

모바일 소프트웨어 프로젝트

XML #1. 마크업 언어 및 XML 개요



2012.03.05.

오 병 우

컴퓨터공학과

기술 동향

- 초고속 정보통신망
 - ◆ 인터넷 발달
 - ◆ 컨텐츠 (Web site) 발달
- 데이터와 출력 분리
 - ◆ XML과 XSL
 - MFC의 Document/View 구조와 유사
- Web site로부터 Web Service로 진화
 - ◆ Data → Function
 - ◆ Publish, Find, Bind
- Tightly coupled로부터 Loosely Coupled

Markup Language

- 문서의 기본 정보에 추가적인 정보를 제공하기 위해 사용하는 언어

◆ HTML과 XML은 마크업 언어의 일종

추가
정보
(tag)

기본
정보

추가
정보
(closing
tag)

◆ <교과목명> 모바일 소프트웨어 프로젝트 </교과목명>

마크업 언어의 역사

● GML (Generalized Markup Language)

- ◆ 1969년 IBM에서 법률문서 관리를 위해 개발

● SGML (Standard GML)

- ◆ 1986년 ISO 인증을 받아 국제표준으로 채택
- ◆ 미국, 캐나다에서 공공문서나 법률문서 관리를 위해 사용됨
- ◆ 문서 구조를 정의하기 위한 복잡한 규칙 제공

● HTML (HyperText ML)

- ◆ 1991년 SGML의 복잡한 기능을 제거하고 인터넷상에서 쉽게 사용 가능한 언어 개발

● XML (eXtensible ML)

- ◆ 1998년 W3C에서 HTML의 데이터 교환 한계를 보완하기 위해 개발

마크업의 분류

순차적 마크업 (Procedural Markup)

Procedural Markup

- ◆ 데이터 + 구조 + 서식
- ◆ Presentation 정보 제공
 - 폰트, 색상, 줄 간격 등
- ◆ E.g., HTML

교과목명	모바일소프트웨어프로젝트
학점	3학점
코드	SC1097-01, SC2097-01

표현의 차이

서술적 마크업 (Descriptive Markup)

Descriptive Markup

- ◆ 데이터 + 구조
- ◆ 문서 내용의 구조에 대한 정보 제공
- ◆ 교과목명, 학점, 코드, 수강인원 등
- ◆ E.g., XML

교과목명: 모바일소프트웨어프로젝트

학점: 3학점

코드: SC1097-01, SC2097-01

순차적 마크업

- 출력 위주
- 표현 정보와 문서 내용 통합
 - ◆ WYSIWYG 편집기에서 주로 사용
 - ◆ 표현정보가 불필요할 수 있음
 - 표에서 실제 내용만을 추출하려고 할 때
 - ◆ 문서 수정으로 표현정보 상실 가능
 - 표 삭제로 페이지 레이아웃 변경
 - ◆ 특정 S/W에서 인식
 - 다른 응용 S/W에서 인식 가능하게 하려면 재편집 또는 변환 필요
- 정보 교환을 위해서는 서술적 마크업이 바람직함

서술적 마크업

- 문서(데이터)의 내용에 구조와 의미를 추가
- 결과적으로 문서구조와 표현 분리
 - ◆ MFC의 Document/View와 유사
- 데이터 교환 형태로 적합
 - ◆ 받아서 쉽게 재가공 가능
 - ◆ E.g., SGML, XML, LaTeX

```

<과목>
  <교과목명>모바일소프트웨어프로젝트 </교과목명>
  <학점>3</학점>
  <코드>SC1097-1</코드>
  <코드>SC2097-1</코드>
</과목>
  
```

마크업 언어의 구성요소

Tag

- ◆ 괄호와 괄호 사이의 이름
- ◆ 빈태그: <태그명/>

Attribute

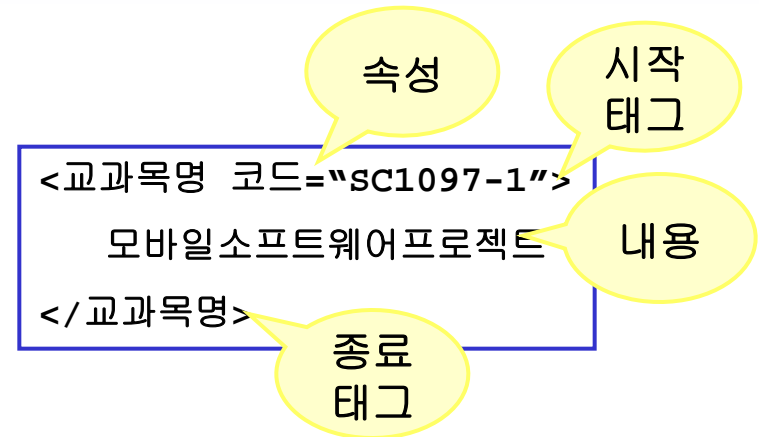
- ◆ 태그와 함께 추가적인 정보 제공
- ◆ Name-value pair
- ◆ XML에서는 value 부분에 반드시 따옴표 필요
 - HTML에서는 중간에 Blank 있을 때만 필요

Content

- ◆ 실제 데이터 (기본 정보)

Element

- ◆ 정보의 최소단위
- ◆ Element = Tag + Attribute + Content



마크업 언어 비교

언어	출시	목적	특징	단점	응용분야
SGML	1986 ISO	<ul style="list-style-type: none"> • 전자출판 	<ul style="list-style-type: none"> • 서술적 ML (tag 생성 가능) • 복잡한 규칙 	<ul style="list-style-type: none"> • Parser 없음 → Parser와 응용 S/W 자체 개발 필요 • 전문가 필요 → 대중화 실패 	<ul style="list-style-type: none"> • 공공기관, 도서관
HTML	1991 W3C	<ul style="list-style-type: none"> • 웹 페이지 제작 	<ul style="list-style-type: none"> • 순차적 ML (제한된 tag) • 사용 용이 → 대중화 	<ul style="list-style-type: none"> • 문서의 구조를 표현하기 어려움 → 비효율적 검색 	<ul style="list-style-type: none"> • 웹 어플리케이션
XML	1998 W3C	<ul style="list-style-type: none"> • 데이터 교환 • Human-readable • Machine-readable 	<ul style="list-style-type: none"> • 서술적 ML (tag 생성 가능) • 시스템 독립적 	<ul style="list-style-type: none"> • 파서는 vendor가 제공 • 응용 S/W는 자체 개발 필요 	<ul style="list-style-type: none"> • 데이터 교환을 위한 모든 분야

XML의 정의

● XML

- ◆ eXtensible Markup Language

- ◆ W3C에서 인간과 응용프로그램간, 또는 응용프로그램간에 정보를 쉽게 교환하기 위해 만든 데이터 교환 포맷

● eXtensible (확장성있는)

- ◆ Tag를 마음대로 사용자가 정의 가능

● 정보 교환 목적

- ◆ Interoperability

- ◆ Reusability

- ◆ 예제

- EDI (Electronic Data Interchange)

- 전자상거래

- Invoice, Receipt 등

XML의 특징

- 표준성
 - ◆ SGML, HTML 의 한계를 극복하기 위해 만든 차세대 표준 인터넷 언어
- 분리성
 - ◆ 표현과 내용이 분리
- 단순성, 호환성
 - ◆ HTML처럼 Text 기반이므로 human-readable
 - ◆ H/W나 S/W에 독립적.
- 수용성
 - ◆ HTML처럼 HTTP에 의해 전달
 - ◆ Web Service 적용 가능 (SOAP: Simple Object Access Protocol)
- 확장성
 - ◆ 확장성이 있어서 어떤 분야의 데이터도 정확하게 기술 가능
 - ◆ 혼란을 방지하려면 특정 분야에 대한 표준화 필요
- 정보검색의 정확성
 - ◆ XML은 정확한 검색 가능

정확한 검색

HTML vs. XML

```
<TABLE border=1>
  <TR>
    <TD>교과목명</TD><TD>모바일소프트웨어프로젝트</TD>
  </TR>
  <TR>
    <TD>학점</TD><TD>3학점</TD>
  </TR>
  <TR>
    <TD>코드</TD><TD>SC1097-01, SC2097-01</TD>
  </TR>
</TABLE>
```

Text-
based
Search

```
<과목>
  <교과목명>모바일소프트웨어프로젝트 </교과목명>
  <학점>3</학점>
  <코드>SC1097-01</코드> <코드>SC2097-01</코드>
</과목>
```

Context-
based
Search