

CBT 동적 문제은행 구성을 위한 XML DTD 정의

정휘웅, 윤애선, 권혁철
부산대학교 인지과학 협동과정

Definition of XML DTD for CBT dynamic item bank construction

Hwi-woong Jeong, Aesun Yoon, Hyukchul Kwon
Department of Cognitive Science, Pusan National university

요약

인터넷에 기반한 학습 콘텐츠가 보편화됨에 따라 이에 대한 사용자들의 요구도 다양화되고 있다. 이중 컴퓨터에 기반한 평가(CBT: Computer Based Test) 시스템은 학술적으로나 경제적으로 최근 그 방법론이 평가 및 개발 분야에 있어 다양하게 진행되고 있다. 그러나 현재 개발된 대부분의 CBT 시스템 환경은 관계형 데이터베이스에 기반하여 개발되었기 때문에 다양한 형태의 문제를 구현하기 어려울 뿐만 아니라, 관리 측면에서도 높은 비용을 요구한다. 본 연구는 이러한 문제점이 문항 정보는 동적인데 비해, 기존 전산 정보 환경은 정적인 것에서 비롯되었다 보고, 이에 대한 구조적 해결책을 제시하기 위해 최근 그 사용이 급증하고 있는 XML(eXtensible Markup Language)을 이용하였다. 컴퓨터 환경 관점과 문항 저작 관점 모두를 충족시키기 위해 XML의 요소(element) 정의는 컴퓨터 친화적(computer friendly) 요소와 콘텐츠 친화적(computer friendly) 요소로 분류하고, 각 요소별로 중복되지 않는 정보 유형을 추출하였다. 본 연구결과를 활용할 경우 기존 관계형 데이터베이스에 비해 동적인 정보를 손쉽게 구현할 수 있을 뿐만 아니라, 다른 XML 문서와 동기화 작업을 거칠 경우 문항 세트를 손쉽게 확장할 수 있어 경제적이고 관리가 용이하여 관련 분야 연구를 촉진할 것이다.

1. 서론

XML 언어가 도입된 이래 인터넷 기술은 기존 HTML(Hyper Text Markup Language) 언어를 이용한 Web service 일반도를 탈피하여 다양한 응용 분야로 그 활용도를 넓혀가고 있다. 초기 XML언어를 활용한 namespace¹는 주로 정보를 표현하는 방법에 대한 연구가 주류를 이루었으나, 최근 XML기반 namespace에 대한 연구는 보안 및 차세대 인터넷 환경인 시맨틱 웹(Semantic WEB)의 기반(RDF: Resource Description Format) 언어, 모바일 환경을 위한 스크립트 언어에 까지 영역을 넓혀, 점차 프로토콜 형식의 정보 단계에까지 진행되고 있다.

이처럼 XML이 그 사용환경을 넓혀가는 것은 기

존의 전산 정보구조가 표현하기 어려웠던 재귀적이고(recursive), 계층적인(hierarchical) 정보를 손쉽게 표현할 수 있기 때문이다.

비록 보안 및 인터넷, 프로토콜까지 XML의 활용도가 높아지고 있으나, 기존 정보의 표현 방법에 대한 연구 역시 매우 활발하게 이루어지고 있다. MathML(Mathematics Markup Language)은 수학 수식정보를 손쉽게 표현할 수 있도록 지원하며, CML(Chemical Markup Language)은 화학 구조도식을 XML로 표현할 수 있도록 지원한다. 다국적 기업 연합체에 의해 연구되고 있는 UDDI(Universal Description, Discovery and Integration)는 비즈니스 의사소통을 위해 막대한 비용을 투자하여 개발되고 있는 언어 구조체다.

이처럼 XML을 이용한 정보 인프라를 각 선진국이 적극적으로 연구를 하는 이유는 표준 정보구조를 선점하는 국가가 향후 정보 인프라 환경을 좌한다.

¹. Namespace란 XML언어의 의미적 특성을 정의하는 하나의 개념적 구조체로서, 인터넷상 하나의 URI를 지정하고, 그곳으로부터 기본 DTD를 차용

우하기 때문이다. 이에 따라 국내에서도 XML 기반 정보 인프라 표준화에 대한 연구가 활발하게 이루어지고 있으나, 외국과는 달리 표준화 보다는 XML에 기반한 어플리케이션 개발에 보다 많은 투자가 이루어지고 있어, 표준화 및 정보 구조화에 대한 연구가 시급한 실정이다.

본 연구는 이러한 정보 구조화 분야 중 학습 콘텐츠 분야에 대한 연구가 국내뿐만 아니라 국외에서도 아직 활발하게 이루어지지 않고 있는 점에 주목하고, 이중 효율성 및 재사용성이 높을 뿐만 아니라 비교적 정보가 정형화되어 있는 문항 정보에 대한 구조화를 시도하게 되었다.

우선 문항 정보를 전산정보에 원활하게 표현하기 위해 컴퓨터의 특성에 맞는 정보구조와, 사용자의 요구에 맞는 정보구조로 정보 형태를 분류하였다. 그리고 각 정보를 표현하는데 의미적으로 동일하게 표현되는 단어를 추출하여 기본 태그셋으로 정의하였다.

2장에서는 컴퓨터 환경에 적합한 정보 형태와 인간의 인지구조에 적합한 정보 형태로 분류하여 필요한 요소정보를 나눈다. 3장에서는 2장에서 분리한 정보 형태를 바탕으로 XML DTD를 정의하고, 이를 바탕으로 한 XML 문서를 제시한다. 4장에서는 본 연구 결과물의 응용방안 및 향후 연구에 대해 언급하겠다.

2. 정보의 구조화

컴퓨터 환경의 급속한 발전에도 불구하고 사용자의 정보 요구는 추상적이고 다계층화 되어있기 때문에, 그것을 전산환경에 적용하기는 매우 어렵다. 그러나 전산환경의 중요성이 날로 부각되어감에 따라 인간의 정보체계를 전산환경에 맞추기 위해 많은 인간중심 정보 구조가 전산환경에 적합하도록 조정되어 왔다. 그러나 XML 언어가 도입되기 시작하고, 컴퓨터 기술이 발전함에 따라 점차 인간 중심 정보구조에 적합한 전산환경을 개발하기 위한 연구가 활발하게 진행되고 있다. 그러나 인간이 생각하는 정보구조를 컴퓨터에 완벽하게 적용할 수 없으므로, 본 장에서는 인간에게 적합한 의미 중심 시험 정보 구조와 컴퓨터 환경에 적합한 형태 중심 시험 정보 구조로 분리하여 알아보겠다.

2.1. 의미 중심 시험 정보 구조

인간의 정보처리체계는 의미망 체계로 구성된 추상적이며 다층적인 형태를 보인다. 인간이 하나의 시험 정보를 접할 때에는 하나 하나의 문자 및 화상을 분리된 개체로 보는 것이 아니라, 하나의 통합된 시험 문제로 인식한다. 하나의 문장 혹은 화상을 읽었을 경우에도 그 의미에 주목할 뿐, 그것이 문자 정보인지 화상정보인지는 중요하게 고려하지 않는다. 오늘날 컴퓨터 환경에서 제시되는 다양한 멀티미디어 정보(화상, 동화상, 벡터 동영상, 음성)기반 시험정보들도 형태의 다양성이 있을 뿐, 전체 정보의 일부분으로 통합하여 사고하게 된다. [1]

이는 인간이 하나의 시험 정보를 주목할 때, 그 의미 및 문제를 풀어야겠다는 목적 의식에 지배를 받는다는 것을 의미한다. 따라서 인간 중심 시험 정보 구조는 의미체계를 중심으로 구성될 수 있다. 그러나 인간의 의미체계를 전산화하기에는 매우 어려우므로 본 연구에서는 시험 문항 정보를 구성하는 각 요소 정보를 추출하였다. 대상 문항은 부산대학교 언어교육원에서 개발한 CyFL 시스템을 활용하였으며, 대상 문항은 무작위로 추출한 150개로 제한하였다. 이를 바탕으로 추출한 정보는 표 1과 같다. [2]

표 1 문항 정보를 바탕으로 한 분류

종류	의미
설명문	문제의 특성을 설명하거나, 주관식인지 객관식인지를 특징짓는 문장 혹은 미디어
대화	대화체 문장에서 한 사람이 이야기하는 구어체 문장. 화자(話者)와 문장으로 구성
문장	하나의 문장을 인용할 때 구성될 수 있는 하나의 문단(paragraph)
예시	예시는 다수의 대화 혹은 문장을 포함하는 구조로서, 인간이 받아들이는 예문 정보
문제 지문	객관식인 경우 각각 문제의 지문

2.2. 형태 중심 시험 정보 구조

지금까지 컴퓨터에 정보를 저장하기 위해서는 파일시스템을 이용하거나 관계형 데이터베이스를 이용해야 했다. 따라서 의미 중심 정보 구조에 비해 유연성이 떨어지고, 구조화를 하기 위해서는 경험이 풍부한 데이터베이스 구조 전문가가 요구되었다.

컴퓨터 환경에서 중요한 정보의 분류 범주는 정보 파일의 유형 등 정보의 형태로서, 정보의 분류는 텍스트, 이미지, 음성, 동화상으로 나눌 수 있다. 인간의 관점에서 고려하는 정보 형태인 문항의 의미적 정보는 컴퓨터 환경에서 단순한 텍스트 정보나 화상 정보로 분류된다. 이러한 형태 중심 시험 정보구조를 분류하면 표2와 같다.

표 2 컴퓨터 중심 정보 분류

유형	종류	요소정보
문자	글꼴, 색상, 크기, 굵기, decoration	색상값, 크기, 각종 Boolean 정보
화상	그림, 벡터 그림, 동영상, flash data	
음성	모든 유형의 정보에 의존적	음성 포맷, URI 정보

표2에서 음성 정보를 독립적으로 분리한 것은, 음성 정보는 반드시 하나의 가시적인 정보와 연계되어야 하기 때문이다. 지필 테스트의 경우 카세트 테이프와 같은 매체 미디어를 활용하나, 컴퓨터 환경에서는 시험 당사자가 정보를 구동시켜야 하기 때문에, 음성 정보는 시각적으로 확인 가능한 매체와 연계되어야만 한다.

3. XML DTD의 정의

본 장에서는 2장에서 제시한 정보 구조를 바탕으로 XML 기반 DTD를 정의한다. XML 문서는 요소 기반 접근법으로 설계하여, 각 노드 정보가 추가적인 자식 요소(child elements)를 손쉽게 포함할 수 있도록 구성하였다. 그러나 화상의 URI(Universal Resource Identifier)정보나 문항의 번호와 같이 최소 정보 단위로 구성되는 항목에 대해서는 attribution 형태로 정보를 나누었다.

3.1. 콘텐츠 기반 정보 구조

콘텐츠 기반 정보 구조는 2장에서 언급한 인간 중심 시험정보 구조를 바탕으로 구성하였다. 각 태그의 설명은 인간의 정보 개념도와 상통할 수 있도록 표3과 같이 구성하였다. 모든 태그의 중심인 ROOT 태그는 XML 문서가 하나의 root 태그만을 가질 수 있기 때문에, 전산 환경에 적합하도록 정의한 태그다. ROOT 태그는 3.2절에서 소개할 다양한 속성(attribute) 정보를 포함함으로써, 전산 환경과 인간 중심 환경의 차이점을 완화시켜주는 역할을 담당한다.

표 3. 정보 범주에 따른 요소 분류

요소명	기능.설명
ITEM	문항이라는 것을 설명하기 위한 개념적 부모 노드. ROOT 태그의 자식 노드로서 구성된다.
QUESTION	하나의 문항에서 질문들을 묶어주기 위한 부모 노드
ANSWER	문항이 주관식일 경우 사용자가 답할 수 있는 가능한 경우를 설명한다. 자연언어 처리 개념을 도입할 경우 그 경우의 수는 줄어들 수 있으나, 현재 주관식 문항 처리 관점에서는 case by case로 처리하는 방안이 가장 타당하다고 본다.
ROOT	XML 문서를 구성하는 바탕 구조. 각 문항정보를 포괄적으로 담으며, 문항 라이브러리의 베이스가 됨.
FEEDBACK	문항의 각 항목에 문항 저작자가 부연 설명 및 학습 안내 정보를 담을 경우 지정한다.
CORRECT	해당 문항에 대한 정답 정보를 의미한다. 주관식일 경우 나열형 정보가 되며, 객관식일 경우 하나 혹은 그 이상의 항목이 구성된다.
CASE	만약 객관식일 경우 하나의 단일 문항 항목을 의미한다.

주요 요소의 주종관계는 ROOT 태그가 ITEM 태그들을 포함하며, ITEM 태그는 질문 영역과 대담 영역으로 분리된다. 대담 영역은 객관식인 경우 CASE 태그로서 분리되며, 주관식인 경우 CORRECT 태그로서 표현된다.

정보 형태별로 구성된 태그세트는 보다 미시적

인 정보들을 포함할 수 있다. 각 의미기반 요소 정보들은 상이한 형태의 정보를 포함할 수 있으며, 표4와 같이 정의하였다. 본 정의에서 DESC, PARA, DIAG와 같이 문자 기반 정보들이 포함된 것은, 의미 중심 태그 분류가 아닌, 형태 중심 분류이기 때문이다. 설명문과 문단, 대화체 문장은 의미적으로는 어떠한 요소에도 들어갈 수 있으나, 전산 환경에서는 본 요소들이 분리될 경우 언어 라이브러리를 구축하는데 소요되는 전산 프로세스 비용을 상당부분 줄일 수 있다.

표 4. 정보형태에 따른 요소 정의

요소명	기능설명
DESC	문항에 대한 간단한 설명문을 저장한다.
PARA	문서에 들어가는 일반적인 문단 정보를 저장한다.
DIAG	대화형 문체 정보를 저장한다. 화자 정보와 구술 정보를 저장한다.
IMG	HTML에서와 같이 화상 정보를 지칭한다. GIF 혹은 Jpg 형태.
ANI	Macromedia Shockwave flash 미디어 정보를 지정한다. 이 경우 재생, 중지, 처음으로 버튼이 구성되도록 한다.
MATHML	MathML(Mathmatics Markup Language) 문서 정보를 지정한다.
CML	CML(Chemical Markup Language) 문서정보를 지정한다.
SVG	(Standard Vector Graphic) 벡터 기반 그래픽 정보를 지정한다.

본 정의에서 MathML과 CML, SVG를 도입한 이유는 본 XML DTD 단독 구조만으로는 다양한 교과 과목의 문항 콘텐츠를 표현할 수 없기 때문이다. MathML과 CML을 도입할 경우 수학 및 이공계 계열의 과목들을 표현할 수 있어 DTD의 확장성이 커질 수 있다.

3.2. 컴퓨터 기반 정보 구조

컴퓨터 기반 정보 구조는 세부적으로 레이아웃 관련 요소와 각 요소에 적용되는 속성 정보로 분류될 수 있다. 레이아웃은 일반적인 텍스트 정보의 형태를 결정짓는 것으로서, 글꼴 정보는 XML 정보가 별도 가공되는 경우 동일적으로 적용될 수 있도록 요소 정보 항목에는 포함시키지 않았다.

표 5. 텍스트 레이아웃을 위한 요소

요소명	기능설명
B	굵은 글씨체
I	기울임 글씨체
U	밑줄 글씨체
C	글자의 색상을 결정한다.
S	가운데 줄이 그어진 문자. (예사)

레이아웃 정보는 다른 태그들을 포함할 수 있으나, XML 노드 구조에서 leaf 위치에 놓이는 정보들은 attribution으로 정의하여 문항 저작자들이 손쉽게 확인할 수 있도록 구성하였다. 각 태그의 정의는 표6과 같다.

표 6. 속성값

이름	기능	관련요소
Src	멀티미디어 정보인 경우 관련 Resource 를 지칭한다.	IMG, ANI, MATHML, CML, SVG
Speaker	대화체인 경우 해당 문장의 화자(話者)를 의미한다.	DIAG
Lang	속성이 속한 요소의 언어정보	(모든 요소에 해당)
Iscorect	해당된 요소가 정답인지 아닌지를 표현한다.	CASE
Color	색상정보를 16진 코드에 의해 표현한다.	C
Level	문항의 난이도	ITEM
Comment	문항 저작자의 추가 설명 혹은 시스템 구성을 위한 추가적인 정보	(모든 요소에 해당)
Seq	객관식인 경우 각 항목의 순차적인 번호. 하나의 문항 내부에서는 unique value를 가진다.	ITEM, CASE
Id	하나의 문항이 가지는 unique값. 다른 문항과 차별되는 독립번호가 된다.	ITEM
pattern	만약 어떤 문항의 형식을 빌려올 경우 다른 문항의 형식 번호를 저장한다.	ITEM
title	문항에 대한 제목을 결정할 경우 저장된다.	ITEM
mode	Feedback이 정답일 때 제시되는지, 오답일 때 제시되는지를 결정한다.	FEEDBACK
sndsrc	음성정보가 연결되는 경우 음성정보의 URI	(모든 요소에 해당)

이러한 각 요소 정의를 바탕으로 XML DTD를 정의할 경우 다음과 같이 표현될 수 있다.

```

<ROOT title="General items collection">
  <ITEM pattern="a01" no="1" level="1">
    <QUESTION>Laquelle est la phrase simple?
    <desc lang="fr">Choisissez une bonne réponse.</desc>
  </QUESTION>
  <answer>
    <CASE seq="1" iscorrect=" true">La jeune fille blonde près
de la cheminée est la nièce de ma marraine.</CASE>
    <CASE seq="2">On se promène sur le quai de la Seine en
chantant haut.</CASE>
    <CASE seq="3">Il va pleuvoir car le ciel se couvre.</CASE>
    <CASE seq="4">Le professeur entre dans la classe et les
élèves s'arrêtent de se parler.</CASE>
  </answer>
</ITEM>
  <ITEM pattern="a01" no="2" level="3">
    <QUESTION>
    <desc lang="fr">Choisissez une bonne réponse.</desc>
    Cherchez la phrase dont le sujet est
souligné.</QUESTION>
  <answer>
    <CASE seq="1">La nuit, il n'y a que des ténébres.</CASE>
    <CASE seq="2">Il fait mauvais aujourd'hui.</CASE>
    <CASE seq="3">Quel est ton nom?</CASE>
    <CASE seq="4" iscorrect=" true">Il se passe quelque chose
à l'école.</CASE>
  </answer>
</ITEM>
</ROOT>

```

하나의 ROOT 태그는 다수의 ITEM 태그를 포함할 수 있으며, 각 ITEM 태그는 하나씩의 QUESTION 태그와 ANSWER 태그를 포함한다. Question과 Answer 요소는 각기 앞 절에서 소개한 다양한 구성요소들을 포함하여 동적인 문항 콘텐츠를 손쉽게 구현할 수 있다.

4. 결론 및 향후 연구

본 연구에서는 CBT 환경을 구축하기 위해 문항 정보를 XML 언어를 이용하여 표현하는 방법을 제시하였다. 비록 본 연구의 결과물이 매우 다양한 형태의 문항정보를 표현하기는 어려우나, 기존의 관계형 데이터베이스 기반 CBT 콘텐츠보다 다양한 정보를 표현할 수 있을 뿐만 아니라, 수학 및 화학 등 범용적인 분야에까지 시험 정보를 손쉽게 구축할 수 있는 장점이 있다. 또한 정보 자체에 의미 정보를 포함하고 있어, 자연언어 처리 등 다양한 분야에 요소 정보를 활용할 수 있다. 그러나 본 연구는 다음과 같은 부분에 있어 추가적인 연구가 필요하다.

첫째, 아직까지 문항을 저작하는 저자들은 대부분 비전산 언어 전문가 혹은 교수학습 전문가들이다. 따라서 이들이 손쉽게 XML 문항 정보를 구축

하기 위해서는 인터넷 기반 저작 인터페이스가 제공되어야 한다. 둘째, 문항 정보 구조는 구성되어 있으나, 평가용 정보로 구성되기 위해서는 자연언어 처리 및 XML 정보처리 기법이 도입된 평가지 구성 시스템이 구축되어야 할 것이다. 셋째, 본 DTD를 이용할 경우 다양한 문항 정보를 표현할 수 있으나, 도표 및 복잡한 다이어그램의 경우 SVG 및 그래픽 정보의 도움을 얻어야 한다. 향후 보다 다양한 문항 콘텐츠 타입을 지원하기 위해서는 HTML 기반 태그세트에 대한 보다 폭넓은 지원이 요구된다. 넷째, 본 DTD는 언어 정의만 되어 있기 때문에, 실제 평가 환경에서 활용하기 위해서는 XSL(eXtensible Style Sheet) 환경 구축에 대한 연구가 요구된다.

[10] MathML society, www.zvon.org

[11] CML society, www.xml-cml.org

5. Acknowledgements

본 연구는 과학재단 2002 지역우수과학자 연구 지원사업의 지원에 의해 이루어졌음을 밝힙니다.
(과제번호: R05-2002-000-00973-0)

참고 문헌

- [1] Stillings, Neil A., *Cognitive Science an introduction – Second edition*, MIT press, 1995, pp.339-342.
- [2] 윤애선, 정희영, 컴퓨터기반 언어 테스트 시스템 구축을 위한 요소 분석 및 그 적용에 관한 연구, 2001 정보통신부 정보통신학술연구과제 연구보고서, 2002.2.
- [3] 부산대학교 전자계산소, *CyFL system* 구축 최종 보고서, 2001.
- [4] 윤애선, 권혁철, 웹기반 원격 언어교육 시스템의 개발과 사용: 언어학, 전산학, 교육공학의 만남, '99 한국 인지과학회 춘계 학술대회 발표 논문집, 1999, pp. 178~185.
- [5] Pardi, William J., *XML in Action – Web technology*, Microsoft Press, 1999.
- [6] Laurant, St., *XML primer – Second edition*, MIS press.
- [7] 나일주, 웹기반 교육, 2001, 교육과학사
- [8] 김정렬, 웹기반 영어교육, 1999, 한국문화사
- [9] WWW consortium, www.w3c.org