```
* Entity
```

- 특별한 문자열에 대한 short cut 지정
- >등 기본 내장 Entity
- internal entity

external entity

XML example:

<author>&writer;&copyright;</author>

## **DTD** - Examples

```
*TVSCHEDULE
  <!DOCTYPE TVSCHEDULE [
  <!ELEMENT TVSCHEDULE (CHANNEL+)>
  <!ELEMENT CHANNEL (BANNER, DAY+)>
  <!ELEMENT BANNER (#PCDATA)>
  <!ELEMENT DAY (DATE, (HOLIDAY | PROGRAMSLOT+)+)>
  <!ELEMENT HOLIDAY (#PCDATA)>
  <!ELEMENT DATE (#PCDATA)>
  <!ELEMENT PROGRAMSLOT (TIME,TITLE,DESCRIPTION?)>
  <!ELEMENT TIME (#PCDATA)>
  <!ELEMENT_TITLE_(#PCDATA)>
  <!ELEMENT DESCRIPTION (#PCDATA)>
  <!ATTLIST TVSCHEDULE NAME CDATA #REQUIRED>
  <!ATTLIST CHANNEL CHAN CDATA #REQUIRED>
  <!ATTLIST PROGRAMSLOT VTR CDATA #IMPLIED>
  <!ATTLIST TITLE RATING CDATA #IMPLIED>
  <!ATTLIST TITLE LANGUAGE CDATA #IMPLIED>
  1>
```

## **DTD** - Examples

```
❖ NEWSPAPER
 <!DOCTYPE NEWSPAPER [</pre>
 <!ELEMENT NEWSPAPER (ARTICLE+)>
 <!ELEMENT ARTICLE (HEADLINE, BYLINE, LEAD, BODY, NOTES)>
 <!ELEMENT HEADLINE (#PCDATA)>
 <!ELEMENT BYLINE (#PCDATA)>
 <!ELEMENT LEAD (#PCDATA)>
 <!ELEMENT BODY (#PCDATA)>
 <!ELEMENT NOTES (#PCDATA)>
 <!ATTLIST ARTICLE AUTHOR CDATA #REQUIRED>
 <!ATTLIST ARTICLE EDITOR CDATA #IMPLIED>
 <!ATTLIST ARTICLE DATE CDATA #IMPLIED>
 <!ATTLIST ARTICLE EDITION CDATA #IMPLIED>
 <!ENTITY NEWSPAPER "Vervet Logic Times">
 <!ENTITY PUBLISHER "Vervet Logic Press">
 <!ENTITY COPYRIGHT "Copyright 1998 Vervet Logic Press">
 1>
```

## Schema - 기본 스키마 문서

```
❖ Simple Schema
 <?xml version="1.0"?>
 <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"</pre>
             targetNamespace="https://www.example.com"
                                                                     note.xsd
             xmlns="https://www.example.com"
             elementFormDefault="qualified">
 <xs:element name="note">
   <xs:complexType>
      <xs:sequence>
        <xs:element name="to" type="xs:string"/>
        <xs:element name="from" type="xs:string"/>
        <xs:element name="heading" type="xs:string"/>
        <xs:element name="body" type="xs:string"/>
     </xs:sequence>
   </xs:complexType>
 </xs:element>
 </xs:schema>
```

# Schema - 기본 스키마 문서

- ❖<schema> 태그의 선언 속성
  - XML Schema 파일의 root element

특성	의미	
	스키마 문서에서 사용될 element와 데이터 타입의 출처	
xmlns:xs	(namespace)를 밝히며 prefix 지정	
	이 스키마에 선언된 태그들(to, from, note,)의 출처	
targetNamespace	(namespace)를 밝힘 → XML 문서에서 사용	
	접두사를 사용하지 않는 문서의 기본 namespace 지정	
xmlns	일반적으로 targetNamespace 사용	
	qualified 또는 unqualified 사용	
	qualified:모든 요소는 targetNamespace에서 선언한 네임스페	
elementFormDefault	이스를 사용해야 함	
	Unqualified: targetNamespace에서 선언한 네임스페이스를	
	사용하지 않아도 됨	

## Schema - 참조

```
❖XML 문서에서 Schema 참조
```

특성	의미
xmlns	문서의 태그들이 기술된 기본 namespace 지정
xmlns:xsi	XML Schema 를 만들 때 사용한 태그들이 기술된 namespace 여기서는 schemaLocation태그를 사용하기 위해 필요
	97/AC 3CHEMIALOCATION 114 PT
xsi:schemaLocation	사용할 namespace와 사용할 xsd 파일을 공백으로 구분해서 표시

## Schema - element 선언

#### **❖ Simple element**

● 단지 텍스트 값으로만 이루어진 element로 다양한 타입 선언 가능

```
<xs:element name="xxx" type="yyy"/>
```

● 타입의 종류

<lastname>Refsnes

- ♦ xs:boolean
- ♦ xs:date <age>36</age>
- 기본 값과 고정값 표현

```
<xs:element name="color" type="xs:string" default="red"/>
<xs:element name="color" type="xs:string" fixed="red"/>
```

## Schema - attribute 선언

```
❖attribute 선언
 <xs:attribute name="xxx" type="yyy"/>
● 타입의 종류(element와 동일)
  xs:string
                      <xs:attribute name="lang" type="xs:string"/>

◆ xs:decimal

  ♦ xs:integer
                      <lastname lang="EN">Smith</lastname>
  ♦ xs:boolean
  ♦ xs:date
  ♦ xs:time
● 기본 값과 고정 값 표현, 필수 표현
 <xs:attribute name="lang" type="xs:string" default="EN"/>
 <xs:attribute name="lang" type="xs:string" fixed="EN"/>
 <xs:attribute name="lang" type="xs:string" use="required"/>
  ◆ 기본은 필수 입력 아님
                                                                     18
```

### **❖** restriction

### ● 제약 타입들

Constraint	Description
enumeration	사용 가능한 값들의 열거
length	문자열 또는 리스트 아이템의 개수로 0 이상의 값
maxExclusive	사용하려는 숫자 값의 상한 지정(포함하지는 않음: >)
maxInclusive	사용하려는 숫자 값의 상한 지정(포함함: >=)
maxLength	문자열 또는 리스트 아이템의 최대 개수로 0 이상의 값
minExclusive	사용하려는 숫자 값의 하한 지정(포함하지는 않음: <)
minInclusive	사용하려는 숫자 값의 하한 지정(포함함: <=)
minLength	문자열 또는 리스트 아이템의 최소 개수로 0 이상의 값
pattern	문자열의 패턴을 정규 표현식 형태로 표현
totalDigits	전체 자리수로 0보다 큰 값
whiteSpace	공백 문자에 대한 처리 규칙(preserve, replace, collapse)

```
* restriction
```

- XML 요소가 가질 수 있는 값의 제약 사항
- 값에 대한 제약

age의 값은 integer 타입인데 0<= age <=120이다.

Element 타입 지정

</xs:simpleType>

```
* restriction
● 입력 가능한 값의 제약
  <xs:element name="car">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="Audi"/>
                                                   car의 값은 문자열인데
        <xs:enumeration value="Golf"/>
                                                   Audi, Golf 중 하나
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
  <xs:element name="car" type="carType"/>
  <xs:simpleType name="carType">
    <xs:restriction base="xs:string">
                                                   별도의 사용자 지정 타입 사용
      <xs:enumeration value="Audi"/>
                                                    - 다른 element에서 재사용 가능
      <xs:enumeration value="Golf"/>
    </xs:restriction>
```

#### \* restriction

● 정규 표현식을 이용한 패턴 지정

- 다양한 패턴의 활용
  - ◆ [xyz]: x 또는 y 또는 z
  - ◆ [a-zA-Z][A-Z][0-9][가-힣]: 대/소문자1, 대문자 1자, 0~9의 숫자 1, 한글 1자
  - ◆ ([a-z])\* : 소문자 0개 이상
  - ◆ ([a-z][A-Z])+: 소문자 하나, 대문자 하나가 1개 이상 : aB, aBcD...
  - ◆ male|female: male 또는 female
  - ◆ [a-zA-Z0-9]{8}: 대소문자, 숫자 8자
    - {8, }: 8자 이상, {8, 10}: 8~10자

#### \* restriction

● 공백 문자 처리

공백을 있는 그대로 처리

- 사용 가능한 값
  - ◆ preserve: 기존의 공백문자를 그대로 유지
  - ◆ replace: line feeds, tabs, spaces, carriage returns 등을 space로 변경
  - ◆ collapse: 맨 앞, 뒤의 공백은 제거되고 나머지 모든 공백 문자들은 공백 하나로 대체

```
* restriction
● 길이에 대한 제약
  <xs:element name="password">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:length value="8"/>
                                                8글자로 한정
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
  <xs:element name="password">
    <xs:simpleType>
      <xs:restriction base="xs:string">
                                                최소 5글자에서
        <xs:minLength | value="5"/>
                                                최대 8글자
        <xs:maxLength value="8"/>
      </xs:restriction>
    </xs:simpleType>
  </xs:element>
```

#### Complex Element

- 다른 element 또는 attribute를 갖는 element

  - ◆ 다른 element를 포함하는 element

```
<employee>
  <firstname>John</firstname>
  <lastname>Smith</lastname>
</employee>
```

- ◆ 단순한 문자열과 속성을 갖는 element
  <food type="dessert">Ice cream</food>
- ◆ element와 text를 갖는 element

```
<description>
  It happenedon <date lang="norwegian">03.03.99</date> ....
</description>
```

```
❖Complex Element의 선언
● element 선언에 직접 선언하는 방식
  <xs:element name="employee">
                                                    complexType: 복합 타입 선언
    <xs:complexType>
      <xs:sequence>
        <xs:element name="firstname" type="xs:string"/>
        <xs:element name="lastname" type="xs:string"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
● 별도의 타입으로 지정하는 방식
  <xs:element name="employee" type="personinfo"/>
                                                    동일한 구성의 여러
  <xs:complexType name="personinfo">
                                                    element 정의에서 재사용 가능
    <xs:sequence>
      <xs:element name="firstname" type="xs:string"/>
      <xs:element name="lastname" type="xs:string"/>
    </xs:sequence>
  </xs:complexType>
```

```
❖Complex Element의 선언
```

```
● 다른 타입의 상속
 <xs:complexType name="personinfo">
   <xs:sequence>
     <xs:element name="firstname" type="xs:string"/>
   </xs:sequence>
 </xs:complexType>
 <xs:complexType name="fullpersoninfo">
   <xs:complexContent>
     <xs:extension base="personinfo"> 
       <xs:sequence>
         <xs:element name="address" type="xs:string"/>
       </xs:sequence>
                                                  complexContent: complexType<sup>©</sup>
     </xs:extension>
                                                  내용을 제한하거나 확장할 때 사용
   </xs:complexContent>
 </xs:complexType>
 <xs:element name="employee" type="fullpersoninfo"/>
```

```
Complex Empty Element
 cproduct prodid="1345" />
 <xs:element name="product">
   <xs:complexType>
     <xs:attribute name="prodid" type="xs:positiveInteger"/>
   </xs:complexType>
 </xs:element>
```

```
Containing Elements Only
xs:sequence
  ◆ 하위 element 들이 순서대로 나와야함
 <xs:element name="person">
   <xs:complexType>
      <xs:sequence>
       <xs:element name="firstname" type="xs:string"/>
        <xs:element name="lastname" type="xs:string"/>
     </xs:sequence>
   </xs:complexType>
 </xs:element>
 <person>
   <firstname>John</firstname>
   <lastname>Smith</lastname>
 </person>
```

### **❖** Text-Only Elements

● xs:simpleContent 내에서 기존 요소에 제약 (restriction) 을 가하거나 확장 (extension)해서 사용

#### **❖with Mixed Content**

</xs:element>

● attributes, elements, text를 모두 포함하는 element <letter> Dear Mr. <name>John Smith</name>. Your order will be shipped on <shipdate>2001-07-13</shipdate>. </letter> <xs:element name="letter"> <xs:complexType mixed="true"> <xs:sequence> <xs:element name="name" type="xs:string"/> <xs:element name="orderid" type="xs:positiveInteger"/> <xs:element name="shipdate" type="xs:date"/> </xs:sequence> </xs:complexType>

### **Schema - Indicators**

### ❖Indicators(지시자)

● 순서 관련 : Element의 순서를 지정

```
indicator 설명

all 하위 요소들이 아무런 순서 없이 모두 등장할 수 있음

choice 하위 요소들 중 하나만 선택해서 등장할 수 있음

sequence 하위 요소들이 순서대로 등장해야 함

<xs:element name="person">
```

### **Schema - Indicators**

### ❖Indicators(지시자)

● 그룹 관련: 관련된 element 또는 attribute의 그룹 지정

```
indicator
              설명
              여러 element 들을 그룹핑해서 하나로 사용
group name
attributeGroup name
              여러 attribute들을 그룹핑해서 하나로 사용
<xs:group name="persongroup">
  <xs:sequence>
    <xs:element name="firstname" type="xs:string"/>
    <xs:element name="lastname" type="xs:string"/>
  </xs:sequence>
</xs:group>
<xs:element name="person" type="personinfo"/>
<xs:complexType name="personinfo">
  <xs:sequence>
    <xs:group ref="persongroup"/>
    <xs:element name="nickname" type="xs:string"/>
  </xs:sequence>
</xs:complexType>
                                                                      33
```

### **Schema - Indicators**

### ❖Indicators(지시자)

● 등장 회수 관련

indicator	설명	
maxOccurs	최대 등장 회수 지정, unbounded를 사용하면 계속 추가 가능	
minOccurs	최소 등장 회수 지정	

◆ 순서 및 그룹 관련 지시자에서 기본 값은 모두 1

# Schema - <any>

```
♦ <any> Element
```

● 기존에 스키마에 선언된 element 이외의 어떤 요소로도 확장 가능

```
<xs:element name="person">
  <xs:complexType>
   <xs:sequence>
      <xs:element name="firstname" type="xs:string"/>
      <xs:element name="lastname" type="xs:string"/>
      <xs:any minOccurs="0"/>
    </xs:sequence>
  </xs:complexType>
</xs:element>
<person>
                                    <person>
  <firstname>Hege</firstname>
                                      <firstname>Stale</firstname>
  <lastname>Refsnes
                                      <lastname>Refsnes
  <children>
                                    </person>
   <childname>Cecilie</childname>
  </children>
</person>
                         <children> element가 선언되어 있다면..
```

# Schema - <anyAttribute>

```
<anyAttribute>
● 기존에 스키마에 선언된 attribute 이외의 어떤 요소로도 확장 가능
  <xs:element name="person">
    <xs:complexType>
     <xs:sequence>
       <xs:element name="firstname" type="xs:string"/>
       <xs:element name="lastname" type="xs:string"/>
     </xs:sequence>
     <xs:anyAttribute/>
    </xs:complexType>
  </xs:element>
  <person eyecolor="green">
                                    <person eyecolor="blue">
    <firstname>Hege</firstname>
                                      <firstname>Stale</firstname>
    <lastname>Refsnes
                                      <lastname>Refsnes
  </person>
                                    </person>
```