

모바일 환경에서 사용자 관심사항을 고려한 XML 문서 변환

송특섭[○] 이진상 손원성 고승규 최윤철 임순범*

[○]연세대학교 컴퓨터과학과,

*숙명여자대학교 멀티미디어학과

{teukseob[○], gr2000, sohnws, skko, ycchoy}@rainbow.yonsei.ac.kr

*sblim@sookmyung.ac.kr

User Centered XML Documents Transcoding for Mobile device

Teuk-Seob Song[○], Jin-Sang Lee, Won-Sung Sohn, Seung-Kyu Ko,
Yoon-Chul Choy, Soon-Bum Lim*

Dept. of Computer Science, Yonsei University

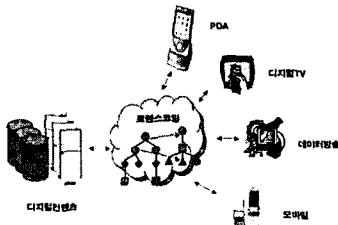
*Dept. of Multimedia, Sookmyung Women's University

요 약

기존의 웹 콘텐츠를 PDA나 핸드폰 등 모바일 디바이스에서 표현하기 위해서는 단말기의 낮은 성능상의 제약으로 인해 적절한 변환이 필요하다. 이와 같이 단말기의 성능을 고려하여 콘텐츠를 변환 하는 것을 트랜스 코딩이라 한다. 현재 까지 트랜스코딩에 대한 연구들은 사용자의 관심사항을 반영하기 보다는 단말기의 성능을 고려한 연구가 주로 이루어지고 있다. 이는 서비스 제공자 중심의 콘텐츠 변환으로 사용자의 관심사항을 반영하기 곤란하다. 현재 웹콘텐츠의 경우 개인화에 대한 연구가 진행 되고 있으나 주로 콘텐츠를 기반으로 하거나 웹로그를 분석하는 등의 방법을 사용하고 있기 때문에 사용자의 명시적인 관심사항을 반영하기 곤란하다. 따라서 본 논문에서는 사용자가 웹문서에서 특정 주제에 대한 관심사항을 명시적으로 표현하는 경우 효과적인 트랜스코딩 기법을 제안한다.

1. 서 론

공간이나 시간에 구애받지 않고 디지털콘텐츠를 사용하고 자 하는 사용자의 욕구로 인하여 무선 인터넷 및 디지털 TV, 데이터 방송, 그리고 전자책 등의 시장은 급격히 성장하고 있다. 그러나 무선 인터넷 및 데이터 방송 등과 같은 환경에서는 휴대폰과 PDA와 같이 단말기의 제약성(낮은 CPU 성능, 작은 출력 화면, 입출력 방법의 단순함 등)이 존재한다. 따라서 이러한 단말기에서 콘텐츠를 서비스하기 위해서는 단말기의 제약성을 고려하여 콘텐츠를 변환해야 한다. 이와 같이 디지털 콘텐츠를 단말기의 환경에 따라 변환하는 것을 트랜스코딩(transcoding)이라고 한다[그림 1].



[그림 1] 트랜스코딩 과정의 예

트랜스코딩과정은 기존의 많은 양의 콘텐츠를 변환 하여야 하므로 많은 양의 시간적, 물질적인 투자를 요구한다. 또한

현재까지 디지털콘텐츠 서비스 분야에서는 HTML기반의 웹 문서만을 고려하고 있으며, 이로 인해 HTML 이외의 다른 양질의 콘텐츠들은 무선 인터넷 등과 같은 환경에서 제공되지 못하고 있다. 따라서 콘텐츠 제공자는 HTML 형태의 콘텐츠만을 제공하고 있으며, 사용자는 제한된 콘텐츠만을 접하고 있다.

최근 들어 사용자의 컨텍스트 정보에 기반한 웹문서 개인화에 대한 연구가 급속히 증가 하고 있다[1,2,3]. 기존의 많은 양의 웹콘텐츠로부터 사용자의 컨텍스트를 고려한 웹문서 개인화는 사용자의 인지적 부담을 줄일 수 있고 사용자의 취향에 맞게 정보를 제공할 수 있는 장점이 있다.

웹 콘텐츠 개인화에 대한 연구는 주로 콘텐츠 기반 기술과 사용자 컨텍스트에 기반한 기술로 분류 한다. 콘텐츠 기반 기술은 키워드나 스트링 매칭을 통해서 웹 콘텐츠를 분류하는 기술이고 사용자 컨텍스트 기반 기술은 사용자의 웹접속 패턴이나 로그 파일에 근거하여 웹콘텐츠를 분류 하는 기술이다.

그러나 기존의 개인화에 대한 연구들은 서비스 제공자의 중심의 연구가 진행 되었고 있기 때문에 사용자가 관심주제에 대해서 명시적인 관심사항을 표현하는 경우 관심사항을 반영하는 기술에 대한 연구는 미흡하다.

따라서 본 연구에서는 HTML 문서뿐만 아니라, XML과 같은 표준문서를 효과적으로 트랜스코딩할 수 있는 기술을 제안하고자 하며, 동시에 사용자의 명시적인 관심사항을 반영할 수 있는 기술을 제안한다.

*본 논문은 2003년도 한국학술진흥재단의 지원에 의하여 연구 되었음(KRF-2003-041-O00533)

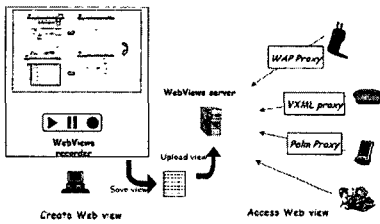
2. 관련 연구

디지털 콘텐츠의 트랜스코딩은 일반적으로 미디어 트랜스코딩과 웹 문서 트랜스코딩으로 분류[4]한다. 본 논문에서 다루고자 하는 웹 문서 트랜스코딩은 PDA와 같은 소형 단말기의 등장에서부터 현재까지 다양한 연구가 이루어져 왔고, 최근에는 W3C Device Independent Working Group를 중심으로 웹 콘텐츠 변환에 관한 연구가 활발하게 진행되고 있다[5]. 트랜스코딩과 웹콘텐츠 개인화에 대한 기존의 연구를 살펴보면 다음과 같다.

Annotation-Based Web Content Transcoding (2000)[6]는 IBM에서 제안한 트랜스코딩 시스템으로 HTML 문서를 대상으로 한 프락시 측 변환 기법이다. 이 연구에서는 서비스 제공자의 의견을 반영하는 어노테이션 기법을 적용하여 페이지 변환 규칙을 작성할 수 있도록 하였다. RDF 기반으로 작성된 외부 어노테이션 파일을 이용하여 트랜스코딩이 이루어진다.

NAC : A Basic Core for the Adaptation and Negotiation of Multimedia Service(2001)[4]는 프랑스 국립연구소 Inria에서 멀티미디어 전자문서 구조 및 표현 모델의 개발을 위한 Opera 프로젝트의 일부로 개발된 트랜스코딩 시스템이다. CC/PP에 기반 하여 UPS(Universal Profile Schema)를 정의하였으며, 이는 서버 측 프로파일과 클라이언트 측 프로파일로 구성된다.

기존의 웹 페이지 환경에서의 콘텐츠 개인화(Contents Personalization)에 관한 연구는 다양한 관점에서 이루어졌으나, 트랜스코딩 환경에서의 연구는 미흡한 실정이다. WebViews 시스템[7]은 2001년 Bell 연구소에서 개발한 트랜스코딩 시스템으로 모바일 환경에서의 콘텐츠 개인화를 목적으로 설계되었다.



[그림 2] WebViews 시스템의 개요

시스템의 설계 목적은 이미지, 플래시, 자바스크립트 등으로 복잡하게 구성된 웹 페이지를 사용자의 접속 장비와 관심에 맞추어 '간략화', '개인화' 하는데 있다.

Web Clip[8]은 최근 PDA를 중심으로 이용되고 있는 트랜스코딩 서비스이다. 주로 신문기사나 날씨, 증시상황, 게시판 등 주기적으로 업데이트되는 웹페이지의 일부분을 클리핑하여 사용자가 원하는 정보를 PDA 상에서 이용하여 온라인이나 오프라인 형태로 콘텐츠를 제공한다. 국내에서는 네이트온의 웹 클립 서비스가 대표적이다.

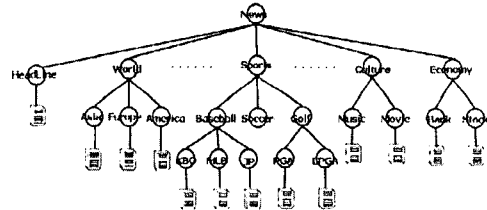
3. 사용자 관심사항을 반영한 XML 문서 변환

본 논문에서 제안하는 개인화된 트랜스코딩 시스템은

주제별 계층 분류가 가능한 뉴스 콘텐츠를 대상으로 한다. 뉴스 콘텐츠는 주제별 계층 분류가 명확하기 때문에 특정 주제에 대한 사용자의 관심사를 반영하기에 적합하다.

3.1 XML 기반 뉴스 콘텐츠의 계층 구조

XML 기반의 구조적인 뉴스 콘텐츠 구성을 위해 뉴스의 주제별 계층 구조를 정의하였다. 뉴스의 주제별 계층 구조는 국내 및 해외 뉴스 서비스의 일반적인 분류 체계를 따르며, 실제 기사의 내용은 뉴스의 저작과 전송을 위한 표준 포맷인 IPTC(International Press Telecommunication Council)의 NewsML을 기반으로 한다. XML 기반 웹 뉴스의 전체 구성은 주제별 계층 구조와 실제 기사 인스턴스들의 집합으로 이루어진다. 뉴스 콘텐츠는 총 11개의 주제로 분류되며, 각각의 주제 대항 세부 항목들이 존재한다. XML 기반으로 정의된 뉴스의 구조 정보 활용을 위해 DOM(Document Object Model) 트리를 이용하여 처리한다. [그림 3]은 뉴스의 주제별 계층 구조를 나타낸 DOM 트리의 예이다.

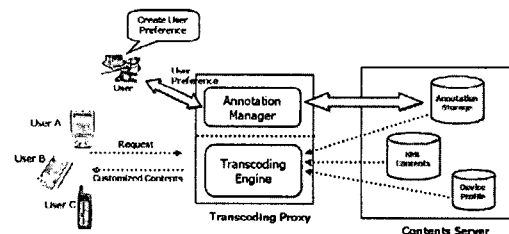


[그림 3] 주제별 계층 구조를 나타낸 DOM 트리의 예

3.2 트랜스코딩 프레임워크

사용자는 데스크 탑 환경의 웹 문서를 대상으로 특정 주제에 대한 관심사항을 명시적으로 표시하는 할 수 있다. 이를 PDA와 같은 모바일 기기가 일반적으로 데스크 탑과 연계하여 사용되고 있으며, 웹 페이지 상에서 특정 항목에 대한 사용자의 관심사를 명시적으로 표현할 수 있기 때문이다.

모바일 단말기에 적합한 실제 콘텐츠 변환은 트랜스코딩 엔진에서 동적으로 처리되는데, 원본 XML 콘텐츠, 디바이스 프로파일, 사용자 관심정보를 종합하여 각 사용자에게 맞는 개인화된 콘텐츠를 제공하게 된다. [그림 4]는 사용자 관심사항을 반영한 트랜스코딩 프레임워크를 나타낸다.

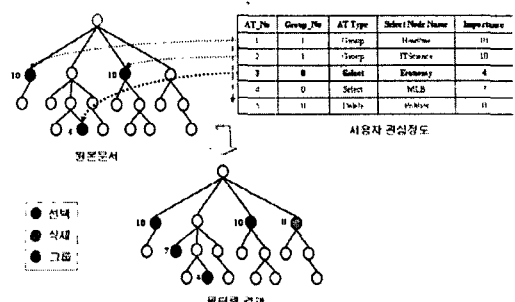


[그림 4] 트랜스코딩 시스템 프레임워크

3.3. 구조정보를 활용한 콘텐츠 변환

문서의 구조정보와 사용자 관심정보를 이용한 트랜스코딩의 과정은 크게 '원본문서의 재구성'과 'XSLT 변환'으로 구성된다. 원본 문서의 재구성 과정은 엘리먼트 필터링, 사용자 지정정보에 근거한 중요도 확산, 주제별 우선순위 결정으로 이루어진다. 엘리먼트 필터링을 통해 선택, 삭제, 그룹, 중요도 등의 정보를 원본문서에 반영하고, 사용자 지정정보에 근거하여 사용자가 지정하지 않은 부분을 처리한다. 다음단계로 모든 노드의 중요도가 결정되면 중요도 값을 이용하여 주제별 우선순위를 결정한다. XSLT 변환 과정에서는 주제별 우선순위 정보를 이용하여 XSLT 변환 스크립트의 동적 생성 및 변환한다.

다음 [그림 5]와 [그림 6]은 엘리먼트 필터링과정으로 트랜스코딩을 위한 첫 번째 단계이다. 이단계에서는 사용자 관심정보를 원본문서 트리에 반영한다. 파싱된 정보는 테이블 형태의 데이터 구조로 변환되어 첫 번째 행부터 순차적으로 처리 된다. 이는 사용자가 선택한 주제에 해당하는 트리상의 노드를 찾아 사용자 지정정보의 유형과 중요도 값을 부여한다.



[그림 6] 엘리먼트 필터링의 예

을 구성하였으며, 최종적으로 XSLT를 거쳐 단일기에 적합한 형태로 최종적으로 변환 한다.

향후 연구로는 W3C의 표준 언어인 CC/PP를 이용하여 다양한 디바이스에 맞게 XML 문서 변환하는 XSLT 스크립트 동적 생성 방법에 대한 연구가 요구된다.

```

useAnnotation [] : Array of user annotation data structure
(atNo, GroupNo, atType, nodeName, Importance)
elementFilter ( userAnnotation [] atData )
{
for ( the number of annotations )
{
case atData.atType == Select :
if atData.nodeName == Node name of Tree
setting of AT Type and Importance in Tree
case atData.atType == Delete :
if atData.nodeName == Node name of Tree
setting of AT Type and Importance in Tree
case atData.atType == StartGroup :
if atData.nodeName == Node name of Tree
setting of AT Type and Group No,
and Importance in Tree
case atData.atType == EndGroup :
if atData.nodeName == Node name of Tree
setting of AT Type and Group No in Tree
}
}
    
```

[그림 5] 엘리먼트 필터링의 과정

5. 참고 문헌

- [1] N. RAMAKRISHNAN, *PIPE: Web Personalization by Partial Evaluation*, IEEE Internet Computing, Vol. 4, No. 6, 21-31, 2000
- [2] S. Perugini and N. Ramakrishnan, *Personalizing Web Sites with Mixed-Initiative Interaction*, Vol. 5, Issue 2, 9-15, 2003
- [3] T. Lemlouma and N. Layaida, *Context-aware adaptation for mobile devices*, Proceeding Mobile Data Management, 2004 IEEE International Conference, 106-111, 2004.
- [4] Lemoluma and N. Layaida, *NAC: A Basic Core for the Adaptation and Negotiation of Multimedia Services*, OPERA Project, INRIA, 2001.
- [5] W3C Device Independence Working Group, <http://www.w3.org/2001/di/>
- [6] M. Hori, G. Kondoh, K. Ono, S. Hirose, and S. Singhal, *Annotation-Based Web Content Transcoding*, 9th World Wide Web Conference, Vol. 3, Issue 1-6, 197-211, 2000.
- [7] J. Freire, B. Kumar and D. Lieuwen, *WebViews: Accessing Personalized Web Content and Services*, 10th international conference on World Wide Web, 2001.
- [8] Nate PDA Service, <http://pda.nate.com>

4. 결론

본 논문은 최근 많은 연구가 진행 되고 있는 웹문서 개인화를 트랜스코딩환경에 적용하는 기술을 제안했다. 기존의 제작자 중심의 개인화와 트랜스코딩에 대한 연구는 사용자의 관심사를 명시적으로 반영하지 곤란했다. 사용자의 명시적인 관심사를 반영하기 위한 모델을 제시하였다. 이를 위해서 주제별 계층 분류가 명확한 NewsML을 대상으로 사용자 관심사항 저장 및 원본문서 변환 미지정 부분 처리를 위한 중요도 확산으로 시스템