

# XTM 문서의 가상문서로의 변환 시스템 설계 및 구현

류금호<sup>0</sup> 김운 김태현 맹성현  
충남대학교 컴퓨터학과

(khryu, wkim)@enya.cnu.ac.kr, (heemang, shmyaeng)@cs.cnu.ac.kr

## Design and Implementation of a System for Converting a XTM Document to a Virtual Document

Kung-Ho Ryu<sup>0</sup>, Woon Kim, Tae-Hyun Kim, Sung-Hyon Myaeng  
Dept. of Computer Science, Chungnam National University

### 요 약

인터넷상의 정보과다로 인하여 사용자는 필요한 정보를 찾는 데 많은 시간을 소모하게 되며 사용자 자신의 정보를 관리하는데 있어서도 어려움을 겪을 수 있다. 이러한 단점을 극복하려는 시도로서 XTM이라는 새로운 문서 표준이[1] 제안되었다. 본 논문에서는 이 XTM 문서와 충남대학교에서 개발한 가상문서가[3] 링크 정보를 통해 문서를 표현하는 공통점을 발견시켜 XTM 문서를 가상문서로 자동 변환하는 시스템을 설계 및 구현하였다. 본 시스템을 통해 가상문서의 처리 능력을 강화함으로써 가상문서 기반 디지털도서관 시스템의 확장을 도모하며 효율적인 정보 구축이 가능하도록 하였다.

### 1. 서 론

1969년 단 4대의 컴퓨터를 연결한 ARPANET에서 시작된 인터넷은 이제 우리 삶의 일부로서 당연시되고 있는 현실이다. 사용자들은 인터넷의 급속한 발전과 저변 확대로 전 세계의 수많은 정보를 실시간으로 손쉽게 접할 수 있게 되었다. 그러나 이러한 정보의 과다는 사용자에게 필요한 정보를 얻는데 많은 시간과 인내력을 요구하며, 사용자 자신의 정보를 관리하는데 있어서도 어려움을 겪게 하는 요인이 된다.

XML에 기반한 TopicMap 문서 표준안[ISO13250]이 제안된 목적은 이런 단점을 극복하여 사용자에게 손쉬운 정보 발견 및 사용자의 정보관리 능력의 향상을 도모하고자 함에 있다.[2] 이 XTM 문서의 구성은 링크정보를 이용한 topic과 topic들간의 관계를 나타내는 association을 통해 표현된다.[1][2] 그런데 이렇게 링크정보를 사용하여 문서를 표현하는 방식은 충남대학교에서 고안한 가상문서라는 개념을 통해 이미 정의되어 있으며, 가상문서 기반 디지털 도서관 시스템을 구축하는 등의 지속적인 연구 수행이 이루어지고 있다.[3] 즉 가상문서의 개념은 실제 내용은 존재하지 않고 기존에 존재하는 콘텐츠로의 Link만으로 구성된 문서를 뜻하며 이는 XTM 문서의 기본 구성단위인 topic이 실제 리소스의 링크를 통해 구성되는 것과 동일한 의미를 가질 수 있다.

본 논문의 주안점은 XTM 문서를 가상문서로 변환하는 시스템을 설계하고 구현하는 것인데, 그 목적은 다음과 같다.

#### (1) 가상문서 시스템의 확장

기존의 가상문서 시스템은 여러 문서형식(텍스트, HTML, XML 등)과 멀티미디어 데이터(오디오, 이미지, 동영상 등)를 처리할 수 있다. 본 논문의 변환 시스템을 통해 웹 상의 XTM 문서를 가상문서로 변환하는 기능을 제공함으로써 가상문서 기반 시스템에 확장성을 부여한다.

\* 본 논문은 과학기술부/한국과학재단 지정 소프트웨어연구센터의 지원을 받았음

#### (2) 가상문서 기반 시스템을 위한 효율적인 정보 구축

본 변환시스템을 통해 XTM의 정제되고 구조화된 정보를 가상문서 기반 디지털 도서관 시스템에 적용함으로써 효율적인 정보 구축을 가능하게 한다.

본 논문의 2장에서는 관련연구로서 가상문서와 XTM 문서에 대하여 간략히 설명하며 3장에서는 변환 시스템의 설계 및 변환과정에 대하여 기술한다. 4장에서는 결론 및 향후 연구방향에 대하여 설명한다.

### 2. 관련연구

#### 2.1 가상문서

실제 문서의 링크만으로 이루어진 가상문서는 새로운 콘텐츠를 만드는데 드는 시간과 저장장소 비용을 낭비하지 않고 기존 분산환경의 콘텐츠를 재사용하는 장점이 있으며 크게 내포링크, 참조링크, 메타데이터로 구성되어 있다. 내포링크란 가상문서 재현시 목적 콘텐츠가 직접 문서 안에 삽입되어 나타나길 원할 때 사용하는 링크이며, 참조링크는 문서 안에 삽입되지 않고 앵커로만 남아 사용자 선택 시에 항해나 브라우징할 수 있도록 표시하는 링크이다. 또한 메타데이터는 가상문서와 관련된 요약정보(저자, 출판사, 날짜 등)를 표현하기 위해 사용된다.[4][5]

#### 2.2 XTM 문서

어떠한 주제(subject)를 컴퓨터 시스템상의 표현 가능한 대상으로 구체화한 것을 토픽(topic)이라 하며, 이러한 토픽들간의 관계를 정의하는데 사용되는 정보 자원들의 구조를 표현하기 위해 제안된 모델을 TopicMap[ISO/IEC 13250:2000]이라 한다. XTM 문서는 TopicMap을 XML에 기반하여 작성한 문서를 의미하며, 문서의 구성은 크게 Topic과 토픽들간의 관계를 정의하는 Association 그리고 토픽과 관련된 정보를 표현하는 Occurrence의 세 부분으로 이루어진다. [2]

XTM 문서는 다음과 같은 응용에 유용하게 쓰일 수 있다.[1]

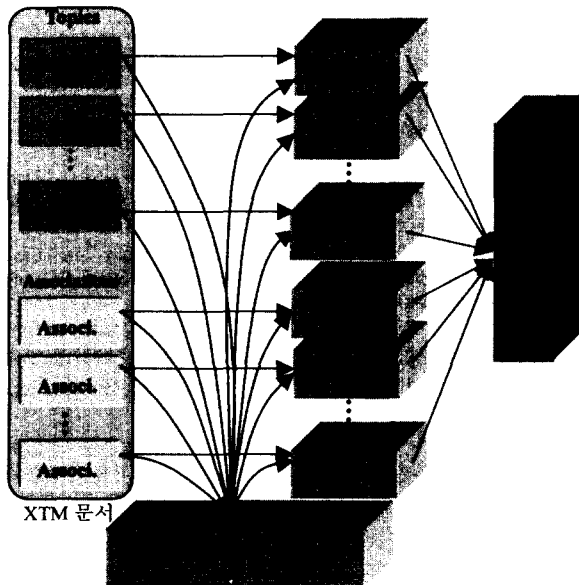
- (1) topic들이 인덱스, 상호참조, 인용 시스템들과 같이 네비게이션 툴로서 사용이 가능하도록 정보 객체들내의 데이터와 콘텐츠를 제공할 수 있다.
- (2) 특정 사용자에게 또는 특정 목적으로 정보 집합을 여과할 수 있다. 이런 여과를 통해 다중언어 문서의 관리나 보안기준에 의거한 접속모드의 관리를 도울 수 있다.
- (3) 비구조화된 정보 객체들을 구조화하는 역할을 제공한다.

3. 변환기 설계

3.1 시스템 설계시의 문제점 및 해결방안

변환기 설계시 XTM 문서가 가상문서로 직접 1:1 매핑되어 변환되지 못하는 문제점이 있다. 또한 변환시 XTM 문서의 일부 의미 표현을 현재의 가상문서 DTD로는 표현할 수 없는 문제가 발생한다. 본 절에서는 이러한 문제점들을 자세히 알아보고 그에 대한 해결방안을 제시한다.

- (1) 하나의 XTM 문서가 가지는 모든 의미를 단일 가상문서만으로 표현하는 것은 불가능하다. [그림 1]에서 볼 수 있듯이 하나의 XTM 문서에는 여러 개의 topic과 association이 존재하며 하나의 topic은 하나의 주제(subject)를 구체화한다. 만일 XTM 문서를 곧바로 가상문서로 매핑하였을 경우 XTM 문서가 각각의 topic이나 association을 통하여 표현하고자 하는 의미와는 전혀 다른 뜻을 지닌 가상문서로 변환되는 문제가 있을 수 있다. 이러한 문제를 해결하기 위하여 [그림 1]에서 처럼 하나의 XTM 문서내의 각각의 topic과 association에 대하여 가상문서를 생성한 후, 각 가상문서의 내포링크와 참조링크만을 취합한 새로운 가상문서를 만들어내므로 XTM 문서와 가상문서가 1:1로 대응되는 효과를 얻을 수 있게 하였다.



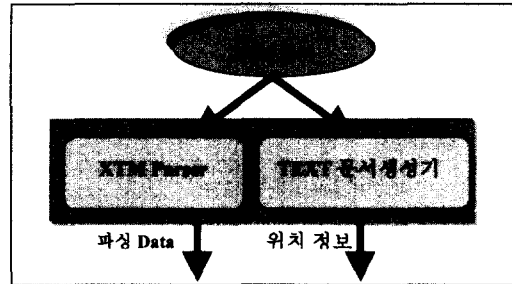
[그림 1] XTM 문서의 가상문서로의 변환과정

- (2) XTM 문서에서 topic은 항상 ID를 갖는다. 그러나 기존의 가상문서 DTD에서는 이를 표현하기 위한 문법적 요소가 정의되어 있지 않다. 따라서, 가상문서 DTD의 ELink, RLink, Source, Destination 엘리먼트 각각에 id라는 속성값을 추가함으로써 id 표현의 문제를 해결한다.
- (3) 가상문서 DTD의 참조링크(RLink)에는 source와 destination 관계가 정의되어 있다. 그러나, XTM 문서에서는 엘리먼트 관계가 단순히 의미상으로만 그룹지어져 있음으로써 가상문서로 변환시 source를 필요로 하지 않는 경우가 있다. 이 와 반대로 XTM 문서내에 실제 content가 표현되는 엘리먼트의 경우 변환시 destination이 필요없다. 따라서 가상문서의 해당 DTD인 'source'를 'source?'로 바꾸며 'destination+'을 'destination\*'로 수정하여 문제를 해결한다.

3.2 변환기 설계 및 변환 과정

변환기는 크게 문서 정보생성기와 문서 자동 변환기로 분류할 수 있다. 문서 정보생성기에서는 XTM 문서를 입력받아 문서내의 topic과 association을 분리하는 전처리과정을 거친 후 문서의 구조정보와 위치정보를 생성하여 문서 자동변환기에 전달하는 역할을 하며, 문서 자동변환기에서는 입력받은 구조정보와 위치정보를 이용하여 가상문서들로 변환한 후 이를 통합하는 최종 가상문서를 생성한다. 각각의 세부 설명은 아래 절들에 기술한다.

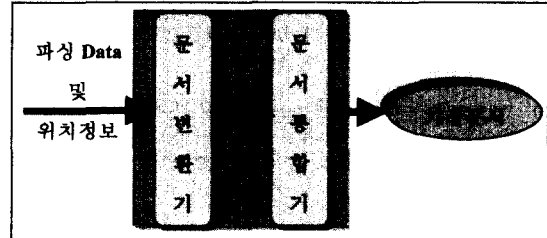
3.2.1 문서 정보생성기



[그림 2] 문서정보 생성기 구조

- XTM 파서 : XTM 문서를 입력받은 후 우선 topic과 association으로 분리한다. 그리고, topic과 association별로 각각의 엘리먼트를 파싱하여 문서변환을 위한 구조 정보를 문서변환기에 전달한다.
- TEXT 문서 생성기 : 가상문서의 특성상 실제 존재하는 물리적 문서를 대상으로 동적으로 연결하여 가상문서가 생성되기 때문에 임의의 물리적 문서를 만들어 줄 필요가 있다. 따라서 이에 해당하는 TEXT 문서를 생성한 후, topic과 association별로 텍스트의 위치정보를 추출하여 문서변환기에 전달한다.

3.2.2 문서 자동변환기



[그림 3] 문서자동 변환 및 통합기 구조

• 문서 변환기

: XTM 파서를 통해 받은 구조정보와 TEXT 문서 생성기를 통해 전달받은 위치정보를 이용하여 topic당 또는 association당 하나씩의 가상문서를 생성한다.

• 문서 통합기

: topic과 associatoin별로 생성된 각각의 가상문서를 내포링크(Elink)와 참조링크(RLink)로 구분하여 통합함으로써 하나의 새로운 가상문서를 생성한다.

3.3 변환 과정

다음은 TopicMap 문서의 가상문서로의 실제적인 변환 과정을 간단한 예제를 통해 살펴보기로 한다.

- (1) [그림 4]는 간단한 XTM 문서를 나타낸 것이다. 변환 시스템은 XTM 문서를 입력받은 후 전처리과정으로서 topic과 association별로 분류하여 이 분류정보를 TEXT 문서 생성기와 XTM 파서에 각각 전달한다.

```
<topicMap>
  <topic id="t36">
    <instanceOf>
      <resourceRef xlink:href="http://www.topicmaps.org/examples/hamlet.html" />
    </instanceOf>
    <baseName>
      <baseNameString>Hamlet</baseNameString>
    </baseName>
  </topic>
  <association>
    <member>
      <roleSpec><topicRef xlink:href="#t-play"/></roleSpec>
      <topicRef xlink:href="#t-hamlet" />
    </member>
    <member>
      <roleSpec><topicRef xlink:href="#t-character"/></roleSpec>
      <topicRef xlink:href="#t36" />
    </member>
  </association>
</topicMap>
```

[그림 4] XTM 문서 예제

- (2) TEXT 문서생성기는 전처리과정을 거친 분류정보를 입력으로 받아 TEXT 문서를 생성하고 분류정보를 저장한다. 그 후 각 topic과 association에 대한 위치정보를 추출하여 문서변환기에 전달한다. 예를 들어 TEXT 문서생성기에서 생성한 TEXT 문서에서 [그림4]에 있는 topic의 시작위치가 11이고 끝나는 위치가 85라고 한다면 TEXT 문서생성기는 "http:// enya.cnu.ac.kr/location/example001.txt#text(11, 85)"와 같은 위치정보를 생성하여 문서변환기에 전달한다.
- (3) XTM 파서는 분류 정보를 입력으로 받아 topic과 association 별로 파싱하여 구조정보를 문서변환기에 전달한다.
- (4) (2)와 (3)의 과정을 통해 얻은 XTM 문서의 구조정보와 TEXT 문서의 위치정보를 가지고 문서변환기를 통해 topic이나 association을 가상문서로 변환한다.
- (5) 변환된 각각의 topic과 association에 대한 가상문서를 문서통합기에 전달하여 최종적인 가상문서를 얻는다  
[그림 5]에서 볼 수 있듯이 최종 변환된 가상문서는 topic과 association의 내포링크를 우선 통합한 후, 이들이 가지는 참조 링크들을 통합하여 만들어진다.

```
<VdocHub>
  <ElinkSeq>
    <Elink id="t36" href="enya.cnu.ac.kr/location/example001.txt#text(11, 85)" role="topic" title="Hamlet" date="08/15/2001"/>
    <Elink href="enya.cnu.ac.kr/location/example001.txt#text(131, 194)" role="association" date="08/15/2001"/>
  </ElinkSeq>
  <RlinkSet>
    <RLink>
      <Source Is-generic="NO" href="enya.cnu.ac.kr/location/example001.txt#text(92, 118)" role="instanceOf/instance"/>
      <Destination href="http://www.topicmaps.org/examples/hamlet.html" role="instanceOf/class" category="resourceRef"/>
    </RLink>
    <RLink>
      <destination href="#t-play" role="member/rolespec/role" category="topicRef"/>
      <destination href="#t-hamlet" role="member/associated/topic" category="topicRef"/>
      <destination href="#t-character" role="member/rolespec/role" category="topicRef"/>
      <destination href="#t36" role="member/associated/topic" category="topicRef"/>
    </RLink>
  </RlinkSet>
</VdocHub>
```

[그림 5] 최종 변환된 가상문서

4. 결론 및 향후과제

본 논문은 웹 문서의 표준인 XTM 문서를 가상문서로 변환하는 시스템에 대한 방법론을 기술한다. 본 변환시스템을 사용하여 가상문서의 처리능력을 강화함으로써 가상문서 기반 디지털 도서관 시스템의 확장성을 향상할 수 있으며 가상문서 기반 디지털 도서관 시스템의 효율적인 정보 구축을 가능하게 할 수 있다.

앞으로는 변환된 가상문서를 효과적으로 브라우징하기 위한 스타일시트의 적절한 활용 문제에 대한 연구가 필요하며, 현 변환시스템에서 XTM 문서의 가상문서로의 변환범위를 가능한 확장할 수 있는 방법론이 필요하다.

참고문헌

[1] ISO/IEC JTC 1/SC34 N0056, <http://www.orml.gov.sg/ml/sc34/document/0058.htm>  
 [2] <http://www.topicmaps.org>  
 [3] 이용배, 이상봉, 맹성현, 김용, 류성섭, 이상홍(2001). "가상 문서 개념을 이용한 교육용 지식서버의 설계 및 구현", 제 11회 통신 정보 합동 학술 대회  
 [4] 맹성현의 8인, "A Digital Library System for Easy Creation/Manipulation of New Documents over Existing Resources", RIAO 2000  
 [5] 맹성현, 강지훈, 이만호, (1999) "Virtual Documents : a New Architecture for Knowledge Management in Digital Libraries", 제 2 회 아시아 디지털 도서관 컨퍼런스 Taipei, Taiwan, 11월, 1999.  
 [6] 이용배, 맹성현, 강지훈(1999), "SGML 문서를 XML 문서로 변환하는 자동변환기 설계 및 구현", 한국정보과학회 춘계학술대회 논문집(B), pp.318-320, 1999년 4월