

MyNews : 모바일 환경에서 사용자 관심사를 고려한 XML 문서 트랜스코딩

송 특 섭[†] · 이 진 상^{**} · 이 경 호^{***} · 손 원 성^{****} · 고 승 규^{*****} · 최 윤 철^{*****} · 임 순 범^{*****}

요 약

기존의 웹 콘텐츠를 PDA나 핸드폰등 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.

MyNews : Personalized XML Document Transcoding Technique for Mobile Device Users

Teuk-Seob Song[†] · Jin-Sang Lee^{**} · Kyong-Ho Lee^{***} · Won-Sung Sohn^{****} · Seung-Kyu Ko^{*****}
Yoon-Chul Choy^{*****} · Soon-Bum Lim^{*****}

ABSTRACT

Developing wireless internet service and mobile devices, mechanisms for web service access are various. However, the existing web infrastructure and content were designed for desktop computers and are not well-suited for other types of accesses, e.g. PDA or mobile phone that have less processing power and memory, small screens, limited input facilities, or network bandwidth etc. Thus, there is a growing need for transcoding techniques that provide the ability to browse the web through mobile devices. However, previous researches on existing web contents transcoding are service provider centric, which does not accurately reflect the user's continuously changing interest. In this paper, we presents a transcoding technique involved in making existing news contents based on XML available via customized wireless service, mobile phone.

키워드 : Annotation(어노테이션), mobile device(모바일 디바이스), transcoding(트랜스코딩), personalization(개인화)

1. 서 론

공간이나 시간에 구애받지 않고 디지털 콘텐츠를 사용하고자 하는 사용자의 욕구로 인하여 무선 인터넷 및 디지털 TV, 데이터 방송, 그리고 전자책 등의 시장은 급격히 성장하고 있다. 아울러 무선 네트워크 및 모바일 컴퓨팅 기술의 발전은 언제 어디서든 인터넷을 사용할 수 있도록 하고 있

다. 그러나 기존의 데스크탑 중심으로 제작된 웹 문서를 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).

기존의 트랜스코딩에 대한 연구들은 주로 HTML환경에서 이루어 졌기 때문에 HTML 이외의 다른 양질의 콘텐츠들은 무선 인터넷 등과 같은 환경에서 제공되지 못하고 있다[1,2,3,4,5,6,7]. XML은 사실상 웹문서의 표준으로 자리 잡고 있으며 기존의 HTML위주의 웹문서는 XML문서로 대체되고 있다. 따라서 XML문서 환경에서와 같이 구조정보를

* 본 논문은 2003년도 한국학술진흥재단의 지원에 의하여 연구 되었음
(KRF 2003 041-D00033)

[†] 준 회 원 : 연세대학교 컴퓨터과학과 박사과정

^{**} 정 회 원 : LG전자기술원 연구원

^{***} 정 회 원 : 연세대학교 컴퓨터과학과 조교수

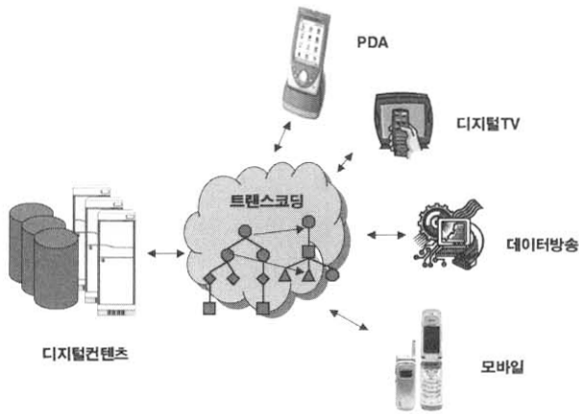
^{****} 준 회 원 : Carnegie Mellon University, Associate Researcher

^{*****} 준 회 원 : University of Maryland Associated Researcher

^{*****} 정 회 원 : 연세대학교 컴퓨터과학과 교수

^{*****} 총신회원 : 숙명여자대학교 멀티미디어학과 교수

논문접수 : 2004년 11월 20일, 심사완료 : 2005년 3월 8일



(그림 1) 트랜스코딩 과정의 예

활용한 트랜스코딩 기술이 필요하다.

한편, PDA나 핸드폰 같은 모바일 디바이스는 데스크탑 컴퓨터 보다 작은 화면을 갖고 있다. 기존의 웹문서를 그대로 모바일 디바이스에서 보게 되면 원하는 정보를 얻기 위해 더 많은 스크롤이나 클릭이 필요하게 된다. 기존의 트랜스코딩에 대한 연구들은 주로 디바이스의 성능에 따른 연구가 주를 이루고 있기 때문에 트랜스코딩환경에서 사용자 중심의 연구가 요구된다. 이와 같이 사용자 위주로 웹컨텐츠를 재구성 하는 웹컨텐츠 개인화(Personalization)에 연구가 이루어지고 있다. 웹컨텐츠 개인화에 대한 연구는 주로 컨텐츠 기반 기술과 사용자 컨텍스트에 기반한 기술로 분류한다[8]. 컨텐츠 기반 기술은 사용자가 자주 사용하는 키워드나 스트링을 사용하여 키워드 매칭이나 스트링 매칭을 통해서 웹컨텐츠를 분류하는 기술이고 사용자 컨텍스트 기반 기술은 사용자의 웹검색 패턴이나 로그 파일에 근거하여 웹컨텐츠를 분류 하는 기술이다.

그러나 기존의 스트링 매칭기법이나 사용자 컨텍스트에 기반한 기술은 서비스 제공자의 중심의 연구이기 때문에 계속적으로 변화하는 사용자의 관심사항을 정확히 반영하기 곤란하다. 따라서 본 연구에서는 사용자가 명시적으로 관심사항을 표시하는 경우 사용자의 의도를 명확히 반영하는 트랜스코딩 기법을 제안하고, HTML 문서뿐만 아니라 XML[9]과 같은 표준문서를 효과적으로 트랜스코딩할 수 있는 인터페이스 및 기법을 제안하고자 한다.

2. 관련연구

디지털 컨텐츠의 트랜스코딩은 일반적으로 미디어 트랜스코딩과 웹 문서 트랜스코딩으로 분류[10]한다. 본 논문에서 다루고자 하는 웹 문서 트랜스코딩은 PDA와 같은 소형 단말기의 등장에서부터 현재까지 다양한 연구가 이루어져 왔고, 최근에는 W3C의 Device Independent working group[11]을 중심으로 웹컨텐츠 변환에 관한 연구가 활발하게 진행되고 있다. 기존의 트랜스코딩 기법들과 모바일 환경에서의

컨텐츠 개인화 기술에 대해 간략히 소개한다.

Annotation-Based Web Content Transcoding(2000)[4]은 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(Resource Description Framework)[12] 기반으로 작성된 외부 어노테이션 파일을 이용하여 트랜스코딩이 이루어지며, 저작단계에서부터 트랜스코딩 된 결과를 확인해 볼 수 있는 틀을 제공한다. 또한 우선순위 부여를 통한 페이지 분할 기법을 적용하여 데이터양을 감소시키며 중요사항을 페이지 상위에 보여줌으로써 컨텐츠 이용의 효율성을 높여준다.

NAC : A Basic Core for the Adaptation and Negotiation of Multimedia Service(2001)[10] NAC는 프랑스 국립 연구소 Inria에서 멀티미디어 전자문서 구조 및 표현 모델의 개발을 위한 Opera 프로젝트의 일부로 개발된 트랜스코딩 시스템이다. CC/PP(Composite Capabilities/Preferences Profiles) [13]에 기반 하여 UPS(Universal Profile Schema)를 정의하였으며, 이는 서버 측 프로파일과 클라이언트 측 프로파일로 구성된다. 서버 측 프로파일에는 문서의 인스턴스(HTML, WML 등), 자원(jpg, gif, au 등) 그리고 어노테이션과 유사한 형태의 변환 규칙(XSLT Style Sheet 등)을 포함하며, 클라이언트 측 프로파일에는 사용자 디바이스 정보에 관한 배달문맥(Delivery Context)을 기술한다. 실제 컨텐츠 변환은 UCM(User Context Module)과 ANM(Adaptation and Negotiation Module)에서 이루어지며, 프로파일 정보를 고려하여 사용자 디바이스에 적합한 형태로 변환한다.

Intel CC/PP Toolkit(2002)[11] 인텔사는 자사의 무선 디바이스 아키텍처인 PCA(Personal Internet Client Architecture)를 탑재한 모바일 기기에 적합한 트랜스코딩 서비스를 제공하기 위해 CC/PP 툴킷을 개발하였다. 아파치 웹서버 모듈에 CC/PP 툴킷을 탑재하여 트랜스코딩 서비스 제공이 가능하며 프로파일의 저장, 관리 및 컨텐츠 변환 모듈이 포함되어 있다. HTTP와 함께 CC/PP 헤더를 요청 받으면, 서버에 저장된 CC/PP정보를 이용하여 기존에 만들어진 데이터를 선택하거나 동적으로 컨텐츠를 변환한다.

WebViews 시스템[3]은 2001년 Bell 연구소에서 개발한 트랜스코딩 시스템으로 모바일 환경에서의 컨텐츠 개인화를 목적으로 설계되었다. 시스템의 설계 목적은 이미지, 플래시, 자바스크립트 등으로 복잡하게 구성된 웹 페이지를 사용자의 접속 장비와 관심에 맞도록 '간략화', '개인화' 하는데 있다. WebClip[14]은 최근 PDA를 중심으로 이용되고 있는 트랜스코딩 서비스이다. 주로 신문기사나 날씨, 중시상황, 게시판 등 주기적으로 업데이트되는 웹페이지의 일부분을 클리핑하여 사용자가 원하는 정보를 PDA 상에서 이용하며 온라인이나 오프라인 형태로 컨텐츠를 제공한다. 오프라인 서비스는 주로 Palm 기반의 PDA에서 이용되는 방식으로, 사용자가 선택한 클립을 PDA에서 이용할 수 자체 포맷으로 미리 변환한다.

Annotation-Based [4]의 경우 저작자의 의견을 반영하는 기법과 저작 틀을 제안하였고 NAC[10] 및 Intel CC/PP 툴킷[15]의 경우 디바이스의 성능에 따른 트랜스코딩 기술을 제안 하였으나 기존의 연구에서는 사용자의 관심사항을 반영하지는 못한다. 벨(Bell) 연구소에서 개발한 WebView나 NateOn[3,14] 서비스의 경우 플래시나 자바 스크립트등 다양한 웹페이지에 대한 트랜스코딩을 제안하고 있으나 사용자의 관심을 명확히 반영하지는 않는다. 앞서 살펴본 기존의 트랜스코딩 기법들과 소프트웨어 기술을 서비스 제공자와 사용자의 의도 반영 여부, 트랜스코딩 기법, 대상 콘텐츠를로 분류하여 각각의 특징을 요약하면 <표 1>과 같다.

<표 1> 기존의 트랜스코딩 관련 연구 요약

Title	서비스 제공자 의도	사용자 의도	트랜스코딩 기법	대상 콘텐츠	연도
Annotation-Based Web Content Transcoding[4]	YES	NO	레이아웃, 어노테이션	HTML	2000
NAC[13]	YES (부분가능)	NO	CC/PP	HTML, XML	2001
WebViews[3]	NO	YES	레이아웃, 디바이스 프로파일	HTML	2001
Intel CC/PP SDK[11]	NO	NO	CC/PP	XML	2002
WebClip[14]	NO	YES	레이아웃	HTML	2004

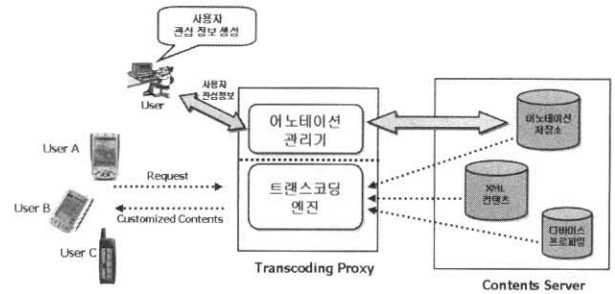
3. 사용자 어노테이션을 이용한 XML 문서 트랜스코딩

본 장에서는 사용자 관심사를 고려한 개인화된 콘텐츠 변환을 위해 어노테이션을 이용한 XML 문서 트랜스코딩 기법을 제안한다. 3.1절에서는 어노테이션을 이용한 트랜스코딩 프레임워크의 개략적인 내용을 설명하고, 3.2절에서는 XML 기반 뉴스 콘텐츠의 주제별 계층 구조를 정의한다. 3.3절에서는 사용자 어노테이션 모델을 정의하고, 3.4절에서는 사용자 어노테이션을 이용한 원본 문서의 재구성 및 주제별 우선순위 결정 규칙에 대해 설명한다.

3.1 MyNews: 트랜스코딩 프레임워크

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. 특정 주제에 대한 사용자의 관심사를 표현하기 위한 방법으로 어노테이션을 이용한다. 사용자는 데스크 탑 환경의 웹 문서를 대상으로 어노테이션을 생성하고 저장한다. 이는 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 어노테이션 정보를 종합하여 각 사용자에 맞는 개인화된 콘텐츠를 제공하게 된다. (그림 2)는 사용자 어노테이션을 이용한 트랜스코딩 시스템의 개요이다. 사용자가 데스크탑 환경에서 생성한 어노테이션을 어노테이션 관리기를 통해 어노테이션 저장소 저장하고 모바일 디바이스를 통해 웹컨텐츠를 요청하게 되면 트랜스코딩 엔진에서 사용자가 생성한 어노테이션에 근거하여 사용자의 요구에 맞는 웹페이지를 제공한다.

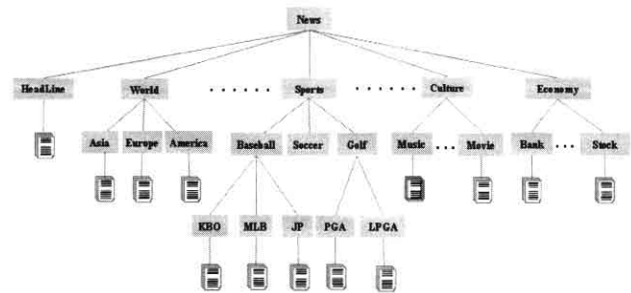


(그림 2) 트랜스코딩 시스템 개요

3.2 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 NewsML[9]을 기반으로 한다. XML 기반 웹 뉴스의 전체 구성은 주제별 계층 구조와 실제 기사 인스턴스들의 집합으로 이루어진다.

뉴스 콘텐츠는 총 11개의 주제로 분류되며, 각각의 주제에 대한 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) 뉴스의 주제별 계층 구조

트리의 노드는 해당 주제별 분류 항목을 나타내며, 실제 기사의 내용은 최하위 노드의 링크 형태로 제공된다.

3.3 사용자 어노테이션(Annotation) 모델

사용자의 관심사를 반영한 의미 있는 트랜스코딩을 위해

서는 특정 주제에 대한 사용자 관심 정보 및 콘텐츠 변환 규칙을 기술한 메타 정보가 필요하다. 본 절에서는 콘텐츠 변환에 필요한 메타 정보 기술을 위해 사용자 중심의 어노테이션 모델을 정의한다. 일반적으로 트랜스코딩에서 의미하는 어노테이션은[4,5] ‘문서 변환에 필요한 부가정보’라고 할 수 있다. 그러나 본 논문에서는 ‘콘텐츠 변환 규칙뿐만 아니라 사용자의 선호도를 표현하기 위한 인터페이스’로 어노테이션의 의미를 확장하도록 한다.

개인화된 트랜스코딩을 위해 필요한 어노테이션의 종류는 선택, 삭제, 그룹의 세 가지로 분류한다. 선택 및 삭제는 특정 주제에 대한 콘텐츠 변환의 유·무를 나타내며, 그룹은 두 가지 이상의 주제에 대한 연관관계를 나타낸다. 선택, 삭제, 그룹을 고려한 사용자 어노테이션의 DTD는 (그림 4)와 같다. 사용자 어노테이션은 단일 어노테이션(Single)과 그룹 어노테이션(Group)의 집합으로 구성된다. 단일 어노테이션은 한 가지 주제에 대한 선택 및 삭제를 의미하며, 그룹 어노테이션은 두 가지 이상의 주제에 대한 연관성을 의미한다.

```

<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT Coordinate EMPTY>
<!ATTLIST Coordinate
  startX NMTOKEN #IMPLIED
  startY NMTOKEN #IMPLIED
  finishX NMTOKEN #IMPLIED
  finishY NMTOKEN #IMPLIED
  x NMTOKEN #IMPLIED
  y NMTOKEN #IMPLIED
  width CDATA #IMPLIED
  height CDATA #IMPLIED
>
<!ATTLIST GAnnotation
  Importance CDATA #REQUIRED
>
<!ELEMENT UserAnnotation (UserID, Single, Group)>
<!ELEMENT UserID (#PCDATA)>
<!ELEMENT Single (SAnnotation+)>
<!ELEMENT Group (GAnnotation)>
<!ELEMENT SAnnotation (Style, Importance, Coordinate, ,
  TopSubject SubSubject*)>
<!ELEMENT GAnnotation (Item+)>
<!ELEMENT TopSubject (#PCDATA)>
<!ELEMENT SubSubject (#PCDATA)>
<!ELEMENT Importance (#PCDATA)>
<!ELEMENT Item (Style, Coordinate, Path)>
<!ELEMENT Path (#PCDATA)>
<!ELEMENT Style (#PCDATA)>
    
```

(그림 4) 사용자 어노테이션 DTD

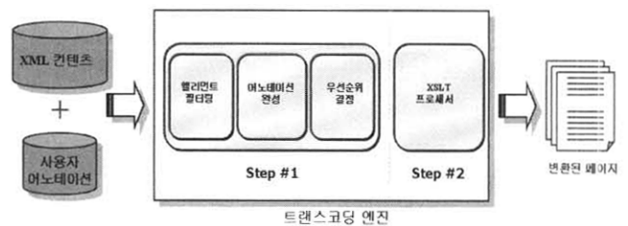
3.3.1 어노테이션 생성

뉴스 콘텐츠를 대상으로 하여 정의된 구조에 유효한 어노테이션 데이터를 생성하기 위해서는 몇 가지 제한 규칙을

정의했다. 어노테이션 생성은 헤드라인과 최근 중요기사를 제외한 나머지 주제별 분류 항목을 대상으로 어노테이션을 생성한다. 헤드라인과 최근 중요기사는 일반적으로 가장 중요한 기사이기 때문에 사용자 관심사와 상관없이 보여준다. 선택(Select)은 특정 주제에 대한 선택은 주제별 대분류 항목뿐만 아니라 하위분류 항목의 중복 선택으로 가능하다. 이는 사용자의 관심사를 좀 더 복합적으로 고려할 수 있게 한다. 삭제(Delete)는 특정 주제에 대한 삭제는 주제별 대분류 항목에서만 가능하며 하위분류 항목은 선택할 수 없다. 그룹(Group)은 두 개 이상의 대분류 항목을 순차적으로 선택하며, 하위분류 항목은 선택할 수 없다. 중요도(Importance)는 사용자가 선택한 특정 주제에 대한 중요도는 0에서 10사이의 값이다. 사용자가 선택 가능한 중요도 값은 ‘매우중요’, ‘중요’, ‘보통’의 3가지로 제한하였다. 삭제에 해당하는 중요도 값은 0으로 자동 지정된다.

3.4 구조정보와 어노테이션을 이용한 콘텐츠 변환

원본 문서의 구조정보와 어노테이션을 이용한 트랜스코딩의 과정은 크게 2단계로 나누어 진행된다. 1단계는 ‘원본문서의 재구성’이고 2단계는 ‘XSLT 변환’과정이다. 원본 문서의 재구성 과정은 다시 ‘엘리먼트 필터링’, ‘어노테이션의 완성’, ‘주제별 우선순위 결정’으로 이루어진다. 엘리먼트 필터링을 통해 선택, 삭제, 그룹, 중요도 등의 어노테이션 정보를 원본문서에 반영하고, 어노테이션 완성 규칙을 적용하여 사용자가 지정하지 않은 부분을 처리한다. 어노테이션의 완성을 통해 모든 노드의 중요도가 결정되면 중요도 값을 이용하여 주제별 우선순위를 결정한다. XSLT 변환 과정에서는 주제별 우선순위 정보를 이용하여 XSLT 변환 스크립트의 동적 생성 및 변환 한다. (그림 5)는 구조정보와 어노테이션을 이용한 트랜스코딩의 과정을 도식화한 것이다. 각 단계에 대한 세부적인 내용은 다음과 같다.



(그림 5) 어노테이션을 이용한 트랜스코딩 과정

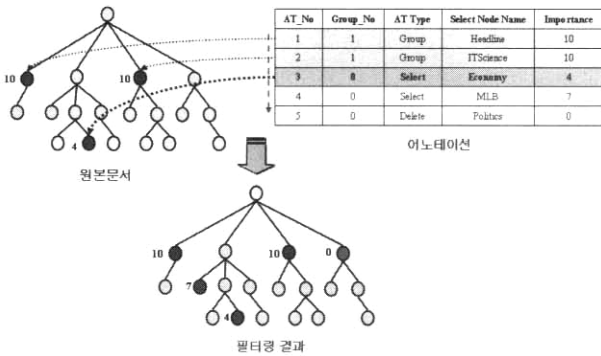
3.4.1 엘리먼트 필터링

엘리먼트 필터링은 트랜스코딩을 위한 첫 번째 단계로 사용자 어노테이션 정보를 원본문서 트리에 반영한다. 파싱된 어노테이션 정보는 테이블 형태의 데이터 구조로 변환되어 첫 번째 행부터 순차적으로 처리 된다. 이는 사용자 어노테이션의 개수만큼 반복되며, 사용자가 선택한 주제에 해당하는 트리상의 노드를 찾아 어노테이션 유형과 중요도 값을 부여한다.

```

useAAnnotation [] : Array of user annotation data structure
(atNo,GroupNo, atType, nodeName, Importance)
elementFilter ( userAAnnotation [] 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 6) 엘리먼트 필터링의 과정



(그림 7) 어노테이션을 이용한 엘리먼트 필터링의 예

3.4.2 어노테이션의 완성

어노테이션 되지 않은 사용자 미지정 노드의 중요도를 부여하기 위해서는 엘리먼트 필터링 결과를 이용한 어노테이션의 완성 규칙이 필요하다. 어노테이션의 완성 규칙은 사용자가 지정한 노드와 중요도 값에 근거하여 인접 노드의 중요도 값을 결정하는 과정이다. 다음은 선택, 삭제, 그룹 및 해당 노드의 깊이를 고려한 어노테이션의 완성 규칙이다.

[규칙 1] 주제별 대분류 노드(Level=2)에 대한 선택, 삭제, 그룹의 경우, 자식 노드의 중요도는 부모 노드의 중요도를 그대로 계승 받는다.

[규칙 2] 선택된 노드가 대분류에 대한 하위 노드(Level=3)이고 자식 노드를 갖지 않은 경우, 부모 노드에는 중요도 값을 그대로 전달하고 선택되지 않은 형제 노드의 중요도는 a ($0 \leq a \leq 10$)만큼 감소시킨다.

[규칙 3] 단말 노드(Level=4)의 경우, 부모와 조상 노드에는 중요도 값을 그대로 전달하고 형제 노드의 중요도는 a 만큼 감소시킨다. 또한 해당 주제 내에서 선택되지 않은 나머지 노드들은 모두 a 감소된 중요도 값을 갖는다.

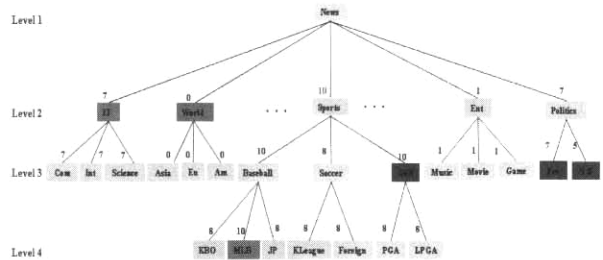
[규칙 4] 특정 주제에 대해 선택된 노드가 전혀 없는 경우, 모든 자식 노드의 중요도 값은 1을 갖는다.

```

useAAnnotation [ ] : Array of user annotation data structure
(atNo,GroupNo, atType, nodeName, Importance)
CompletionOfAnnotation ( userAAnnotation [ ] atData )
{
if ( Level of Current Node == 2 && Importance == P )
for ( the number of child nodes )
Importance of Child Node = P
else if ( Level of Current Node == 3 && Importance == P )
{ Importance of Parent Node==P;
for ( the number of sibling nodes )
Importance of Sibling Node = P - a ; // a (0 ≤ a ≤ 10)
}
else if ( Level of Current Node == 4 && Importance == P )
Importance of Parent Node = P;
Importance of Ancestor Node = P;

for ( the number of sibling nodes )
Importance of Child Node = P - a ; // a (0 ≤ a ≤ 10)
for ( the number of remain nodes )
Importance of remain Node = P - a;
else if ( Current Node is not Annotated Node )
for ( the number of all child nodes )
Importance of Child Node = 1;
}
    
```

(그림 8) 어노테이션 완성 규칙의 적용에



(그림 9) 어노테이션 완성 규칙의 적용 결과의 예

3.4.3 주제별 우선순위 결정

엘리먼트 필터링과 어노테이션 완성 규칙을 적용하여 모든 노드의 중요도가 결정되면, 각 노드의 중요도 값에 따라 출력에 필요한 우선순위가 결정된다. 대분류 및 하위 주제별 우선순위는 노드의 중요도가 큰 순서에 따라 결정되며

<표 2> 주제별 우선순위 결정의 예

주제별 우선순위	대분류 주제	세부 주제에 대한 우선순위	Path
1	Headline	없음	News/Headline
2	Top Stories	없음	News/Topstories
3	Sports	1	News/Sports/Baseball/MLB
		2	News/Sports/Baseball/KBO News/Sports/Baseball/JP
		3	News/Sports/Golf/PGA News/Sports/Golf/LPGA
		4	News/Sports/Soccer/KLeague News/Sports/Soccer/Foreign
4	Politics	1	News/Politics/President
		2	News/Politics/NorthandSouth
5	IT	없음	News/IT/Computer News/IT/Internet News/IT/Science
6	Ent	없음	News/Ent/Music News/Ent/Movie News/Ent/Game

노트간의 중요도가 같을 경우, $\sum(\text{선택된 하위 노트의 깊이} * \text{중요도})$ 값의 순서대로 주제별 우선순위가 결정된다. 만일 대분류 노트만을 선택한 [규칙 1]과 같은 경우는 자식 노트의 개수에 따라 우선순위를 결정한다.

<표 2>는 주제별 우선순위를 결정한 예이다. 이와 같이 대분류 및 하위 주제별 우선순위가 결정되면 콘텐츠 변환에 필요한 XSLT 스크립트가 동적으로 생성된다.

4. MyNews: 시스템의 구현 및 평가

본 장에서는 사용자 관심사를 고려한 XML 문서 트랜스코딩 시스템을 구현하고 그 결과를 분석 및 평가한다. 시스템은 JAVA를 기반으로 구현하였으며 스윙(Swing), JAXP (Java API for XML Processing), 서블릿(Servlet), XSLT를 사용하여 세부 모듈을 구성하였다.

4.1 시스템 구성

트랜스코딩 시스템은 크게 사용자 관심 정보 생성을 위한 어노테이션 인터페이스와 실제 콘텐츠 변환을 처리하는 트랜스코딩 엔진으로 구성된다.

(그림 10)은 제안하는 트랜스코딩 시스템의 전체 구조를 나타낸다. 사용자는 데스크탑 환경에서 제공되는 웹 페이지를 대상으로 뉴스의 특정 주제에 대한 어노테이션(선택, 삭제, 그룹)을 생성하며, 생성된 어노테이션 정보는 정해진 구조에 유효한 XML 문서로 저장된다. 이후 모바일 기기에서 해당 페이지를 요청하면, 트랜스코딩 엔진에서 접속 디바이

스의 프로파일 정보, 사용자 어노테이션 정보, XML 콘텐츠를 종합하여 주제별 우선순위 결정하고 XSLT 스크립트 생성하여 각각의 사용자에게 적합한 개인화된 페이지를 전송한다.

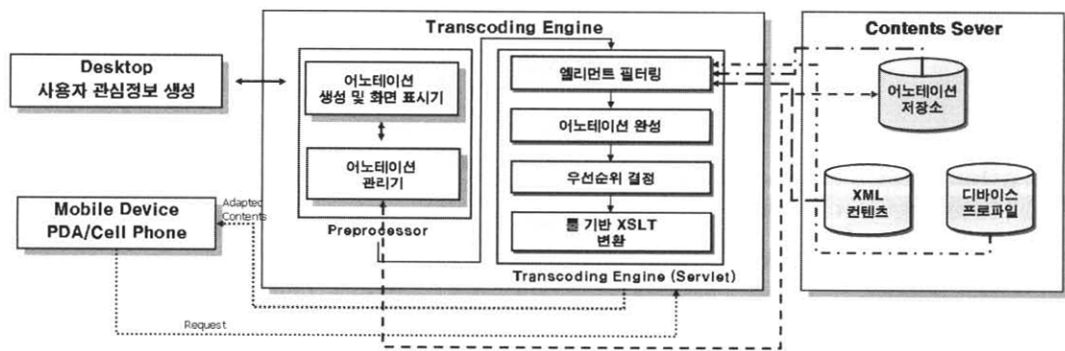
어노테이션 인터페이스는 일반적인 웹 브라우저와 유사한 형식으로 구현하였으며, '웹 문서 브라우징'과 '어노테이션 생성'의 두 가지 모드를 제공한다. 웹 문서 브라우징 모드에서는 웹 브라우저와 동일한 기능을 제공하며, 어노테이션 생성 모드로 변환하여 해당 문서에 사용자 관심 정보를 생성한다.

어노테이션 유형은 별도의 팝업창을 이용하여 선택하고 각각의 유형에 해당하는 스타일로 화면에 출력된다. 기존에 사용자가 생성한 어노테이션이 존재할 경우, 서버에 저장된 어노테이션 파일을 분석하여 화면상에 보여준다. 각각의 어노테이션은 특정 주제 영역을 드래킹하여 선택하고 중요도 및 세부 주제에 대한 선택은 별도의 팝업 메뉴를 통해 이루어진다. (그림 11 (a))는 구현된 어노테이션 인터페이스이고, (그림 11 (b))는 선택한 주제에 대한 중요도 및 세부 주제 선택을 위한 팝업 메뉴의 예를 나타낸다. (그림 11 (c))는 주제별 중요도를 선택하기 위한 윈도우이다.

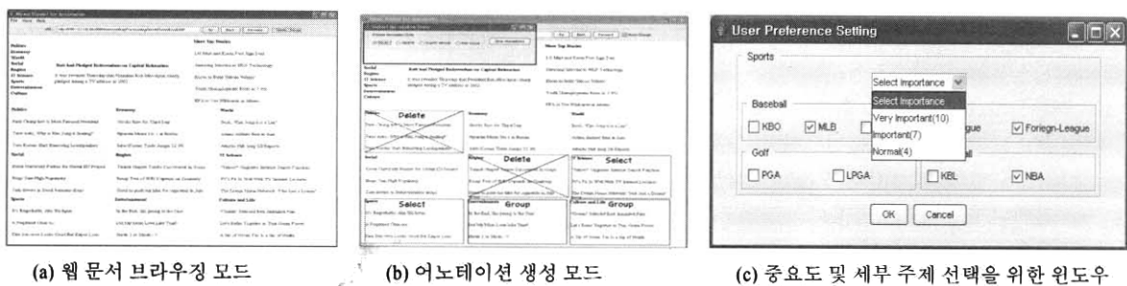
사용자 관심사가 반영되지 않은 경우, 시스템에 정해진 규칙에 따라 모든 콘텐츠가 순차적으로 변환되며, 사용자 관심사가 반영된 경우, 어노테이션 유형과 우선순위 정보를 고려하여 주제별 출력 순서가 결정된다.

4.2 시스템 분석 및 평가

본 절에서는 제안하는 트랜스코딩 시스템의 평가를 위하



(그림 10) MyNews: 시스템 구조도



(a) 웹 문서 브라우징 모드

(b) 어노테이션 생성 모드

(c) 중요도 및 세부 주제 선택을 위한 윈도우

(그림 11) 어노테이션 인터페이스



(a) 사용자 관심사가 반영되지 않은 경우
(시스템의 정해진 규칙에 따라 모든 내용이 출력)

(b) 사용자 관심사가 반영된 경우
(사용자 관심 정보에 따라 주제별 출력 순서가 결정)

(그림 12) 포켓PC 에뮬레이터를 대상으로 한 트랜스코딩 결과

여 기존 연구와의 비교 분석 및 사용자 평가를 통한 차별성을 검증한다. 제안하는 트랜스코딩 시스템은 데스크 탑 환경의 웹 페이지를 대상으로 사용자 관심 정보를 생성하고, 사용자 관심 정보에 따라 개인화된 트랜스코딩을 수행하는 시스템이다. 본 연구에서 제안한 시스템과의 비교 대상은 관련연구에서 제시한 'WebViews'[3]와 '네이트온의 'WebClip'[14]으로 사용자 관심 정보를 고려한 개인화된 트랜스코딩 시스템이라는 관점에서 상호 비교 분석이 가능하다. 본 연구에서 제안하는 시스템과 WebViews와 WebClip의 시스템의 차이점은 다음 <표 3>과 같다

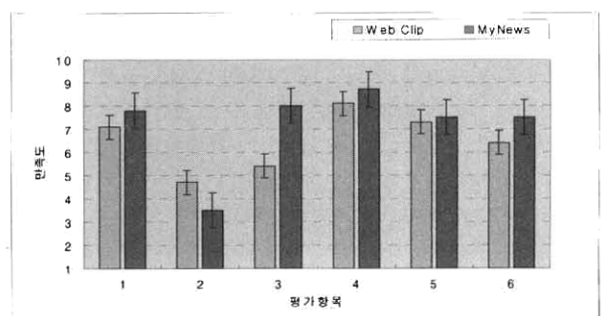
<표 3> 기존 시스템과의 비교

비교 항목	WebViews[3]	WebClip[14]	제안 시스템
트랜스코딩 대상 문서	HTML	HTML	XML
구조정보 이용	불가능	불가능	가능
사용자 관심 정보 생성 방법	메뉴 선택 방식	테이블 단위로 관심 정보를 선택	주제별 선택 영역 지정 및 팝업 메뉴를 통한 세부 주제 선택
선택한 정보들 간의 중요도	불가능	가능	가능

WebViews와 WebClip은 HTML 문서를 대상으로 하기 때문에 웹 페이지 상에 표현되는 1차적인 정보만을 대상으로 한다. 따라서 XML과 같이 구조 문서 환경에서의 구조 정보를 활용하여 사용자의 관심사항을 정교하게 반영하지 못한다. 사용자 관심정보 생성 방식을 살펴 보면 WebViews는 메뉴선택 방식을 이용하여 관심 정보를 생성하며 해당 페이지의 URL과 선택한 항목에 관한 정보를 통합하여 'Smart Bookmark'라는 자체 정의된 형식으로 저장된다. WebClip은 웹 페이지의 테이블 단위로 관심 정보를 선택하고 선택된 정보는 WebClip 서버에 저장된다. WebViews는 메뉴 방식을 통해 원하는 관심 정보를 생성할 수는 있지만, 특정 주제에 대한 중요도 값을 부여할 수는 없다. 따라서

시스템의 정해진 규칙에 따라 페이지가 생성된다. WebClip의 경우 웹상에서 관심 정보를 클릭평한후 선택된 항목에 대한 리스트를 보고 우선순위를 보여 주기 때문에 여러 단계를 거쳐야 한다. 반면에 제안하는 시스템은 사용자가 관심주제에 대해 선택과 동시에 우선순위를 부여 할 수 있으며, 각각의 주제 항목에 대해 중요도 값을 부여하도록 할 수 있기 때문에 사용자의 관심정보를 정교하게 반영 할 수 있다. WebClip의 경우, 출력이 필요한 우선순위를 사용자가 수동적으로 정의하지만, 제안 시스템은 선택한 주제별 중요도 값에 근거하여 내부 규칙을 통해 주제별 우선순위가 결정된다.

제안하는 시스템의 효율성을 검증하기 위해 사용자 평가를 수행하였다. 본 실험에서는 사용자 관심 정보 생성 인터페이스와 변환된 트랜스코딩 결과에 대한 만족도를 측정하기 위해 네이트온의 WebClip 서비스를 대상으로 비교 평가 하였다. 본 실험에는 20명의 피실험자(PDA 사용 경험군 10인 + 비 경험군 10인)가 참여하였으며, 마이크로소프트 포켓 PC 2003 에뮬레이터 환경에서 실험하였다. 평가는 크게 사용자 관심 정보 생성을 위한 어노테이션 인터페이스와 트랜스코딩 결과에 대한 만족도를 측정하는 관점에서 실시하였다. 실험 결과는 (그림 13)과 같이 WebClip보다 대체로 높은 만족도를 보였다.



(그림 13) 사용자 평가 결과

〈표 4〉 분산분석(ANOVA TEST) 결과

변동의 요인	제공합	자유도	제곱 평균	F 비	P-값	F 기각치
인자 A(행)	21.31667	5	4.263333	5.585153	0.041115	5.050329
인자 B(열)	1.333333	1	1.333333	1.746725	0.243529	6.607891
관차	3.816667	5	0.763333			
계	26.46667	11				

각 시스템에 따른 사용자들의 정확도를 분석하기 위하여 본 실험에서는 Two-way ANOVA(analysis of variance)를 사용하였다. <표 4>는 사용자들의 평가에 대한 평균을 나타내고 있으며 대체적으로 제안 기법이 기존 기법보다 높은 만족도를 유지할 수 있었다. 또한 실험결과 본 논문에서 제안하는 시스템이 기존의 시스템유 의한 차이가 $F(0.05, 5, 5) = 5.05$, $F(0.05, 1, 5) = 6.61$ 이고 $F비 = 5.58 > 5.05$ 있음을 확인 할 수 있었다.

평가 항목은 인터페이스의 편리성 3개 항목(1. 사용자 관심 정보 생성 및 저장 인터페이스는 편리한가? 2. 인터페이스 사용에 있어 부가적인 학습 과정이 필요한가? 3. 특정 주제에 대한 관심 정보를 명확히 생성할 수 있는가?)과 트랜스코딩 결과에 대한 만족도 3개 항목(4. 사용자 관심 정보가 반영되지 않은 결과에 비해 만족하는가? 5. 사용자의 관심사를 정확히 반영한 트랜스코딩 결과인가? 6. 페이지 간 네비게이션은 편리한가?)으로 구성하였다. 평가 항목 중 인터페이스에 대한 편리성에 대한 질문에 대한 질문 중 “관심사항을 명확하게 표현 할 수 있었는가?”에 대한 만족도 평가에서 WebClip에 비해 우수한 결과를 얻었다. WebClip의 경우 특정 항목에 대한 선택은 편리한 반면, 각 항목들 간의 우선 순위 부여와 저장의 과정에 있어 제안하는 시스템보다 절차가 복잡하기 때문인 것으로 분석된다.

두 시스템간 트랜스코딩 결과에 대한 만족도를 평가하기 위한 질문인 “사용자의 관심사를 정확히 반영한 트랜스코딩 결과인가?”에 대한 질문에서는 WebClip과 비슷한 만족도를 얻었다. 이는 시스템에서 의도하는 중요도 값과 사용자가 판단하는 중요도 값이 의미상으로 정확히 일치하지 않기 때문인 것으로 분석된다.

5. 결론 및 향후 연구 방향

본 논문에서는 모바일 환경에서의 개인화된 웹 콘텐츠 제공을 위해 사용자의 관심사를 고려한 XML 문서 트랜스코딩 기법을 제안하였다. 이를 위해 뉴스 콘텐츠를 대상으로 하는 사용자 어노테이션 모델 및 인터페이스를 설계하였으며, 어노테이션 정보를 이용한 XML 문서 변환 기법을 제안하였다. 본 논문에서 제안하는 트랜스코딩 기법은 크게 3가지 모듈로 구성되는데, 이는 특정 주제에 대한 사용자 어노테이션의 생성, 어노테이션을 이용한 원본 문서의 재구성 및 주제별 우선순위 결정, 단말기 제약성을 고려한 XSLT 동적 변환이다. 데스크 탑 환경의 웹 문서를 대상으로 하는

어노테이션의 생성은 사용자 관심사를 표현하는 명시적인 방법으로 사용자가 원하는 주제의 선택 및 해당 주제에 대한 중요도 값을 부여하게 된다. 원본문서의 변환 과정은 어노테이션에 기반한 엘리먼트 필터링, 사용자가 지정하지 않은 부분의 처리를 위한 어노테이션의 완성, 주제별 우선순위 결정으로 이루어지며, XSLT 프로세서를 거쳐 단말기에 적합한 형태로 최종 변환된다.

본 연구에서는 HTML 레이아웃 기반 트랜스코딩이나 CC/PP 기반의 일방적인 트랜스코딩이 아닌 사용자 중심의 개인화 된 콘텐츠 변환으로 서비스 이용의 효율을 높일 수 있었다. 또한 사용자 평가를 통해 기존의 시스템과 비교하여 우수한 만족도를 얻게 되었다. 본 연구 결과를 바탕으로 뉴스 콘텐츠뿐만 아니라 주제별 계층 구조를 가진 다양한 XML 문서의 트랜스코딩에 적용할 수 있을 것으로 기대된다.

참고 문헌

- [1] T.W. Bickmore and B.N. Schilit, "Digester: Device-independent access to the World Wide Web", *Proc. of the 6th International World Wide Web Conference, Computer Networks and ISDN Systems*, Vol. 29, 1075-1082, 1997.
- [2] M. Butler, F. Giannetti, R. Gimson, and T. Wiley, "Device Independence and the Web", *IEEE Internet Computing*, Vol. 6, Issue 5, 81-86, 2002.
- [3] J. Freire, B. Kumar, D. Lieuwen, "WebViews: Accessing Personalized Web Content and Services", *10th international conference on World Wide Web*, 2001.
- [4] 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.
- [5] M. Hori, K. Ono, T. Koyanagi, and M. Abe, "Annotation by Transformation for the Automatic Generation of Content Customization Metadata", *Proc. the First Inter. Conference on Pervasive Computing, Lecture Notes in Computer Science* Vol. 2414, 267-281, 2002.
- [6] A.W. Huang and N. Sundaresan, "A Semantic Transcoding System to Adapt Web Services for User with Disabilities", *Proc. fourth international ACM conference on Assistive technologies(ASSET'00)*, 156-163, 2000.
- [7] Y. Hwang, J. Kim, and E. Seo, "Structure-aware Web transcoding for mobile devices", *IEEE Internet Computing*, Vol. 7, Issue 5, 14-21, 2003.
- [8] N. Ramakrishnan, "PIPE: Web Personalization by Partial Evaluation", *IEEE Internet Computing*, Vol. 4, No. 6, 21-31, 2000.
- [9] NewsML Homepage, <http://www.newsml.org>

- [10] T. Lemoluma and N. Layaida, "NAC: A Basic Core for the Adaptation and Negotiation of Multimedia Services", *OPERA Project, INRIA, 2001.*
- [11] W3C Device Independence Working Group, <http://www.w3.org/2001/di/>
- [12] Resource Description Framework (RDF), <http://www.w3.org/RDF/>
- [13] CC/PP Information Page, <http://www.w3c.org/Mobile/CCPP/>
- [14] Nate PDA Service, <http://pda.nate.com>
- [15] Intel CC/PP Toolkit, http://www.intel.com/pca/developernetwork/download/bsp/ccpp_WP_finaldraft.pdf
- [16] T. Lemoluma and N. Layaida, "Universal Profiling for Content Negotiation and Adaptation in Heterogeneous Environments", *W3C Workshop on Delivery Context, 2002.*
- [17] T. Lemoluma and N. Layaida, "Context-Aware Adaptation for Mobile Devices", *2004 IEEE Inter. Conf. Mobile Data Management (MDM'04)*, 19-22, 2004.
- [18] W.Y. Lum and F.C.M. Lau, "A Context-Aware Decision Engine for Content Adaptation", *IEEE Pervasive Computing*, Vol. 1, Issue3, 41-49, 2002.
- [19] L. Suryanarayana and J. Hjelm, "CC/PP for content negotiation and contextualization", *Lecture Notes in Computer Science*, Vol. 1987, 239-245, 2001.
- [20] W. Ma, I. Bedner, G. Chang, A. Kuchinsky, and H.J. Zhang, "A Framework for Adaptive Content Delivery in Heterogeneous Network Environments", *Proc. MMCN*, Vol. 3969, 86-100, 2000.
- [21] 신회숙, 마평수, 조수선, 이동우, "소형 화면 단말기를 위한 웹 문서 변환 기법(Web Document Transcoding Technique for Small Display Devices)", *한국정보처리학회 논문지-D* Vol. 9 No. 6, 1145-1156, 2002.
- [22] IBM, WebSphere Transcoding Publisher, <http://www-3.ibm.com/software/webservers/transcoding/html>
- [23] Oracle, iAS WE, <http://otn.oracle.com/products/jdev/htdocs/partners/addins/exchange/iaswireless/index.html>
- [24] XSL Transformations (XSLT) Version 1.0 W3C Recommendation 16 November 1999, <http://www.w3c.org/TR/xslt>



이진상

e-mail : gr20000@rainbow.yonsei.ac.kr
 2001년 연세대학교(학사)
 2004년 연세대학교(석사)
 2004년~현재 LG전자기술원 연구원
 관심분야: XML 문서처리, GUI, Car Navigation System

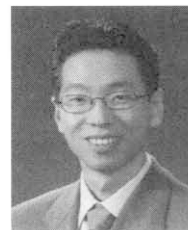


이경호

e-mail : khlee@cs.yonsei.ac.kr
 1995년 연세대학교 전산과학과 졸업(이학사)
 1997년 연세대학교 컴퓨터과학과 졸업(공학석사)
 2001년 연세대학교 컴퓨터과학과 졸업(공학박사)

2001년 National Institute of Standards and Technology(NIST) 객원연구원

2002년~현재 연세대학교 컴퓨터과학과 조교수
 관심분야: 멀티미디어 문서처리, XML, 웹 서비스

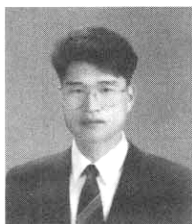


손원성

e-mail : sohnws@cmu.edu
 1998년 동국대학교 컴퓨터공학과(학사)
 2000년 동국대학교 컴퓨터공학과(석사)
 2004년 연세대학교 컴퓨터과학과(박사)
 2004년 4월~8월 University of Washington, Post Doc.

2004년 8월~현재 Carnegie Mellon University, Associate Researcher

관심분야: 웹 Annotation, 스마트 홈, 튜터링 인터페이스



송특섭

e-mail : teukseob@rainbow.yonsei.ac.kr
 2001년 연세대학교 수학과 (이학박사)
 2002년~현재 연세대학교 컴퓨터과학과 박사과정
 관심분야: Transcoding, Annotation, XML, Semantic Web, Cyber Class, 가상환경 네비게이션, Process Annotation



고승규

e-mail : skko@cs.umd.edu
 1992년 연세대학교 컴퓨터과학과(학사)
 1994년 연세대학교 컴퓨터과학과(석사)
 2002년 연세대학교 컴퓨터공학과(박사)
 2002년~2004년 산업기술연구소 객원 연구원

구원

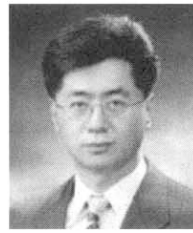
2004년~현재 University of Maryland Associated Researcher
 관심분야: 웹 문서 처리, 데이터베이스



최 윤 철

e-mail : ycchoy@rainbow.yonsei.ac.kr
1973년 서울대학교(학사)
1975년 Univ. of Pittsburgh(석사)
1976년 Univ. of California, Berkeley(석사)
1979년 Univ. of California, Berkeley(박사)
1979년~1982년 Lockheed사 및 Rockwell International사 연구원

1984년~현재 연세대학교 컴퓨터과학과 교수
1990년~1991년 University of Massachusetts 교환교수
2002년~2003년 Keio University 방문교수
관심분야 : 멀티미디어 문서처리 (SGML/XML), GIS, eLearning, 3D User Interface, Avatar Interface, Computer Graphics



임 순 범

e-mail : sblim@sookmyung.ac.kr
1982년 서울대학교 계산통계학과(학사)
1983년 한국과학기술원 전산학과(석사)
1992년 한국과학기술원 전산학과(박사)
1989년~1992년 (주)휴먼컴퓨터 이사/연구소장

1992년~1997년 (주)삼보컴퓨터 부장
1997년~2001년 건국대학교 컴퓨터과학과 교수
2001년~현재 숙명여자대학교 멀티미디어과학과 교수
관심분야 : 컴퓨터 그래픽스, 웹 멀티미디어 응용, 가상환경, 전자출판(폰트, 전자책, 사이버 교재)