

모바일 소프트웨어 프로젝트

XML #3. DTD



2012.03.23.

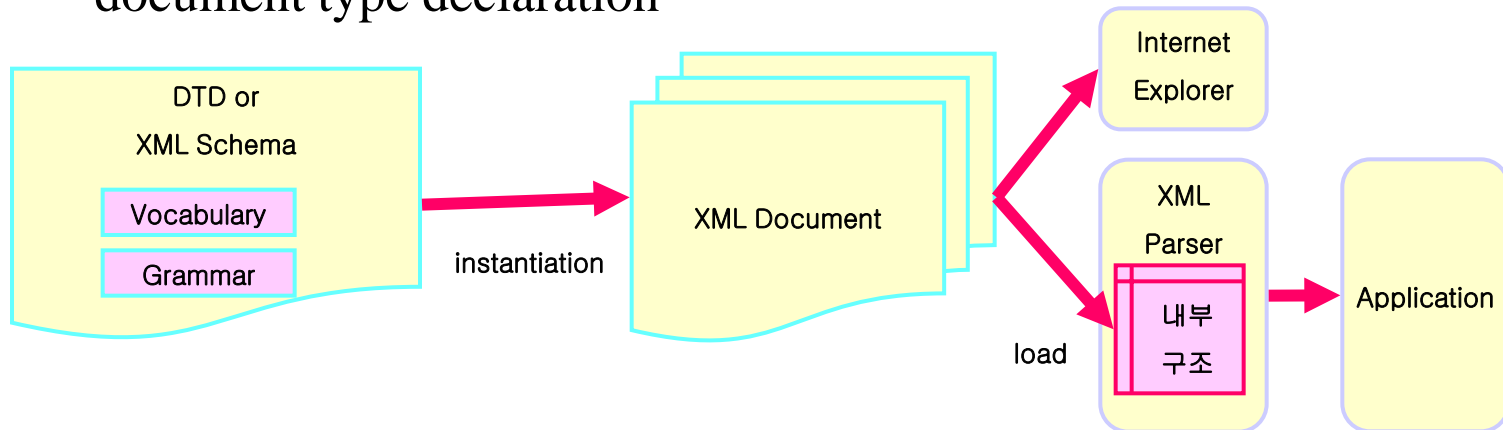
오 병 우

컴퓨터공학과

Valid XML

● DTD 또는 XML Schema를 만족시키는 Well-formed XML 문서

- ◆ Must be well formed
- ◆ Must have a document type declaration
- ◆ Root element must be the one specified by the document type declaration
- ◆ Must satisfy all the constraints indicated by the DTD specified by the document type declaration



Document Type Definition

- XML 문서의 문법 정의
- 외부 DTD를 선언하는 경우

◆ SYSTEM

- Intended for private DTDs used by a single author or group

◆ PUBLIC

- Intended for broad use and has a name made by a Organization of Standardization such as ISO

- DTD 선언 내부에 DTD가 있는 경우

```
<!DOCTYPE first [
  <!ELEMENT first (#PCDATA)>
]>
```

- 둘 모두 사용하는 경우

```
<?xml version="1.0" encoding="euc-kr"?>
<!DOCTYPE name SYSTEM "name.dtd" [
  <!ELEMENT name (first, middle, last)>
  <!ELEMENT middle (#PCDATA)>
  <!ELEMENT last (#PCDATA)>
]>
<name>
  <first>John</first>
  <middle>F.</middle>
  <last>Kennedy</last>
</name>
```

first tag는
"name.dtd"
File에서 정의

name.dtd
file

```
<?xml version="1.0" encoding="euc-kr"?>
<!ELEMENT first (#PCDATA)>
```

Root
Element
이름

URL
지정하
겠다.
c.f.) PUBLIC

"../name.dtd"
"/dtd/name.dtd"
"dtd/name.dtd" 등
가능

```
<!DOCTYPE first SYSTEM "name.dtd">
```

Element 선언

- `<!ELEMENT name (content model)>`
- “`<!ELEMENT`” 와 “`>`” 사이에 요소의 이름 및 내용 유형
 - ◆ ELEMENT는 항상 대문자
- 예제

```
<!ELEMENT first (#PCDATA)>
```



```
<first>John</first>
```

■ PCDATA

◆ Parsed Character Data

- Parsing의 대상
- Element의 데이터 타입

■ CDATA

◆ Character Data

- Attribute의 데이터 타입 (대부분의 경우 Parsing 대상)
- CDATA section의 데이터 (Parsing 대상 아님)

Content Model

- Element의 content로 어떤 것이 올지 선언
- Content의 출현 횟수 규정
- 5가지
 - ◆ Children Elements
 - ◆ PCDATA
 - ◆ Mixed Content
 - ◆ Empty Content
 - ◆ Any Content (위의 4가지 모두)
- EBNF Notation 사용
 - ◆ ?, *, +

```
<!ELEMENT discoverer (name+, year)>
<!ELEMENT weapon (part-of-body, description)?>
```

Children Elements

- 계층 구조

- ◆ Element

- EBNF를 사용하여 Cardinality (몇 개인지) 표현

- 2가지

- ◆ Sequence: , (comma)

- 자식요소들이 주어진 순서에 따라 오는 경우

```
<!ELEMENT discoverer (name+, year)>
<!ELEMENT name (#PCDATA)>
<!ELEMENT year (#PCDATA)>
```

```
<?xml version="1.0" encoding="euc-kr"?>
<!DOCTYPE discoverer SYSTEM "dino.dtd">
<discoverer>
  <name>Roniewicz</name>
  <name>Barshold</name>
  <year>1972</year>
</discoverer>
```

- ◆ Choice: | (vertical bar)

- 자식요소들이 순서에 상관 없이 선택적으로 오는 경우

```
<!ELEMENT dinosaurs (carnivore | herbivore | omnivore)*>
```

PCDATA

- 실제 데이터
 - ◆ Leaf node
 - ◆ 최소 0개 이상의 문자 데이터 (공백 가능)
- EBNF 사용 불가능
- Parsed Character

```
<!ELEMENT first (#PCDATA)>
```

Mixed Content

- Children Elements와 PCDATA 모두 포함

```
<!ELEMENT para (#PCDATA | link | emp)*>
```

`<para><emp>XML</emp>은 <link>W3C</link>가 추천하는 <emp> 차세대 인터넷 언어</emp>입니다</para>`

EMPTY and ANY Content

EMPTY

◆ Empty Element

`<!ELEMENT 구분 EMPTY>`

`<구분></구분>`
 또는
`<구분/>`

ANY

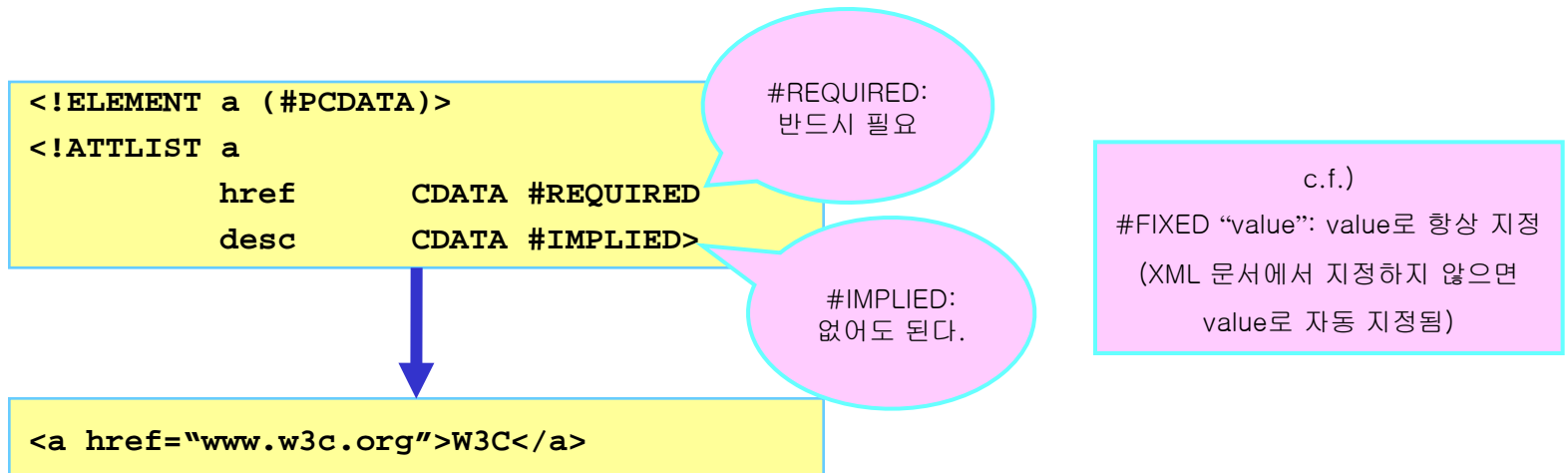
- ◆ DTD에 선언되어 있는 모든 Content Model 사용 가능
- ◆ `<!ELEMENT 의견 ANY>`
- ◆ 가능하면 사용하지 말 것

Attribute 선언

- NV-pair
 - ◆ name-value 쌍으로 속성 표현
- 속성값의 형식 지정
- 속성의 선택 사항 여부
- 속성의 기본값 부여 가능
- Attribute List

Attribute List 선언

❶ <!ATTLIST element_name (attr_name attr_type default)+>



Attribute Types

● 크게 2가지

- ◆ 문자 데이터 (1가지)
 - CDATA
- ◆ 정의된 개체 사용 등(9가지)
 - Enumeration
 - ID, IDREF, and IDREFS
 - NMTOKEN and NMTOKENS
 - ENTITY and ENTITIES
 - NOTATION

● Predefined Attributes

- ◆ xml:space
 - White space 처리
- ◆ xml:lang
 - 언어 설정

Enumeration

Attribute Value 제한

```
<!ELEMENT car (#PCDATA)>
<!ATTLIST car
    color      CDATA #REQUIRED
    maker      (현대 | 기아 | 대우 | 쌍용) "현대"
    model      CDATA #REQUIRED
    year       CDATA #IMPLIED>
```

주의

- ◆ Attribute Value 내의 공백은 오류 (다음 두 예제 모두 오류)

```
<!ATTLIST car maker (현대 |GM 대우) "현대">
<!ATTLIST car maker (현대 |"GM 대우") "현대">
```

- 인용 부호(“”)를 사용하지 않음

ID, IDREF, and IDREFS

Element Identity

```
<!ELEMENT 교수 (#PCDATA)>
<!ATTLIST 교수 번호 ID #REQUIRED>
```

- ◆ OLE/DB의 bookmark 기능
- ◆ RDB의 primary key (단, 한 개의 field일 경우)
- ◆ ID 값
 - 두 개의 Element가 같은 값 가질 수 없음
 - 대소문자 구분
 - 숫자로 시작할 수 없음

```
<교수 번호="F1234" >오병우</교수>
<과목 교수-번호="F1234">모바일SW프로젝트</과목>
```

Reference

```
<!ELEMENT 과목 (#PCDATA)>
<!ATTLIST 과목 교수-번호 IDREFS #REQUIRED>
```

- ◆ RDB의 foreign key에 해당

NMTOKEN and NMTOKENS

- White space로 구분되는 스트링

```
<!ATTLIST user      loginID NMTOKEN #REQUIRED
                password NMTOKEN #REQUIRED>
```

- 제한 조건

- ◆ CDATA와 유사함
- ◆ White space 없어야 함
- ◆ 숫자로 시작해도 상관없음

- NMTOKENS인 경우 각 token을 공백으로 분리

ENTITY and ENTITIES

- External binary 데이터 연결 시에 사용 가능

```

<!ELEMENT 학교 (#PCDATA)>
<!ATTLIST 학교 href ENTITY #IMPLIED>
<!ENTITY kumoh "http://www.kumoh.ac.kr">
<!ENTITY seonju "http://www.seonju.es.kr">
  
```

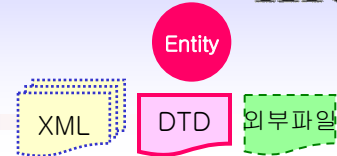


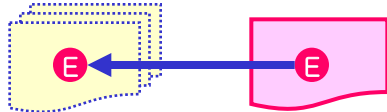
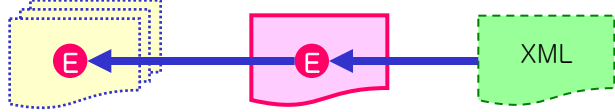
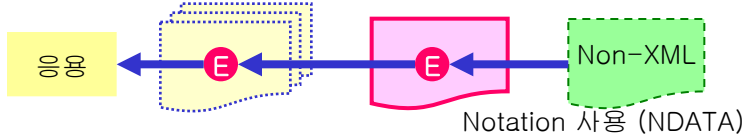
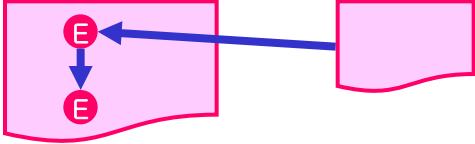
```

<학교 href="kumoh">금오공대</학교>
  
```

- Entity는 미리 선언되어 있어야만 사용 가능
- ENTITIES인 경우는 White space로 분리

참고: ENTITY의 분류



<p>Predefined Entity -XML 표준에서 정의</p>	<p>&lt; (<), &gt; (>), &amp; (&), &quot; (”), &apos; (’)</p>		
<p>General Entity -XML에서 사용 -&ent;</p>	<p>Parsed</p>	<p>Internal -DTD 내부</p>	
		<p>External -DTD 외부</p>	
	<p>Unparsed</p>	<p>External -DTD 외부</p>	 <p>Notation 사용 (NDATA)</p>
	<p>Parameter Entity -DTD에서 사용 -%ent;</p>	<p>Parsed</p>	<p>Internal -DTD 내부</p>
<p>External -DTD 외부</p>			

NOTATION

외부 응용을 위한 방법

- ◆ PI(Processing Instruction)에서 target 선언
- ◆ Attribute Type으로 사용
- ◆ General-Unparsed-External Entity로서 외부 개체 선언

```
<!NOTATION PLAY SYSTEM "tts.exe">
<!NOTATION CONV SYSTEM "Txt2Wav.exe">
<!ATTLIST text SOURCE NOTATION (PLAY | CONV) #REQUIRED>
```