

동적 프로파일과 어노테이션을 이용한 XML 문서 트랜스코딩

정쌍용, 손원성, 이진상 임순범*, 최윤철
연세대학교 컴퓨터과학과
*숙명여자대학교 멀티미디어학과

XML Document Transcoding using Dynamic Profile and Annotation

Ssang Yong Jung, Won-Sung Sohn, Jinsang Lee, Soon-Bum Lim*, Yoon-Chul Choy*
Dept of Computer Science, Yonsei University
(E-mail:gahul, sohnws, gr20000, ycchoy)@rainbow.yonsei.ac.kr
*Dept of Multimedia, Sookmyung Women's University
E-mail:sblim@sookmyung.ac.kr

요 약

현재 유선에서 지원되는 웹 콘텐츠를 개인용 단말기에서 지원하기에는 단말기의 성능상 한계(screen size, memory size, bandwidth 등) 때문에 여러 가지 문제가 있다. 트랜스코딩이란 이러한 기존 환경에서 제공되는 웹 콘텐츠를 특정 환경에 적합한 형태로 변환 하는 것을 의미한다. 그러나 이와 관련된 기존 연구에서는 사용자가 요구하는 사항만을 변환 하거나 서비스 제공자가 일반적으로 변환하여 웹 콘텐츠를 제공하고 있어 이슈변화에 따른 사용자의 대처능력이 떨어지고 사용자의 사용성이 저하되며, 사용자에게 무의미한 정보 제공의 가능성이 있다. 이러한 문제점들을 해결하기 위해 본 논문에서는 멀티미디어 뉴스 제작을 위한 표준인 NewsML을 대상으로 사용자의 동적 프로파일과 서비스제공자의 어노테이션을 이용하여 사용자가 요구하는 기사와 서비스 제공자가 제공하는 기사를 같이 변환하는 기법을 제안 한다. 본 논문의 결과 갑자기 발생하는 사회적 이슈변화에 따른 사용자의 대처능력이 향상 되고 사용자가 불필요한 정보에 과다하게 노출되는 것을 막을 수 있다.

1. 서론

무선 용 개인 단말기기를 이용하여 기존 웹 콘텐츠에 접속할 경우 단말기의 제약(화면의 크기, 메모리 크기 그리고 제한된 대역폭 등)[1][2] 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 이루어지고 있다[3].

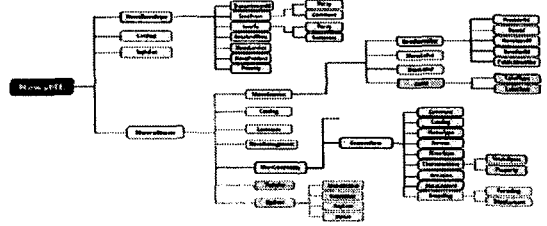
그러나 트랜스코딩에 관한 기존 연구에서는 기술적인 측면과 사용성 측면에서 문제가 있어 추가적인 연구가 필요하다. 기술적인 문제점으로는 주로 HTML의 스타일 정보를 이용하여 변환이 이루어지기 때문에 불필요한 정보가 포함되어 사용자에게 혼란을 주거나 상호 연관성이 없는 정보가 같은 카테고리에 포함 될 가능성이 있다. 그리고 현재 활발히 사용되고 있는 XML 과 같은 구조문서에 대해서는 많은 연구가 이루어지지 않고 있다. 또한 사용성 측면에서는 사용자가 요구하는 사항만을 변환 하거나[4] 서비스 제공자가 일반적으로 변환하여 웹 콘텐츠를 제공하고 있어[5], 이슈변화에 따른 사용자의

대처능력이 떨어지고 사용자의 사용성이 저하되며, 사용자에게 무의미한 정보 제공의 가능성이 있다. 위에서 언급한 기술적인 문제를 해결하기 위해 본 논문에서는 멀티미디어 뉴스를 제작, 전송 및 저장하기 위해 IPTC(International Press Telecommunications Council)에 의해 표준으로 지정된 XML 기반의 NewsML(News Markup Language)[6]을 대상으로 XML 문서의 구조정보를 효율적으로 이용하여 트랜스코딩하는 기법을 제안한다.

사용성 측면에서는 사용자가 요구하는 기사와 서비스 제공자가 제공하는 기사를 같이 변환하여 사용자에게 서비스 함으로써, 사용자가 갑자기 발생하는 사회적 이슈변화에 대처토록 하며 불필요한 정보에 과다하게 노출되는 것을 막을 수 있다. 또한 사용자가 이전에 접근했던 정보들을 바탕으로 사용자의 관심사를 추출하고 추출된 관심사를 사용자 프로파일에 자동으로 등록한다. 이 방법은 직접적인 피드백을 필요로 하지 않기 때문에 사용자의 의식적 작용에 따른 번거로움을 극복할 수 있고, 사용자의 관심사 변화도 반영할 수 있다. 그리고, 사

용자가 검색했던 내용 중에서 차후에 사용하기 위해 저장하고자 하는 내용은 부가적인 씨택트 정보들과 함께 저장할 수 있는 기법도 제공한다.

본 논문의 구성은 1장 서론에 이어 2장에서는 트랜스코딩과 NewsML에 관한 기존 연구에 대해 알아보고 3장에서는 본 논문에서 제시하는 트랜스코딩 기법에 대해 설명한다. 마지막 4장에서는 전체 논문에 대한 요약 및 향후 연구 방향에 대해 논하도록 한다.



[그림 1] NewsML의 구조[12]

2. 관련연구

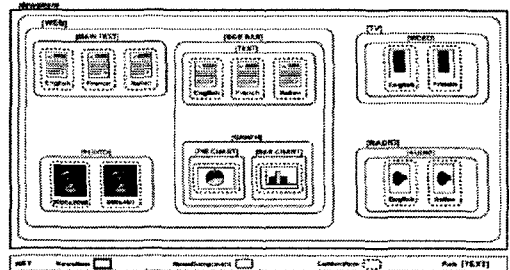
웹 콘텐츠를 변환하기 위한 방법에는 여러 가지가 존재한다. 2.1 장에서는 웹 문서 변환기법에 대해 알아보고 2.2 장에서는 본 논문의 적용 대상인 NewsML에 대해 알아 보겠다.

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

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 이루어지는 장점이 있지만, 제한된 정보만이 제공되는 단점이 있다. 클라이언트에서 변환이 이루어지면, 사용자 개인별 특성에 맞는 변환이 이루어 지지만, 단말기의 성능 상 제한사항으로 인해 시간과 자원의 효율적인 이용이 어렵게 된다. 프락시에서 변환이 이루어지면 반드시 프락시를 거쳐야 한다는 단점이 있지만 여러 가지 웹 콘텐츠들을 다양한 단말기에 제공할 수 있다. 따라서 현재 대부분의 변환 시스템은 프락시에 기반하고 있다.



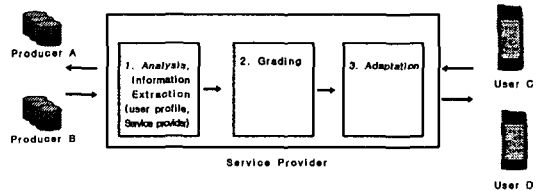
[그림 2] NewsItem의 구성 예[12]

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

사용자의 의견 반영정도에 따른 분류로는 프락시에서 사용자 프로파일을 가지고 사용자가 원하는 정보만 단말기에 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.

3. 시스템 구조 및 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 수동 추출하는 것으로 나눌 수 있다. 서비스 제공자의 의견을 어노테이션 기법을 이용하여 추출하는 것은 스타일 정보로 표현되어 구조정보를 이용할 수 없기 때문이다. 서비스제공자의 의견을 추출하기 위해 신문의 중요 내용을 머리기사, 속보 및 중요기사로 구분하여 어노테이션을 모델링 한다. 이렇게 추출된 정보는 우선순위 부여 규칙에 의해 우선순위가 매겨지고, 우선순위에 따라서 단말기에 맞게 변환규칙에 의해 변환된다.



[그림 3] 전체 변환 절차도

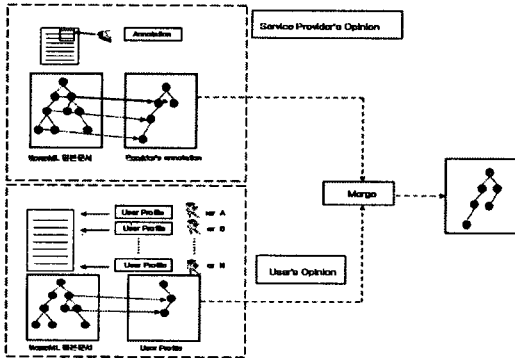
2.2 NewsML

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.

3.1 분석 및 정보 추출

[그림 3]의 1단계에서는 서비스 제작자(신문사)에 의해 제공된 기사를 서비스 제공자가 분석하여 사용자가 원하는 내용과

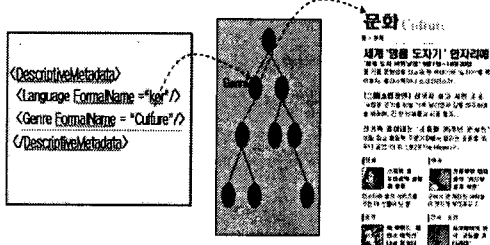
서비스 제공자의 의견을 추출 한다.



[그림 4] 분석 및 정보 추출

3.1.1 사용자 의견

서비스 제공자는 컨텐츠 제작자로부터 컨텐츠를 받으면 제공된 기사의 NewsML 구조를 분석하여 사용자 프로파일에 등록된 사용자 요구사항과 일치하는 내용을 추출한다. 아래 [그림 5]와 같이 사용자가 문화 분야를 사용자 프로파일에 등록한 경우, 서비스 제공자는 제공된 기사의 NewsML 구조를 분석하여 문화 분야에 해당하는 기사를 추출한다.



[그림 5] 사용자 프로파일을 이용한 정보추출[12]

(1) 동적 프로파일

기사 내용 중 프로파일에 등록된 내용이 아니더라도 사용자가 기사 검색 중 자주 방문하는 장르는 사용자의 프로파일에 동적으로 추가하여, 사용자가 차후 접속 시 사용성을 증대하고 사용자의 관심분야가 바뀔 때마다 매번 수작업으로 프로파일을 수정해야 하는 번거로움을 제거한다.

사용자의 프로파일을 동적으로 변환하기 위한 규칙은 아래와 같다.

[정의 1] 3일 동안 사용자의 방문회수(링크 히트 수)가 10회 이상이고 70% 이상 같은 장르를 방문하면 그 장르를 d_genre 로 생성한다.

$$\text{if } (N \geq 10) \text{ and } \left(\frac{n_s = (N-n_e)}{N} \times 100 \geq 70 \right)$$

$$d_genre = n_s$$

N : 3 일간 총 방문회수(총 링크 히트 수)

n_s : 3 일간 특정 장르 방문 회수

n_e : 3 일간 기타 장르 방문 회수

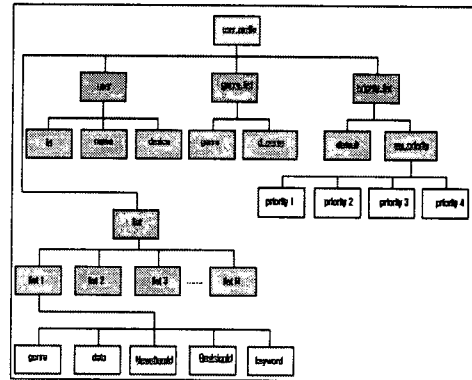
d_genre : 사용자 프로파일에 동적으로 추가되는 장르

[정의 2] 3일 동안 사용자의 방문 회수(총 링크 히트 수)가 10 회 미만이거나 특정 장르의 방문 회수가 70% 미만일 경우, d_genre는 생성되지 않는다.

$$\text{if } (N < 10) \text{ or } \left(\frac{n_s = (N-n_e)}{N} \times 100 < 70 \right)$$

$$d_genre = 0$$

[그림 6]에서 보듯이 사용자 프로파일에는 사용자가 원하는 장르, 사용자 ID 그리고 단말기 종류 등을 등록하고 우선순위를 시스템에서 선언된 내용으로 할 수도 있으며, 사용자가 원하는 순서로 수정 할 수도 있다. 그리고 위에 정의된 [정의 1,2]에 따라 프로파일을 동적으로 생성 할 수 있다.



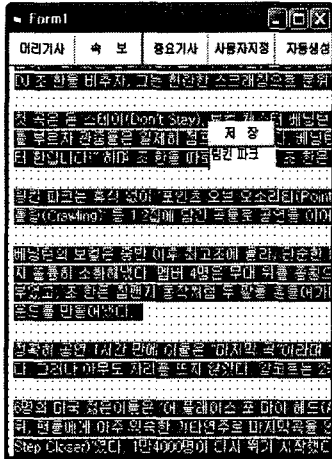
[그림 6] 사용자 동적 프로파일 DTD 구조

(2) 사용자 어노테이션

사용자가 방문한 기사의 내용 중 차후 검색이나, 편집 및 사용을 위해 저장하기 원하는 중요 내용은 여러 가지 씨맨틱 정보들과 함께 프로파일에 추가된다.

사용자가 저장하기 원하는 내용은 간단한 어노테이션 기법을 활용하여 사용자의 프로파일에 저장 된다. 씨맨틱 정보 중 장르, 날짜, NewsItemId, RevisionId는 NewsML 문서의 구조정보를 이용하여 시스템에서 자동으로 저장 되지만, keyword 정

보는 사용자가 원하는 내용으로 설정해 주어야 한다. 씨맨틱 정보들을 추가하는 이유는 사용자가 차후 이용 시 검색을 용이하게 하기 위해서 이다. [그림 7]은 사용자 어노테이션의 예를 나타낸다.



[그림 7] 사용자 어노테이션 예

3.1.2 서비스 제공자의 의견

서비스 제공자의 의견은 머리기사, 속보, 중요기사로 나누어 서비스 제공자가 프락시 서버에서 어노테이션을 이용하여 내용을 추출한다.



[그림 8] 서비스제공자 어노테이션 예

3.2 우선순위 부여

1단계에서 사용자가 원하는 내용과 서비스 제공자의 의견에 따라 정보를 추출하면, 2단계에서는 추출된 정보를 변환시키기 위해 우선순위를 규칙에 따라 부여한다.

3.3 변환

분석 및 정보추출 단계에서 사용자와 서비스 제공자의 의견에 따라 정보를 추출하고, 그 추출된 정보를 우선순위 부여규칙에 따라 우선순위를 부여한 후 마지막 단계인 변환 단계로

넘어간다. 기존 변환에 관한 연구에서는 작은 단말기 화면에 많은 양의 정보를 표현하거나 과도하게 내용을 생략하고 링크로 대체하여 정보를 찾을 때 시각적 부담을 주거나 검색 시 길을 잃을 가능성이 많았다. 그러나 본 논문의 변환은 사용자 중심의 변환이 이루어 지고 단말기 화면에 맞추어 한번에 제한된 내용을 표현해 주어 사용자가 시각적 부담감 없이 정보를 찾을 수 있다.

4. 결론 및 향후연구

트랜스코딩에 관한 기존 연구에서는 사용자의 요구사항과 서비스 제공자의 의견을 같이 반영 할 수가 없어 이슈변화에 따른 사용자의 대처능력이 떨어지고 사용자의 사용성이 저하되며, 사용자에게 무의미 한 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 사용자의 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.

참고문헌

- [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 정보통신기편 제7절 PDA", 정보통신정책연구원, 2001.
- [3] T. Bickmore, W. Schilit, "Digestor : 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, 2001.
- [5] M. Hori, G.Kondoh, K. Ono, S Hirose and S. Singhal, "Annotation-Based Web Content Transcoding", 9th World 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, www-3.ibm.com/software/webservers/transcoding/index.html
- [10] T. Bickmore, A.Gingensohn 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.