

# 사용자 프로필 및 어노테이션을 이용한 XML 문서 트랜스코딩

정쌍용<sup>o</sup>, 손원성, 이진상, 임순범\*, 최윤철  
연세대학교 컴퓨터과학과  
숙명여자대학교 멀티미디어학과  
{gahul<sup>o</sup>, sohnws, gr20000, ycchoy}@rainbow.yonsei.ac.kr  
sblim@sookmyung.ac.kr\*

## XML Document Transcoding using User Profile and Annotation

Ssang Yong Jung<sup>o</sup>, Won-Sung Sohn, Jinsang Lee, Soon-Bum Lim, Yoon-Chul Choy  
Dept. of Computer Science, Yonsei University  
Dept. of Multimedia, Sookmyung Women's University

### 요 약

트랜스코딩이란 기존 유선 환경에서 제공되는 웹 콘텐츠를 특정 환경에 적합한 형태로 변환 하는 것을 의미한다. 그러나 이와 관련된 기존 연구에서는 사용자가 요구하는 사항만을 변환 하거나 서비스 제공자가 일방적으로 변환하여 웹 콘텐츠를 제공하고 있다. 따라서 이슈변화에 따른 사용자의 대처능력이 떨어지기 때문에 사용자의 사용성이 저하되며, 사용자에게 무의미한 정보 제공의 가능성이 있다. 이러한 문제점들을 해결하기 위해 본 논문에서는 사용자 프로필에 의한 요구사항과 서비스 제공자의 의견을 함께 제공 할 수 있는 변환 기법을 제안 하고, 특히 멀티미디어 뉴스 제작을 위한 표준인 NewsML을 대상으로 적용하였다. 사용자 프로필에 의한 요구 사항은 XML 문서의 구조정보를 이용하여 자동으로 추출하고, 서비스 제공자의 의견은 문서의 레이아웃(Layout) 정보를 가지고, 어노테이션(Annotation) 기법을 활용하여 수동으로 추출한다. 그 결과, 사용자 관점에 부합하는 변환이 이루어지고, 다양한 이슈변화에 대한 대처능력이 향상되어 사용자의 사용성이 증대 되었다.

## 1. 서 론

무선으로 웹 콘텐츠에 접속 가능한 PDA, Hand-held computer, 스마트 폰 등과 같은 무선 단말기의 증가[1][2]로 인해 장소와 시간의 제약 없이 웹 콘텐츠를 사용하고자 하는 요구가 증가 하고 있다. 하지만 기존 웹 콘텐츠에 개인용 단말기를 이용하여 접속 할 경우 화면의 크기, 메모리 크기 그리고 제한된 대역폭 때문에 여러 가지 문제점이 발생한다. 따라서 이러한 문제점들을 해결하기 위해, Desktop 환경에 맞게 디자인 된 웹 콘텐츠를 개인용 단말기에 맞게 변환 시켜주는 트랜스코딩에 대한 연구가 활발히 이루어 지고 있다[3].

그러나 기존 연구에서는 사용자가 요구하는 사항만을 변환 하거나[4] 서비스 제공자가 일방적으로 변환하여 웹 콘텐츠를 제공하고 있어[5], 이슈변화에 따른 사용자의 대처능력이 떨어지고 사용자의 사용성이 저하되며, 사용자에게 무의미한 정보 제공의 가능성이 있다. 따라서 이러한 문제점들을 해결하기 위해 본 논문에서는 사용자 profile에 의한 요구사항과 서비스 제공자의 의견을 함께 제공 할 수 있는 변환 기법을 제안 하고, 멀티미디어 뉴스 제작을 위한 표준인 NewsML(News Markup Language)[6]을 대상으로 적용하였다.

사용자 profile에 의한 요구 사항은 XML[7] 문서의 구조정보를 이용하여 자동으로 추출하고, 서비스 제공자의 의견은 문서의 레이아웃 정보를 가지고, 어노테이션 [8] 기법을 활용하여 수동으로 추출한다. 그 결과, 사용자 관점에 부합하는 변환이 이루어지고, 다양한 이슈변화에 대한 대처능력이 향상되어 사용자의 사용성이 증대 되었다.

본 논문의 구성은 1장 서론에 이어 2장에서는 트랜스코딩과 NewsML에 관한 기존 연구에 대해 알아보고 3장에서는 본 논

문에서 제시하는 트랜스코딩 기법에 대해 설명한다. 마지막 4 장에서는 전체 논문에 대한 요약 및 향후 연구 방향에 대해 논 하도록 한다.

## 2. 관련연구

웹 콘텐츠를 변환하기 위한 방법에는 여러 가지가 존재한다. 2.1 장에서는 웹 문서 변환기법에 대해 알아보고 2.2 장에서는 본 논문의 적용 대상인 NewsML에 대해 알아 보겠다.

### 2.1 트랜스코딩

트랜스코딩의 분류에는 크게 트랜스코딩 시점에 의한 분류, 처리형태에 의한 분류, 사용자 의견 반영정도에 의한 분류로 나누어 볼 수 있다.

시점에 의한 분류에는 서버측 변환[9], 프락시측 변환[4][10], 클라이언트 측 변환[11]이 있다. 서버측에서 변환이 이루어 질 경우 정확한 변환이 이루어 지는 장점이 있지만, 제한된 정보만이 제공되는 단점이 있다. 클라이언트에서 변환이 이루어 지면, 사용자 개인별 특성에 맞는 변환이 이루어 지지만, 단말의 성능상 제한사항으로 인해 시간과 자원의 효율적인 이용이 어렵게 된다. 프락시에서 변환이 이루어 지면 반드시 프락시를 거쳐야 한다는 점이 있지만 여러 가지 웹 콘텐츠를 다양한 단말기에 제공할 수 있다. 따라서 현재 대부분의 변환 시스템은 프락시에 기반하고 있다.

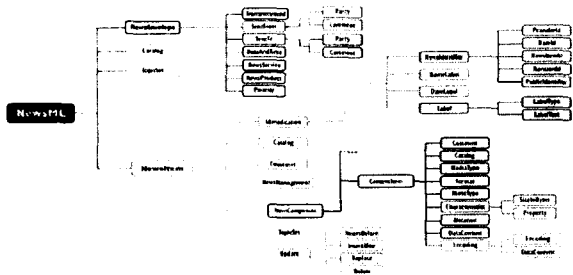
처리 형태에 따른 분류는 자동 변환[8][9]과 수동 변환으로 나눌 수 있다. 시스템에 따라 자동 변환과 수동 변환을 이용할 수 있다. 자동 변환과 수동 변환을 같이 사용 할 수도 있다.

사용자의 의견 반영정도에 따른 분류로는 프락시에서 사용자

profile을 가지고 사용자가 원하는 정보만 단말에 보내어 주는 경우[4]와 프락시에서 전체 내용을 전부 변환[5]하여 보내어 주는 경우가 있다. 전자의 경우 사용자에 따른 특화된 정보 제공이 가능하지만 이슈 변화에 따른 대처 능력이 떨어져 사용자의 사용성이 저하될 수 있다. 후자의 경우에는 사용자 관점에 부합하지 않는 무의미한 정보가 제공될 가능성이 크다. 따라서 본 논문에서는 이러한 두 가지 경우의 문제점을 해결 할 수 있는 변환 기법을 제안 한다.

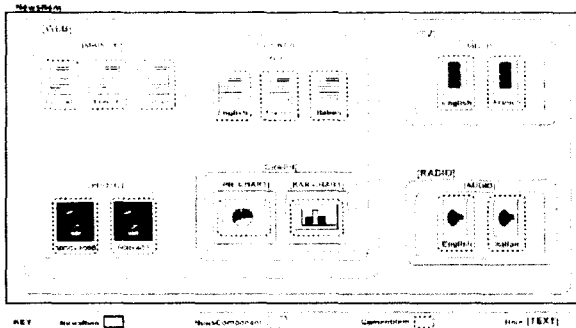
### 2.2 NewsML

NewsML은 IPTC(International Press Telecommunications Council)에 의해 정의된 XML기반의 멀티미디어 뉴스를 제작, 저장 및 전송하기 위한 표준이다. IPTC에서 발표한 NewsML의 스펙은 DTD와 XML 스키마 형식으로 발표 되었다. NewsML의 구조는 다음 [그림 1]과 같다.



[그림 1] NewsML의 구조[12]

NewsML의 핵심은 NewsItem이란 개념이고 NewsItem은 텍스트, 동영상, 이미지 등 다양한 미디어를 포함 할 수 있다.

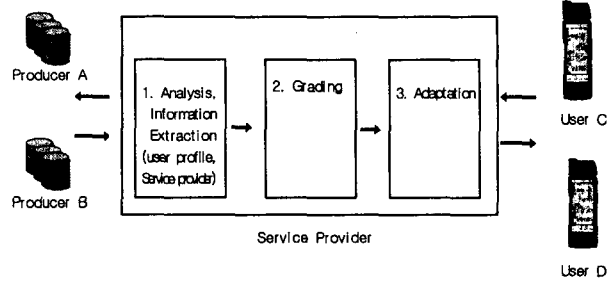


[그림 2] NewsItem의 구성 예[12]

### 3. 사용자 프로파일 및 어노테이션을 이용한 트랜스코딩 기법

본 논문에서 제안하는 시스템은 프락시기반의 변환 시스템으로써, 사용자에게 제공되는 내용은 프락시에 미리 등록된 사용자 프로파일에 의해 NewsML의 구조 정보를 이용하여 자동 추출되는 것과 서비스 제공자가 어노테이션 기법을 활용하여 자동 추출하는 것으로 나눌 수 있다. 서비스제공자의 의견을 어노테이션 기법을 이용하여 추출하는 것은 스타일 정보로 표현되어 구조정보를 이용할 수 없기 때문이다. 서비스제공자의 의

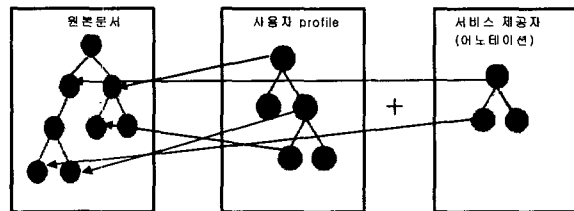
견을 추출하기 위해 신문의 중요 내용을 머리기사, 속보 및 중요기사로 구분하여 어노테이션을 모델링 한다. 이렇게 추출된 정보는 우선순위 부여 규칙에 의해 우선순위가 매겨지고, 우선순위에 따라서 단말에 맞게 변환규칙에 의해 변환된다.



[그림 3] 전체 변환 절차도

### 3.1 분석 및 정보 추출

[그림 3]의 1단계에서는 서비스 제작자(신문사)에 의해 제공된 서비스 제공자가 내용을 분석하여 사용자가 원하는 내용과 서비스 제공자의 의견을 추출 한다.



[그림 4] 분석 및 정보 추출

사용자의 요구사항은 NewsML의 구조정보를 분석하여 사용자 프로파일에 등록된 장르의 내용과 일치하는 부분을 자동으로 추출 하는 내부적인 방법을 사용한다. 예를들어, 사용자가 문화 분야를 등록했으면 NewsML의 장르를 분석하여 문화분야의 기사를 추출한다.

```
<DescriptiveMetadata Confidence="High" Importance="5">
  <Language FormalName="en"/>
  <Genre FormalName="Current"/>
  <SubjectCode>
    <Subject FormalName="1100000"/>
  </SubjectCode>
  <TopicOccurrence Topic="#person1" HowPresent="Prominent"/>
  <TopicOccurrence Topic="#person2" HowPresent="RelatesTo"/>
  <TopicOccurrence Topic="#location1" HowPresent="Passing"/>
</DescriptiveMetadata>
```

[그림 5] 사용자 프로파일을 이용한 정보추출[12]

뉴스기사는 중요도에 따라 스타일 정보를 이용하여 구분되기 때문에 서비스 제공자의 의견은 NewsML의 구조정보를 이용 할수 없다. 그러므로 서비스 제공자의 의견을 추출하기 위해서는 스타일 정보를 이용한다. 그러기 위해, 어노테이션을 머리기사, 속보, 중요

기사로 나누어 모델링하고, 모델링 된 어노테이션을 활용하여 정보를 추출하는, 외부적인 방법을 사용한다. 따라서 정보를 분석하고 추출하는데 내부적인 방법과 외부적인 방법을 같이 이용하는 혼합 기법을 이용한다.

### 3.2 우선순위 부여 규칙

1단계에서 정보를 추출하면 2단계에서는 추출된 정보의 변환을 위해 각 정보에 우선순위를 부여한다. 추출된 정보에 대한 우선 순위 부여규칙은 [표 1]과 같다.

[표 1] 우선순위 부여 규칙

전체 Grading Rules		
우선 순위	내 용	비고
1	머리기사	다음의 우선 순위는 Default로 하고 사용자가 우선순위를 변경할 수 있다.
2	속 보	
3	중요기사	
4	사용자 요청 기사	

속보 Grading Rules			
구분	새로운 기사	개정된 기사	기존의 기사
User가 원하는 항목	2-1	2-2	삭제
User가 원하지 않는 항목	2-3	2-4	삭제

중요기사 Grading Rules			
구분	새로운 기사	개정된 기사	기존의 기사
User가 원하는 항목	3-1	3-2	삭제
User가 원하지 않는 항목	3-3	3-4	삭제

사용자 요청기사 Grading Rules			
구분	새로운 기사	개정된 기사	기존의 기사
사용자 요청기사	4-1	4-2	삭제

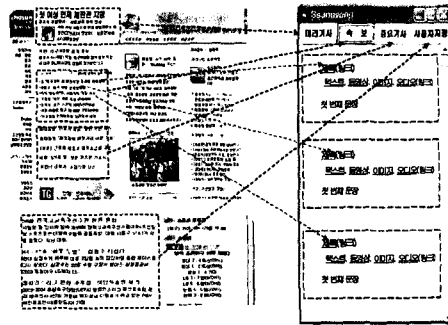
Grade가 같을 경우 가장 최근의 기사를 우선으로 한다.

### 3.3 변환 규칙

우선순위 부여규칙에 의해 순위가 정해지면, 다음 단계로 넘어가 아래와 같은 변환 규칙에 따라 레이아웃이 정해진다.

- \* 머리기사, 속보, 중요기사, 사용자지정 기사는 스크린 상단에 우선순위에 따라 왼쪽부터 콘텐츠를 탭으로 배치한다.
- \* 상단의 콘텐츠를 탭을 선택하면 해당 분야 기사가 스크린에 뜬다
- \* 위에서 아래로 우선순위에 따라 배열된다.
- \* 타이틀은 링크가 되고 동영상, 텍스트, 이미지, 오디오와 같은 미디어 타입을 선택하면 해당 미디어를 볼 수 있다.
- \* 타이틀 밑에는 해당기사의 첫 번째 문장이 뜬다.
- \* 좌, 우 스크롤은 없고 상, 하 스크롤은 존재한다.

기존의 변환에 관한 연구에서는 작은 단말기 화면에 많은 양의 정보를 표현하거나 과도하게 내용을 생략하고 링크로 대체하였다. 따라서 사용자가 시각적으로 부담감을 가지고 정보를 찾거나 네비게이션 하는데 많은 단계를 거쳐, 길을 잃을 가능성이 있었다. 또한 정보를 그룹핑 하는데 있어서 레이아웃 정보에 의존 하였기 때문에 서로 의미적 연관성이 없는 정보가 같은 그룹에 포함 될 가능성이 많았다. 그러나 본 논문에서 제시하는 기법은 단말기 화면에 맞추어 한번에 제한된 내용을 표현해 주고 사용자 중심의 변환이 이루어지기 때문에 사용자가 시각적으로 부담 없이 네비게이션 할 수 있을 뿐만 아니라 우선순위에 맞게 사용자가 원하는 내용을 찾아볼 수 있다. 또한 구조정보를 이용하여 사용자가 원하는 내용을 자동으로 추출하고 서비스 제공자가 제공하고 하는 내용은 수동으로 어노테이션 기법을 활용하여 변환하기 때문에 서로 의미적 연관성이 없는 정보가 같은 그룹 내에 포함될 가능성이 없다. 따라서 사용자가 혼란 없이 정보를 얻을 수 있다.



[그림 6] 변환 예

### 4. 결론 및 향후 연구

트랜스코딩에 관한 기존 연구에서는 사용자의 요구사항과 서비스 제공자의 의견을 같이 반영 할 수가 없어 이슈변화에 따른 사용자의 대처능력이 떨어지고 사용자의 사용성이 저하되며, 사용자에게 무의미한 정보 제공의 가능성이 있었다. 따라서 위의 문제를 해결하기 위해 본 논문에서는 NewsML을 대상으로 하여 사용자의 요구사항과 서비스 제공자의 의견을 함께 충족 시킬 수 있는 변환 기법을 제안 하였다. 사용자의 의견은 XML 문서의 구조정보를 이용하여 정보를 추출 하였고, 서비스 제공자의 의견은 스타일 정보에 기반을 둔 어노테이션 기법을 사용 하였다. 그 결과, 사용자 관점에 부합하는 변환이 이루어지고, 다양한 이슈변화에 대한 대처능력이 향상되어 사용자의 사용성이 증대 되었다.

향후 연구 과제로는 다양한 XML 문서에 적용 가능한 재사용 트랜스코딩 기법을 연구하는 것이다. XML 문서의 레이아웃과 논리 정보간의 관계를 정의하여 둘 간의 관계를 추출 및 정의 하면 한번 사용된 어노테이션을 다른 분야에서도 사용 가능할 것이다.

### 참 고 문 헌

- [1] E. A. Brewer, R. H. Katz, Y.Chawathe, et al. "A Network Architecture for Heterogeneous Mobile Computing", IEEE Personal Communications, Vol.5, No.5, pp.8-24, October 1998.
- [2] 배찬권, "정보통신산업동향 정보통신기기제 7절 PDA", 정보통신정책연구원, 2001.
- [3] T. Bickmore, W. Schilit, "Digester : Device-Independent Access to the World Wide Web", Computer Networks and ISDN Systems, Vol.29, No.8, pp.1075-1082, 1997.
- [4] Katashi Nagao, Kevin Squire, "Semantic Annotation and Transcoding : Making Web Content More Accessible", IEEE Multimedia Magazine, Vol.8, Issue 2, April-June
- [5] M. Hori, G.Kondoh, K. Ono, S Hirose and S. Singhal, "Annotation-Based Web Content Transcoding", 9<sup>th</sup> World Wide Web Conference, 2000.
- [6] NewsML, <http://www.newsml.org>
- [6] XML, <http://www.w3.org/XML>
- [7] IPTC, <http://www.iptc.org>
- [8] 정필모, 문헌정보학원론, 구미무역(출판사), pp254-257, 1996.
- [9] IBM, Websphere Transcoding Publisher, <http://www-3.ibm.com/software/webservers/transcoding/index.html>
- [10] T. Bickmore, A.Girgensohn and J.W. Sullivan, "Web Page Filtering and Re-Authoring for Mobile Users", The Computer Journal, Vol.42, No.6, pp.534-546, 1999.
- [11] N. Milic-Frayling and R. /sommerer, "SmartView : Flexible Viewing of Web Page Contents", World Wide Web Conference 2002, 2002
- [12] NewsML Version 1.0, Functional Specification, 24 October, 2000.