

모바일 사용자를 위한 웹 서비스 페이지 개인화 기법

(Personalizing Web Service Pages for Mobile Users)

전 영 호 * 황 인 준 **

(Yeonghyo Jeon) (Eenjun Hwang)

요약 최근 PDA나 차세대 이동통신 단말기와 같은 인터넷 접속이 가능한 무선 단말기들의 보급이 크게 증가하면서 무선 인터넷 접속이 보편화되고 있다. 그러나 대부분의 기존 웹 콘텐츠와 서비스들은 데스크탑 환경에 최적화되어 무선 단말기를 통한 접근이 쉽지 않다. 무선 단말기의 작은 화면이나 제한된 입력장치와 같은 특성을 고려해 볼 때 효과적인 웹 접근을 위한 새로운 기법이 요구되고 있다. 한편, 개인화는 웹 콘텐츠나 서비스를 맞춤형으로 제공하는 데 중요한 역할을 한다. 본 논문에서는 무선 단말기 사용자를 위한 프락시 기반의 웹 서비스 페이지 개인화 기법을 제안한다. 이를 위해 로그 데이터뿐만 아니라 웹 페이지의 서비스 관련 특징을 고려하여 사용자가 자주 이용하는 서비스를 식별한다. 식별된 서비스 페이지들은 무선 단말기에 알맞은 형태로 개인화된 리스트로 제공된다. 결과적으로 사용자들은 여러 제약을 가지는 무선 단말기 상에서 개인화된 웹 서비스를 최소한의 검색만으로 제공받을 수 있다. 마지막으로 프로토타입 시스템에서 다양한 실험을 통하여 기법의 효용성을 입증하였다.

키워드 : 개인화, 무선 인터넷, 콘텐츠 변환

Abstract Recent popularity of web-ready mobile devices such as PDA and cellphone has made mobile Internet access very popular. However, most of existing Web contents and services are optimized for desktop computing environment and not well-suited for mobile devices. Considering different system features of mobile devices such as small display and limited input capability, an alternative scheme to access the Web efficiently is required. On the other hand, personalization plays an important role in the tailored access to the Web contents and services. In this paper, we propose a proxy-based personalizing scheme of Web service pages for mobile users. For that purpose, in addition to log data, service related features of Web pages are considered for the correct identification of popular services. It automatically provides mobile users with the tailored list of Web services well suited for diverse mobile devices. Consequently, mobile users can utilize customized Web services with minimum navigation on the mobile devices with limited capability. In order to show its effectiveness, we have performed several experiments on the prototype system and reported some of the results.

Key words : personalization, wireless Internet, content transcoding

1. 서론

무선 단말기의 수요 증가와 기술발전에 힘입어 무선 인터넷 접속이 폭발적으로 증가하고 있다. 특히 선진국에서는 데스크탑을 이용한 인터넷 연결의 보조적인 수단으로 무선 인터넷 접속이 꾸준히 확산되고 있다. 컴스

코어 네트워크(comScore Networks, Inc.)의 조사에 따르면, 약 1천만명의 미국인들이 휴대폰이나 PDA를 통해 인터넷에 접속하는 것으로 나타났다. 1천910만 명이 PDA를 소유하고 있으며, 그 중 500만 명이 PDA로 인터넷에 접속하고 있다. 또한 휴대폰을 소유한 6천 720만 명의 사용자 중 580만 명이 휴대폰으로 인터넷에 접속하고 있다[1]. 50개국의 인터넷 사용자 11억 2천만 명을 분석한 CIA(Computer Industry Almanac Inc)은 2005년 말까지 상당수의 사용자들이 웹 사용이 가능한 휴대폰이나 PDA 등의 무선 단말기를 통해 온라인에 접속할 것으로 전망했다[2].

* 본 연구는 대학 IT 연구센터 육성·지원사업의 연구결과로 수행되었음

† 비회원 : 삼성전자 무선사업부
yeonghyo.jeon@samsung.com

** 종신회원 : 고려대학교 전자컴퓨터공학과 교수
ehwang04@korea.ac.kr

논문접수 : 2003년 6월 12일

심사완료 : 2004년 11월 10일

그러나 무선 단말기를 통한 인터넷 서비스의 이용은 아직까지 상당히 불편하며 사용 가능한 서비스의 수가 제한적이다. 무선 단말기의 작은 화면과 제한된 입력장치를 고려한 전용 콘텐츠가 각 통신사업자별로 제공되고 있지만 그 내용과 서비스의 수는 극히 빈약하며 제한적이다. 그러므로 기존의 데스크탑을 통해 사용하던 콘텐츠 및 서비스를 무선 단말기를 이용하여 그대로 사용하고자 하는 요구가 증가하고 있다. 이런 요구에 부응하여 근래에 출시되고 있는 무선 단말기들의 하드웨어적 성능은 비약적인 발전을 거듭하고 있다. 예를 들면, 최근 출시된 Dell Axim™ X5와 Compaq iPAQ H3900과 같은 PDA는 기존 제품들보다 더욱 큰 용량의 메모리와 고속 프로세서가 탑재되어 있다. 아울러 인텔, 모토로라, 히타치와 같은 주요 마이크로프로세서 제조업체들이 경쟁적으로 저전력, 고속 프로세서를 개발하고 있다.

하지만 화면 크기와 제한된 입력장치는 무선 단말기의 가장 큰 장점인 휴대성과 상충관계에 있기 때문에 쉽게 개선되지 못하고 있다. 데스크탑에 최적화된 웹 페이지를 무선 단말기 상에서 접근하기 위해서 사용자는 많은 스크롤이 필요하며 탐색구조의 불편함도 감수하여야만 한다. 이런 단점을 보완하여 제한된 무선 환경하에서 웹 콘텐츠를 효율적으로 제공하기 위해서 많은 연구들이 진행되어 왔다. 대표적인 연구로서 Power Browser[3] 나 Digestor[4]는 무선 단말기를 위해 웹 콘텐츠와 구조를 요약하고 재구성한다. 또한 일부 기업[5, 6]들은 기초적인 콘텐츠 변환 서비스를 제공하고 있다. 그러나 변환 과정에서 과도한 오버헤드가 발생하거나 변환이 수동으로 이루어지는 등 크고 작은 문제들을 안고 있으며, 특히 사용자 성향을 고려하지 않은 콘텐츠 변환은 사용자가 꼭 필요로 하는 콘텐츠나 서비스가 제거될 수 있는 단점을 가지고 있다.

콘텐츠 변환과정에 개인화 프로세스를 추가함으로써 위와 같은 단점을 극복함과 동시에 사용자에게 필요한 정보만을 선별하여 제공할 수 있다. 개인화 프로세스는 콘텐츠 변환과정에 사용자의 선호도를 반영함으로써 콘텐츠가 의도와 다르게 제거되는 단점을 극복하고 보다 효율적으로 웹 콘텐츠 및 서비스를 요약하고 재구성할 수 있다.

개인화는 사용자 정보 수집 방법에 따라 크게 묵시적 개인화 방법(implicit personalization)과 명시적 개인화 방법(explicit personalization)의 두 종류로 나눌 수 있다. 전자는 사용자가 제공한 이름, 주소, 전화번호, 선호도, 취미, 나이 등을 이용하는 방법이며, 후자는 사용자와 웹 콘텐츠간의 상호작용 분석 자료를 토대로 개인화를 구현한다[7]. 특히 명시적 개인화 방법은 사용자 데

이타를 로그 분석을 통해 얻어냄으로써 자동화된 개인화 시스템 구현에 주로 사용된다.

그러나 로그 분석만을 기반으로 한 웹 콘텐츠 및 서비스 개인화 시스템들은 다음과 같은 문제점들을 가지고 있다. 첫째, 콘텐츠 저자가 설계한 서비스 탐색 구조를 반영하지 못한다. 개인화 시스템이 재구성한 웹 콘텐츠들은 실제와는 다른 전혀 새로운 탐색 구조를 갖기 때문에 사용자가 기존에 사용하던 콘텐츠의 구조를 일관되게 사용할 수 없다. 둘째, 접근 빈도만을 중점적으로 측정하기 때문에 실제로 사용자가 접근하고자 하는 페이지를 정확하게 제공할 수 없다. 예를 들어 사용자가 $P_1 \rightarrow P_2 \rightarrow P_3$ 순으로 중간 단계의 페이지를 거쳐 목표 페이지에 접근한다면 세 개 페이지의 접근 빈도는 모두 1이 된다. 그러나 P_1 과 P_2 는 실제로 사용자가 원하는 페이지가 아니다. 셋째, 개인화와 관련된 모든 데이터를 실시간으로 처리하기 힘들다. 대용량의 로그를 전처리하여 분석하는 과정은 상당한 시간이 요구된다. 또한 콘텐츠 변환 과정에서 발생하는 오버헤드는 개인화 시스템의 실시간 운영에 커다란 장애 요소이다.

이런 단점을 극복하기 위해 본 논문에서는 서비스 페이지 중심의 개인화 기법을 제안한다. 웹 사이트 내 서비스 페이지는 사용자가 원하는 서비스 직접적으로 가지고 있는 웹 페이지로서 사용자가 서비스를 이용하기 위해 웹 사이트에 접속했다면 링크나 검색을 통한 탐색을 거쳐 서비스 페이지에 도달하게 된다. 본 논문은 이런 페이지가 포함하고 있는 서비스들을 모바일 전용 서비스와 유사한 형태로 제공하는 것에 목적을 두었다. 무선 사용자들은 유선에 비해 상대적으로 비싼 통신비용을 지불하기 때문에 가능한 빨리 서비스를 이용하기를 원한다. 그러므로 모바일 전용 서비스들은 빠른 전송을 위해 텍스트 기반의 메뉴를 사용하고 있으며 각 메뉴 항목은 직접 서비스와 연결되어 있다. 메뉴 기반 인터페이스는 제한된 입력 장치를 이용한 서비스 검색을 용이하게 해준다. 그러므로 개인화 시스템이 웹 콘텐츠와 서비스를 모바일 전용 서비스와 유사한 형태로 제공해 준다면 사용자들은 기존 탐색 구조를 그대로 사용하거나 불필요한 링크탐색이 제거된 형태로 개인화된 웹 콘텐츠와 서비스를 손쉽게 이용할 수 있다.

우리는 사용자들이 가장 자주 접근하는 서비스 페이지들을 제공하기 위해 서비스 지향적 개인화 시스템을 구현하였다. 프락시에 누적되는 접근 로그 분석을 통한 사용자의 접근빈도와 페이지 코드의 서비스 관련 요소 평가를 통해 서비스 페이지만을 추출한 후 이를 기반으로 자동으로 생성된 서비스 페이지 리스트를 텍스트 메뉴 형태로 제공함으로써 사용자가 서비스에 쉽게 접근할 수 있도록 하였다.

본 논문의 구조는 다음과 같다. 2장에서는 개인화 시스템 및 콘텐츠 변환 시스템에 대한 관련 연구에 대해 논의한다. 3장에서는 서비스 페이지 및 서비스 구역을 정의하고 이를 추출하기 위한 서비스 페이지 모델을 제시한다. 4장에서는 본 논문에서 제안된 시스템의 전체적인 구조를 설명하고, 서비스 페이지와 서비스 구역을 추출하기 위한 알고리즘을 살펴본다. 5장에서는 실험결과를 기술하고 마지막 장에서는 결론과 함께 향후 계획을 논한다.

2. 관련 연구

지금까지의 웹 콘텐츠나 서비스의 개인화에 대한 연구들은 주로 로그 데이터 분석에 초점을 두고 있다. PROTEUS[8,9]는 프락시나 웹 서버상의 접근 로그(access log) 분석을 통해 사용자의 패턴과 성향에 따라 웹 페이지를 개인화해서 보여준다. PROTEUS는 접근 로그를 분석하여 사용자 모델을 생성한 후 각 사이트를 기대값(Expected Utility)[8]을 최대로 하는 형태로 변환한다. 하지만 개인화 과정에 작업 부하가 많아 실시간 서비스가 불가능하다. 또한 사용자 단말기의 제약에 대한 협상 절차가 없기 때문에 다양한 무선 단말기에 유연하게 대처할 수 없다. 제안된 기법은 거대한 로그를 한번에 분석하지 않고 시스템의 성능을 고려하여 지정한 시간 단위로 분석하여 데이터를 축적하는 방식을 채용하여 부하를 줄였다. 또한 페이지 접근 빈도 외에도 페이지의 소스코드 분석을 통해 실제 서비스를 제공하는 페이지들을 찾아내어 변환함으로써 콘텐츠 변환 프로세스의 부하를 줄였다. Daily Learner[10]는 내용기반 기계 학습 알고리즘(content-based machine learning algorithm)을 이용하여 데스크탑과 팜(Palm)의 사용자 요구를 분석한다. 이를 기반으로 지능형 뉴스 에이전트가 사용자의 뉴스 관심도를 학습한 뒤 열악한 무선환경(대역폭, 화면크기)에 적합하게 정보의 양을 축소시켜 제공해준다. 하지만 Daily Learner는 뉴스 서비스에만 한정하여 작동한다. iMobile[11]은 프락시 기반의 멀티미디어 서비스를 제공한다. iMobile은 단말기의 성능을 고려한 멀티미디어 파일을 서비스 하기 위해 별도로 정의한 사용자와 단말기 프로파일을 유지하고 관리한다. 본 논문에서는 사용자와 단말기의 프로파일을 현재 표준화 진행 단계에 있는 CC/PP(Composite Capability/Preference Profiles)[12]를 이용하여 기술하였다. WebViews[13]와 WebVCR[14]은 기존 연구와는 달리 로그를 이용하지 않고 구현된 개인화 시스템이다. WebViews는 WebVCR의 확장된 버전으로써 사용자가 원하는 서비스로 이동하기까지 요구되는 정보를 저장함으로써 Smart Bookmark(SMB)를 생성한다. 이를 바탕으

로 콘텐츠를 가공하여 WebView를 만들게 된다. 그러나 Smart Bookmark를 만드는 과정을 WebViews Recorder를 이용하여 사용자가 수동으로 직접해 주어야 하는 단점이 있다. 반면 제안된 시스템은 최초 이용시 간단한 프로파일 제출 외에는 사용자의 부수적인 조작을 최소화하였다.

본 연구에 관련된 또 다른 분야로는 이기종 무선 단말기들을 위한 웹 콘텐츠의 요약과 재구성이 있다. 대부분의 개인화 시스템은 무선 단말기를 위한 웹 콘텐츠 변환 시스템을 포함하고 있다. Power Browser[3]와 Digestor[4]는 웹 프락시(Web proxy)를 이용하여 동적으로 웹 페이지의 요약된 뷰(View)를 생성하여 사용자에게 전달해준다. Web Stream Customizers(WSC)[15]는 서버 측과 클라이언트 상에 동적으로 위치할 수 있는 소프트웨어 모듈로써 웹 콘텐츠를 무선 단말기의 성능이나 대역폭에 따라 변환해주는 미들웨어이다. Transcoding Proxy[16]에서는 각 웹 페이지마다 WYSIWYG 도구를 사용해 주석 처리를 한 후 Transcoding proxy를 이용하여 모바일 디바이스의 작은 화면에 표시할 수 있도록 처리하였다. 결과적으로 보다 정확하게 웹 페이지를 요약할 수 있지만 서비스 제공자가 직접 주석처리를 해야 하는 단점을 가지고 있다. PageMap[17]은 웹 페이지의 구조를 분석하여 유사한 콘텐츠를 가진 요약된 구역으로 나누었다. 나뉜된 구역을 바탕으로 페이지 맵을 생성하여 사용자가 작은 화면에서 손쉽게 페이지의 대부분을 검색할 수 있게 하였다. 본 연구에서는 기존의 페이지 맵 기법을 확장하여 서비스 페이지 중 사용자의 이용빈도가 높은 특정 구역만을 추출하여 전송함으로써 콘텐츠의 변환을 최소화 하였다.

마지막으로 데이터 마이닝의 한 분야인 웹 사용 마이닝(Web usage mining)은 개인화 시스템, 상품 추천 시스템, 웹 사이트 설계 최적화 등을 위해 널리 연구되고 있다. 특히 사용자의 접근패턴 및 성향을 분석하기 위해 웹 서버의 로그를 마이닝하는 기법이 주로 연구 되었다. J. Srivastava et al.[18]은 웹 사용 마이닝을 위한 절차와 데이터 소스, 기존 프로젝트와 제품을 분류하였다. B. Mobasher[19]는 웹 서버 로그로부터 웹 사용 패턴을 마이닝하여 자동화된 개인화 시스템을 구축하는 절차를 설명하고 있으며 사용자가 원하는 새로운 콘텐츠를 추천하는 시스템을 제안하였다. Fenstermacher and Ginsburg[20]는 서버 측의 데이터 수집정책의 부정확성을 밝히며 클라이언트측에서 사용자의 Activity를 마이닝할 수 있는 프레임워크를 제안하였다. S. Dua et al.[21]은 어느 정도 구조화된 웹 서버 로그로부터 사용자의 모든 예측 가능한 접근 시퀀스(access sequence)집합을 추출하는 새로운 로그 분석 알고리즘

을 제안하였다.

3. 서비스 페이지 정의 및 추출기법

이 장에서는 서비스를 직접적으로 제공하는 페이지의 특징과 이를 이용한 서비스 관련 페이지의 추출 기법에 대해 기술한다.

3.1 서비스 페이지 및 구역

현재 웹에서 제공되고 있는 대부분의 서비스는 사용자의 요청에 따라 대화식으로 진행된다. 사용자가 서비스를 제공받기 위해 웹 페이지에 접속하고 필요한 정보를 입력하여 서버로 전송하면 이를 처리하여 결과를 돌려준다. 서비스를 제공하는 웹 페이지는 서버 측 프로세싱에 필요한 정보를 입력 받기 위한 폼이나 직접적인 서비스를 위한 웹 스크립트 등을 다수 포함하고 있다. 우리는 이런 웹 페이지를 서비스 페이지로 정의하였다. 서비스 페이지의 주된 특징을 살펴보면 먼저 사용자로부터 필요한 정보를 얻기 위해 폼을 가지고 있거나 특정 서비스를 직접 제공할 수 있는 웹 스크립트를 구동시키는 링크 등 서비스 제공에 관련된 코드를 가지고 있다. 아울러 이런 요소들은 사용자와의 데이터 교환이나 서버 측의 처리가 필요한 경우가 많다. 그러므로 네트워크 대역폭이 일정하다면 서버의 응답 시간이 길고 전송되는 데이터 양이 다른 페이지에 비해 많게 된다. 또한 요즘 폭발적인 증가추세에 있는 멀티미디어 서비스는 데이터 전송량이 많은 특징이 있다. 이런 특징들을 기반으로 서비스 페이지는 사용자와의 상호작용 정도와 데이터 송수신량에 따라 크게 네 가지 종류로 구분할 수 있다.

많은 상호작용이 요구되며 데이터 전송량이 적은 경우 : 많은 양의 데이터 입력이나 많은 횟수의 조작이 요구되는 페이지이다. 예를 들면 각종 예약 페이지나, 주식, 신용카드, 모바일 뱅킹 등 금융 서비스 페이지들이 있다. 이런 페이지들의 특징은 사용자의 입력을 받기 위한 폼을 많이 가지고 있으며 주고 받는 데이터는 주로 텍스트 형태이므로 양이 적다. 또한 서버 측 프로세싱 시간만큼 응답시간이 지연된다.

많은 상호작용이 요구되며 데이터 전송량이 많은 경우 : 소형 멀티미디어 콘텐츠 다운로드 및 스트리밍 서비스(음악, 뮤직비디오, 영화클립, 스포츠, 뉴스 등)가 이 범주에 속하게 된다. 또한 데이터 및 소프트웨어 구매 서비스(전자책, 자료 등), 온라인 갤러리 등도 유사한 특징을 갖는다고 할 수 있다. 주로 검색을 기반으로 한 멀티미디어 서비스를 제공하는 페이지로써 수신 데이터의 양이 많은 특징을 가진다. 스트리밍 서비스를 위한 미디어 플레이어 임베드 태그나 이미지 태그 등을 많이 가지고 있다.

적은 상호작용이 요구되며 데이터 전송량이 적은

경우 : 단순히 몇 개의 데이터의 입력이나 조작으로 서비스가 가능한 페이지들로서 간단한 검색, 텍스트 위주의 뉴스, 커뮤니티(뉴스그룹) 게시판 등이 있다. 주로 단순히 링크를 클릭함으로써 텍스트를 통한 서비스를 제공하는 곳으로 사용자의 데이터 송수신량이 매우 적고 별다른 서비스 요소들을 가지고 있지 않다. 하지만 자주 업데이트 되는 내용을 서비스 하고 있기 때문에 사용자들의 접속 빈도가 잦은 특징이 있다.

적은 상호작용이 요구되며 데이터 전송량이 많은 경우 : 주로 단순 자료 다운로드 서비스나 검색이 필요 없는 소규모의 VOD 서비스 등이다. 사용자의 접속 빈도가 낮고 폼이나 웹 스크립트는 적게 포함되어 있으나 데이터 송수신량이 큰 특징이 있다. 인터넷 방송이나 웹 자료실 서비스 또는 웹 하드 서비스 등이 있다.

앞서 살펴본 바와 같이 서비스 페이지들은 일반 페이지와 달리 서비스를 제공하기 위한 공통된 코드와 동작 특징들을 가지고 있다. 우리는 서비스 페이지가 가지고 있는 코드나 사용 중에 발생하는 동작 특징들을 서비스 요소로 정의하였다. 입력 및 선택 폼, 미디어 재생기, 웹 스크립트 링크 등의 서비스 요소들은 대부분 전체 서비스 페이지 레이아웃 중에서 일부만을 차지하고 있다. 그러므로 사용자가 특정 서비스를 사용하고자 할 때 전체 페이지 레이아웃 중에서 서비스 요소들이 위치한 특정 부분만이 요구된다. 예를 들면 우리가 웹 검색을 위해 야후나 구글등의 검색 페이지에 접근한다면, 우리는 단지 검색 키워드 입력을 위한 폼만이 필요하다. 다른 모든 서비스 요소들은 불필요하게 된다. 일반적으로 웹 페이지 저자들은 서비스 요소들은 페이지 내 특정 프레임이나 테이블 셀 또는 레이아웃에 위치시킨다. 우리는 이러한 작은 구역을 서비스 구역으로 정의하였다. 서비스 구역은 서비스 페이지내 서비스 요소들이 밀집되어 있는 작은 구역이다. 그러므로 서비스 구역은 간단한 변환을 통해 무선 단말기의 작은 화면에 표시될 수 있으며 낮은 대역폭의 네트워크를 통해서 효율적으로 전달될 수 있다.

3.2 서비스 페이지 모델

제안하는 기법은 사용자의 웹 페이지 접근 빈도와 웹 페이지 및 로그로부터 분석된 서비스 요소들을 평가하여 페이지 및 서비스 구역을 추출하게 된다. 서비스 페이지와 구역 추출에 사용되는 가중치 계산을 위한 서비스 페이지 모델은 다음과 같다.

사용자들의 집합 $U=(u_1, u_2, \dots, u_n)$ 를 가정하자. 각각의 사용자 u_i 는 선호도(D)와 자주 접근한 페이지의 집합(F)으로 구성된다. 여기서 F 는 접근 로그 마이너(Access log miner)에 의해 분석 된다. 선호도(D)는 사용자가 초기에 제출한 값이다.

F 는 오름차순으로 정렬된 자주 접근한 페이지의 집합이다. 각 페이지들은 접근 빈도(f)와 URL(l) 그리고 각 페이지 내에 포함되어 있는 서비스 요소들의 집합(s)으로 구성된다. 서비스 요소의 집합은 웹 페이지를 파싱하여 얻어진다. 각 서비스 요소(e_i)는 종류(c)와 가중치(w)로 구성된다. 서비스 요소와 각 요소에 대한 가중치는 관리자에 의해 선택되고 결정된다. 서비스 요소 집합 내에 가중치의 합은 항상 1이 되어야 하므로 관리자가 특정 서비스 요소의 가중치를 조작하면 다른 서비스 요소의 가중치에 영향을 미치게 된다. 즉 한가지 서비스 요소의 가중치를 증가 시키면 다른 요소들의 가중치는 감소하고, 그 반대도 마찬가지이다. 예를 들어 서비스 요소 e_0 의 가중치가 $w(e_0)$ 에서 $w'(e_0)$ 로 바뀐다면 다른 서비스 요소들의 가중치 $w'(e_i)$ 는 아래 식 (1)에 의해 조절된다.

$$w'(e_i) = \left[\frac{1 - w'(e_0)}{\sum_{e \neq e_0} w(e)} \right] w(e_i) \quad (1)$$

위 식은 모든 서비스 요소의 가중치 합을 1로 유지하며 관리자가 0으로 설정한 서비스 요소의 가중치를 그대로 유지시킨다. 관리자는 서비스 요소의 종류에 대한

가중치 크기의 조절을 통해 특정 요소나 페이지가 선택될 확률을 증가시킬 수 있다. 가중치 계산을 위한 서비스 페이지 모델을 정리하면 표 1과 같다.

사용자가 자주 접근하는 페이지들 중 페이지 내부의 모든 서비스 요소에 대한 가중치의 합을 계산한 후 일정 임계값(threshold)을 넘은 페이지들은 서비스 페이지로 선택된다. 식 (2)과 같이 서비스 페이지(S)는 페이지 정보 p_i 와 페이지 가중치로 구성된다.

$$S = \{ \langle p_i, weight(p_i) \rangle \mid p_i \in F, weight(p_i) \geq \tau \} \quad (2)$$

$$weight(p_i) = \sum_{e=1}^n ew(p_i, e_i) \quad (3)$$

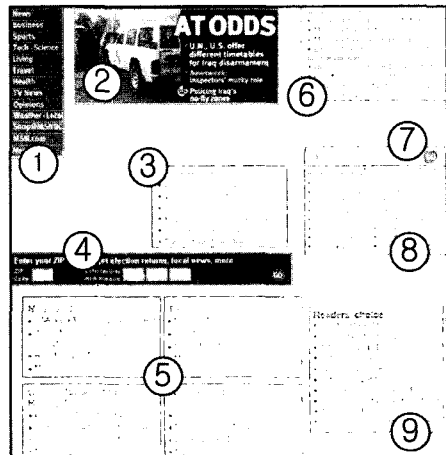
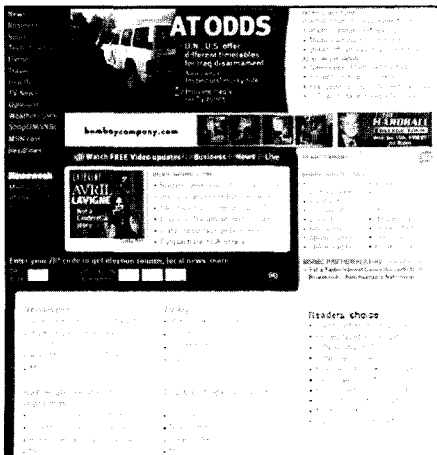
가중치는 식 (3)과 같이 자주 접근되는 페이지 내에 있는 서비스 요소의 가중치 $ew(p_i, e_i)$ 의 합이 되며 τ 는 관리자가 설정할 수 있는 임계값이다.

3.3 서비스 페이지 및 구역 추출 예제

그림 1(a)는 MSNBC 뉴스 서비스 페이지이다. 페이지의 서비스 요소를 살펴보면 하위 뉴스 카테고리를 위한 링크메뉴, 뉴스 동영상이나 하이라이트 뉴스를 볼 수 있는 웹 스크립트 링크, 로컬뉴스와 탐색 서비스를 위한 품이 위치하고 있다. 표 2는 MSNBC 뉴스 페이지에 포함되어 있는 서비스 요소들의 수를 나타낸 것이다. 특히 뉴스 서비스를 위한 ASP(Active Server Page) 링크가

표 1 가중치 계산을 위한 서비스 페이지 모델

$U = \{u_1, u_2, \dots, u_m\}$	사용자 집합
$u_i = \{D, F\}$	각 사용자는 선호도(D)와 자주 접근하는 페이지 집합(F)로 구성된다.
$F = \{p_1, p_2, \dots, p_i\}$	자주 접근하는 페이지 집합내 각 페이지들은 접근 빈도에 따라 정렬된다.
$p_i = \{f, l, s\}$	각 페이지는 접근 빈도, URL 그리고 포함하고 있는 서비스 요소들의 집합으로 구성된다.
$s = \{e_1, e_2, \dots, e_n\}$	서비스 요소의 집합은 각 페이지가 포함하고 있는 n 개 서비스 요소로 구성된다.
$e_j = \{c_j, w_j\}$	각 서비스 요소는 종류(c)와 가중치(w)로 구성된다.



(a) (b)
그림 1 서비스 페이지와 서비스 구역(MSNBC)

표 2 MSNBC 메인 페이지의 서비스 요소 분포

서비스 요소 (Service element)	서비스 요소의 개수 (Number of service elements)
Form	3
ASP	120
JavaScript	7

표 3 각 서비스 구역당 서비스 요소의 분포

(서비스 요소) Service element	서비스 구역(Service region)								
	1	2	3	4	5	6	7	8	9
Form				2			1		
ASP	13	1	6		16	8	1	16	10

많이 존재하고 있으며, 뉴스 검색을 위한 폼이 존재하고 있다. 그림 1(b)는 그림 1(a)에서 추출한 서비스 구역을 나타낸 것이다. 표 2는 각 서비스 구역별 서비스 요소의 분포를 나타내고 있다. 표 3은 구역별로 분포되어 있는 서비스 요소들을 평가한 것이다. 1, 3, 5, 6, 8, 9 번 서비스 구역은 뉴스 및 VOD(Video On Demand) 서비스를 위한 ASP 링크로 구성되어 있고, 4, 7번 서비스 구역은 검색을 위한 폼과 ASP 링크가 사용되었다. 2번은 한 개의 ASP 링크로 구성되어 있지만 사용자의 접근 빈도가 높기 때문에 서비스 요소로 추출되었다.

4. 시스템 구조

시스템의 전체 구조는 그림 2와 같다. 시스템은 다음과 같은 컴포넌트들로 구성된다 : i) 사용자 프로파일 서버(User profile server)는 사용자의 인증과 장치 프

로파일을 관리하며 주기적으로 누적되어 갱신되는 서비스 페이지 및 구역 리스트에 대한 정보를 제공한다. ii) 접근 로그 마이너(Access log miner)는 주기적으로 사용자의 사이트 접속 빈도 및 패턴을 로그로부터 분석하여 자주 접속하는 페이지를 발견한다. 분석주기가 짧을 수록 사용자의 선호도가 빠르게 반영될 수 있다. iii) 서비스 페이지 및 구역 추출 모듈(Service page & region detector)은 접속 빈도와 서비스 요소를 평가하여 사이트내 서비스 페이지를 발견한다. 또한 발견한 페이지를 서비스 구역으로 나눈 후 구역내 링크 클릭빈도와 서비스 요소를 평가하여 사용자가 자주 사용한 서비스 구역 정보를 추출하여 사용자 프로파일 서버에 저장한다. iv) 서비스 페이지 및 구역 캐쉬(Service page & region cache)는 사용자에게 이미 제공되었던 페이지 및 구역을 캐싱함으로써 반복적으로 사용되는 서비스에 대해 응답시간을 줄여준다. v) 콘텐츠 변환 모듈(Contents converter)은 사용자 프로파일 서버로부터 제공 받은 사용자별 서비스 페이지 및 구역 정보를 바탕으로 서비스 페이지로부터 구역을 추출하여 간단한 변환 과정을 거쳐 사용자에게 제공한다.

4.1 사용자 프로파일 서버(User profile server)

웹 개인화에 있어 접속한 사용자를 식별하는 문제는 큰 이슈중의 하나이다. 특히 패턴을 분석하기 위한 서버 모니터링은 캐쉬에 영향을 받는 로그 데이터의 부정확성으로 인해 사용자의 식별이 어렵다. 보통 IP 주소가 간편하고 일반적인 해결책으로 사용되지만, 이 방법은 심각한 단점을 안고 있다. 대부분의 ISP(Internet Service Provider)는 IP 주소를 효율적으로 이용하기

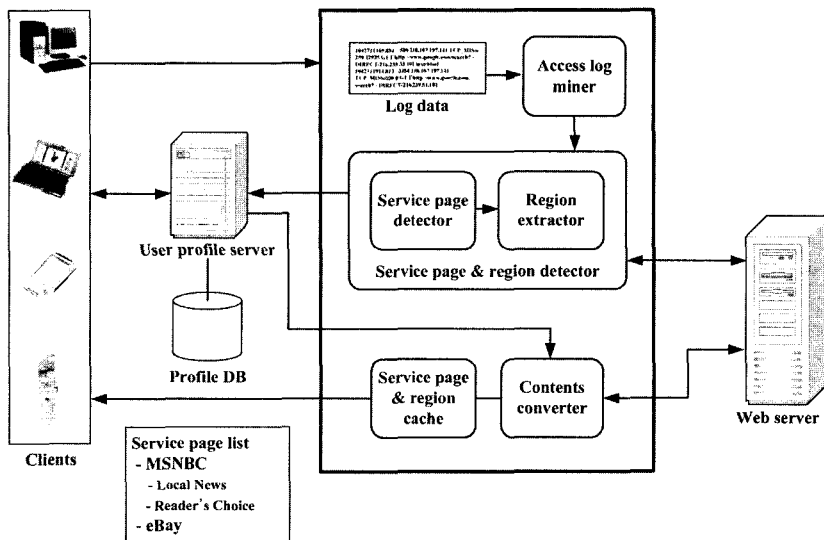


그림 2 전체 시스템 구조

위해 DHCP(Dynamic Host Configuration Protocol)를 사용하고 있다. 그러므로 각 사용자마다 IP 주소가 임의로 할당된다. 결과적으로 서로 다른 사용자에게 같은 IP 주소가 할당될 수 있다. 이는 심각한 로그 데이터의 부정확성을 야기시킨다. 본 논문에서는 사용자를 정확하게 식별하기 위해 프락시와 함께 동작할 수 있는 사용자 프로파일 서버를 사용하였다. 사용자 프로파일 서버는 프락시의 작업부하를 줄이고 사용자를 정확하게 식별할 수 있다. 또한 프로파일 서버는 사용자의 성향과 디바이스에 대한 프로파일 및 서비스 페이지와 서비스 구역정보를 관리한다. 사용자는 제한된 시스템의 사용을 위해서 사용자 프로파일 서버에 접속하여 기본적인 인증 정보와 소유하고 있는 무선 단말기에 대한 간단한 정보를 제공하게 된다.

그림 3은 사용자 프로파일 전송을 위한 간단한 입력 폼을 나타내고 있다. 사용자 프로파일은 아이디, 패스워드 등의 인증정보, 사용하는 무선 단말기의 하드웨어 및 소프트웨어 프로파일, 사용자의 선호옵션 등을 포함하고 있으며 사용자 프로파일 서버에 저장되게 된다. 추가적으로 프로파일 서버에는 서비스 페이지 및 구역 분석 모듈로부터 분석된 사용자 서비스 선호도 데이터(랭크, 페이지 및 구역 정보)가 저장된다. 인증 정보는 사용자가 어떠한 인터넷 디바이스로 접속해 들어오더라도 사용자를 고유하게 식별하며 개인화된 서비스를 제공하는 데 사용된다. 디바이스 프로파일은 콘텐츠 변환 과정에

서 사용된다.

입력된 사용자 프로파일은 CC/PP를 이용하여 프로파일 서버로 전송된다. CC/PP는 Content Negotiation에 관련된 표준으로 W3C(World Wide Web Consortium)[22]에서 표준화 과정에 있다. CC/PP는 XML 기반의 메타 정보를 교환하기 위한 언어인 RDF(Resource Description Framework)[23]를 사용하여 디바이스의 프로파일 및 사용자의 성향(Preference)을 HTTP Request와 함께 보내는 방법이다. CC/PP를 이용하여 프로파일 서버에 디바이스와 사용자의 성향을 저장하게 된다. 본 논문에서 우리는 제정한 RDF schema를 사용하였다. 그림 4는 제안된 시스템에서 사용된 CC/PP 프로파일의 일부인 디바이스 부분을 나타내고 있다. ①은 사용자가 프로파일을 입력하지 않았을 때 사용될 기본 프로파일을 나타내고 있다. ②는 사용자가 실제로 입력한 내용 중 무선 단말기의 사양을 나타내는 디바이스 프로파일을 나타내고 있다.

4.2 서비스 페이지 및 구역추출(Service page and region detection)

서비스 페이지 및 서비스 구역을 찾기 위한 과정은 크게 두 단계로 나누어진다. 첫번째 단계에서는 로그 정제와 분석을 통해 자주 접근하는 페이지를 찾는다. 두번째 단계에서는 자주 접속했던 페이지들 중에서 서비스 페이지와 서비스 구역을 추출하게 된다. 그림 5는 자주 접속한 사이트와 페이지를 찾아내기 위한 알고리즘을

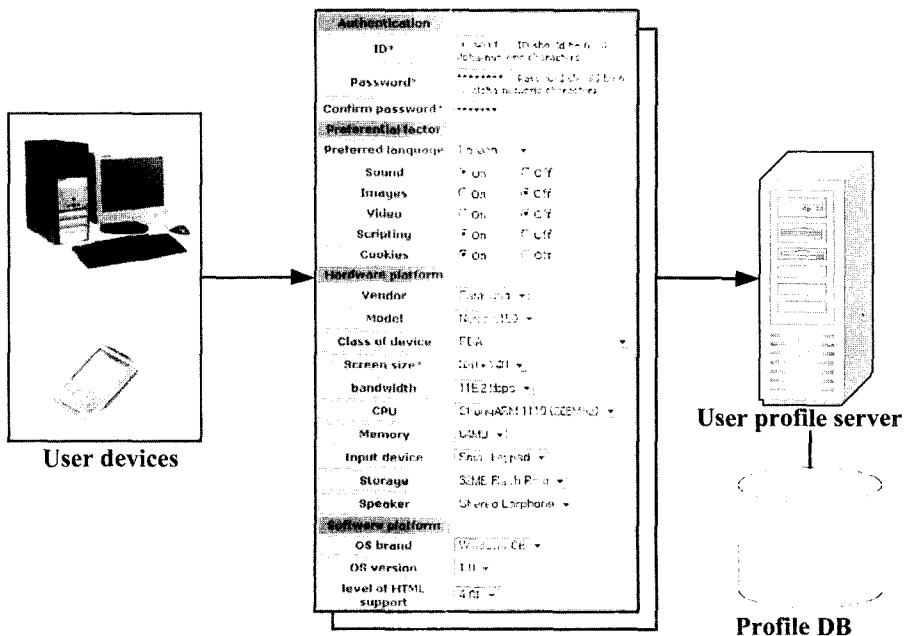


그림 3 간단한 사용자 프로파일 입력 폼

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ccpp="http://www.w3.org/2002/11/08-ccpp#"
  xmlns:prf="http://adtl.ajou.ac.kr/Schema#">
  . . . . .
  <rdf:Description rdf:about="http://adtl.ajou.ac.kr/MyProfile">
    <ccpp:component>
      <rdf:Description rdf:about="http://adtl.ajou.ac.kr/TerminalHardware">
        <rdf:type rdf:resource="http://adtl.ajou.ac.kr/Schema#HardwarePlatform"/>
        <ccpp:defaults> ----- ①
          <rdf:Description rdf:about="http://adtl.ajou.ac.kr/HWDefault">
            <rdf:type rdf:resource="http://adtl.ajou.ac.kr/Schema#HardwarePlatform"/>
            <prf:cpu>PPC</prf:cpu>
            <prf:displayWidth>320</prf:displayWidth>
            <prf:displayHeight>200</prf:displayHeight>
          </rdf:Description>
        </ccpp:defaults>
        <prf:vendor>Samsung</prf:vendor> -----②
        <prf:model>nexio s150</prf:model>
        <prf:classofdevice>PDA</prf:classofdevice>
        <prf:displayHeight>320</prf:displayHeight>
        <prf:displayWidth>240</prf:displayWidth>
        <prf:bandwidth>115.2</prf:bandwidth>
        <prf:cpu>strongARM 1110</prf:cpu>
        <prf:memory>64</prf:memory>
        <prf:input>small keypad</prf:input>
        <prf:storage>32</prf:storage>
        <prf:speaker>yes</prf:speaker>
      </rdf:Description>
    </ccpp:component>
    .
    .
    .
```

그림 4 CC/PP 프로파일 예제(일부분)

```
Function FindFrequentlyAccessedPage()
ft = frequency threshold value
T = time interval
fap[][] = array for frequently accessed pages
load changed log since the last running time
clean loaded log
repeat
  foreach user begin
    find frequently accessed sites
    foreach site begin
      find frequently accessed page
      if ft < frequency then
        insert page URL and frequency into fap[][]
    end
  end
  call FindServicePage(fap[][])
sleep(T)
until dearon stop
```

그림 5 접근 빈도가 높은 페이지의 추출(알고리즘 I)

```
Function FindServicePage(fap[][]))
k = sum of weight for each element
kt= threshold value of k
ri[] = region data
fin = 1
repeat
  foreach pages in fap[][] begin
    download page from web server
    calculate k
    if k > kt then {
      ri[] = call pagemap()
      estimate service elements for each region and rank region
    }
  end
  call SavetoProfile()
  fin = 0
until fin=0
```

그림 6 서비스 페이지 추출(알고리즘 II)

나타내고 있다.

개인화된 서비스를 제공하기 위해서 먼저 사용자가 자주 접속하는 사이트와 페이지가 분석되어야만 한다. 특히 접근 로그는 로그 마이너에 의한 작업부하를 줄이기 위하여 시스템의 성능과 사용자들의 접속량을 고려하여 관리자가 설정한 주기 동안 추가적으로 누적된 부분만 분석되며 각 사용자별로 분석된다. 사용자별로 분석된 페이지 정보는 접근 빈도에 따라 배열에 저장된다.

그림 6은 서비스 페이지 추출을 위한 알고리즘을 나타내고 있다. 서비스 페이지 및 구역 분석 모듈은 배열에 저장된 URL을 바탕으로 웹 서버로부터 사용자가 자주 접근했던 페이지를 다운로드 받는다. 다운로드 받은

```
Function ServiceList()
foreach login begin
  generate service and region information from profile DB
  provide service list to client
  if request then
    call ContentConverter()
  end
end

Function ContentConverter()
foreach request begin
  if request service page then
    call pagemap()
  else if request service region then
    extract region from service page by region information
    send to client with content conversion by device profile
  end
end
```

그림 7 서비스 리스트 생성과 콘텐츠 변환(알고리즘 III)

웹 페이지로부터 서비스 요소들의 가중치를 평가한 후 임계값과의 비교를 통해 서비스 페이지를 찾아낸다. 이

과정에서 서비스 요소들의 가중치 및 임계값의 조절을 통해 특정 요소나 페이지가 선택될 확률을 높일 수 있다. 또한 임계값이 넘는 서비스 페이지로부터 페이지 맵 알고리즘[17]을 사용하여 서비스 구역을 추출하게 된다. 페이지 맵 알고리즘은 테이블 태그의 정보로부터 구역을 추출한다. 추출된 구역의 서비스 요소를 평가하여 최종 서비스 구역을 결정하게 된다. 마지막으로 추출된 구역은 접근 빈도와 가중치에 따라 랭크된다.

4.3 콘텐츠 변환

무선 단말기 사용자가 제안된 시스템에 접속하게 되면 사용자 프로필 서버는 개인화 서비스 리스트를 제공한다. 리스트는 트리 형태로 구성된다. 사용자가 서비스 구역 링크를 선택하게 되면 프로필 서버가 해당 서비스 페이지 및 구역정보를 콘텐츠 변환기에 전송하며 콘텐츠 변환기는 이정보를 이용하여 서비스 페이지로부터 서비스 구역을 추출하여 간단한 변환 과정만을 거쳐 사용자에게 전송하게 된다. 그림 8은 서비스 리스트와 선택된 구역을 보여준다. 그림 8의 (a)는 서비스 페이지 및 구역 리스트를 (b)는 CNN 비디오 서비스 페이지 중 정치 뉴스 비디오 서비스 구역을 나타내고 있다. 정치 뉴스 서비스 구역은 크기가 작으므로 최소한의 변환 과정만을 거쳐 작은 화면에 표시할 수 있다.

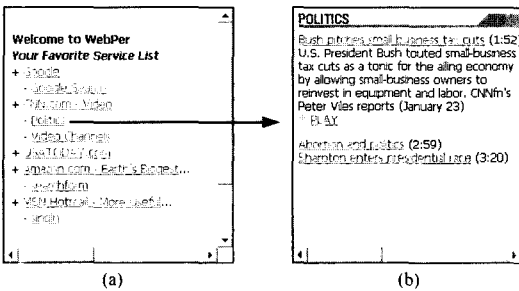


그림 8 서비스 리스트와 선택된 구역

5. 실험 결과

이 장에서는 제안한 시스템의 효율성을 확인하기 위해 수행한 실험의 환경과 결과에 대해 기술한다. 우리는 시스템의 프로토타입을 리눅스 기반의 Squid 웹 프락시[24]를 이용하여 구현하였다. Squid는 유닉스 계열에서 구동되는 오픈 소스 웹 프락시 캐쉬로서 연구목적으로 소스의 수정이 자유롭다. 또한 다양한 로그를 제공하며 분석 또한 용이하다. 사용자 프로필 데이터베이스로는 eXcelon[25]을 사용하였다. eXcelon은 XML 문서를 표준 XML 형태로 저장, 검색 가능한 XML 데이터베이스이며 CC/PP를 이용한 사용자의 인증 정보와 프로필 및 서비스 페이지, 구역정보 관리를 위해 사용되었다.

정확한 로그를 생성하기 위하여 우리는 사용자의 모든 요청이 제안된 시스템을 통해 이루어지도록 설정하였고, 웹 브라우저의 캐쉬 설정을 조정하여 영향을 최소화하였다. 일반적으로 무선 단말기에 사용되는 캐쉬는 로그파일에 미치는 영향이 적기 때문에 본 실험에서는 별도로 고려하지 않았다. 결과적으로 사용자가 접근한 거의 모든 페이지와 링크가 로그에 기록되었다. 사용자가 서비스를 이용하기 위해 웹에 접근한다는 가정하에 브라우저의 Forward and Back 버튼으로 인한 링크의 재구성은 하지 않았다. 로그를 분석하기에 앞서 우리는 ICQ, MSN 등의 메신저 프로그램과 애드웨어(광고를 보는 대신 무료로 쓰는 소프트웨어)가 생성하는 인터넷 접근을 제거하였다. 이는 최대한 사용자의 의도에 의해 생성된 접근 정보만을 추출하기 위해서이다.

우리는 20명의 사용자에게 다음과 같은 설문문을 조사하였다. 먼저 사용자가 가장 자주 접속하는 웹 사이트 20개를 조사하였다. 그런 후 첫 번째 질문에 대답한 각 사이트 내에서 가장 많이 이용하는 인터넷 서비스를 하나씩 총 20개를 조사하였다. 설문 조사 후 사용자들에게 PC와 무선 단말기를 이용해 7일 동안 제안된 시스템의 프로토타입을 거쳐 자유롭게 인터넷 서비스를 이용하게 하여 기본 로그 데이터를 수집하였다. 사용자들이 인터넷을 사용하는 동안 제안된 시스템은 지속적으로 서비스 요소를 평가하여 서비스 페이지 및 구역을 추출하였다.

우리는 제안된 시스템의 효율성을 검증하기 위하여 재현율(recall)과 정확율(precision)[26]을 계산하였다. 재현율은 추출된 페이지들 중 사용자의 설문과 관련된 페이지의 비율이며 정확율은 관련된 페이지들 중 실제로 추출된 페이지의 비율이다. 재현율과 정확율을 계산하기 위해 다음과 같은 변수가 사용되었다.

각 사용자에 대해

- (1) 모든 페이지(All pages) : 제안된 시스템이 추출한 모든 서비스 페이지
- (2) 관련 페이지(Relevant pages) : 사용자가 설문에 대답한 서비스 페이지. 단, 우리는 항목 (1)에 포함되지 않은 페이지를 제거하였다. 이러한 페이지는 사용자의 잘못된 대답이나 기준에 사용자에게 인기가 있었지만 인기가 없어진 페이지 때문에 주로 발생한다. 예를 들면 사용자는 물건을 고르고 사기 위해 온라인 쇼핑 사이트를 자주 접속한다. 하지만 일단 물건을 구입한 뒤에는 일정기간 동안 쇼핑 사이트에 접속하지 않게 된다.
- (3) 추출된 페이지(Retrieved pages) : 제안된 시스템이 추출한 사용자가 자주 접속한 서비스 페이지. 우리는 추출한 서비스 페이지의 수를 1에서 20까지 5단위로 증가시켰다. 이는 여러 종류의 무선 단말기에

서 화면 스크롤 없이 한 화면상에서의 적중률(hit rate)을 시뮬레이션 하기 위한 것이다. 서비스 리스트에서 한 개 서비스 페이지는 링크를 가지고 있는 한 줄의 텍스트로 표현된다. 일반적으로 셀 폰은 5-10줄을 표시할 수 있으며 스마트 폰은 15줄 정도를 표시할 수 있다. PDA는 내장된 브라우저를 이용하여 약 20줄까지 표시할 수 있다. 그러므로 우리는 무선 단말기의 다양한 화면 크기를 고려하여 4개의 precision을 계산하였다.

- (4) 추출된 페이지 중 관련 페이지(Relevant and retrieved pages) : 사용자가 질문에 대답한 서비스 페이지중 제안된 시스템이 추출한 페이지와 적중된 개수이다.

위에서 언급한 변수들을 정리하면 표 4와 같다.

표 4 재현율, 정확율 변수

P	시스템이 추출한 n 개 페이지
$Pq \in P$	질문(질의 q)에 의해 사용자로부터 추출된 관련 페이지
추출된 페이지	시스템에 의해 추출된 사용자가 자주 접속한 페이지 리스트 $\{p_1, p_2, \dots, p_k\}$
k	k 는 추출된 페이지 수이며 1부터 20까지 5단위로 증가

위 데이터를 이용하여 우리는 추출된 페이지가 적중한 경우를 1로 그렇지 않은 경우를 0으로하여 관련 리스트($r_1, r_2, r_3, \dots, r_n$)를 계산하였다. $k \geq 1$ 일 때 재현율, 정확율, 평균 정확율은 다음과 같이 각각 정의된다.

$$recall(k) = \frac{1}{|P_q|} \sum_{1 \leq i \leq k} r_i$$

$$precision(k) = \frac{1}{k} \sum_{1 \leq i \leq k} r_i$$

$$average\ precision = \frac{1}{|P_q|} \sum_{1 \leq k \leq |P|} r_k \times precision(k)$$

그림 9는 20명의 사용자가 질문 대답한 서비스 페이지에 대한 제안된 시스템이 추출한 페이지 수를 5 단위로 증가시켜감에 따른 재현율과 정확율을 백분율로 나타내고 있다. 예를 들면, 총 20개의 서비스 페이지를 추출했을 때 정확율은 약 64%이고 재현율은 약 76%였다. 이 정확율은 약 20줄의 텍스트를 표시할 수 있는 240×320의 해상도를 가진 PDA에 적용할 수 있다. 그림 10은 5단위 추출 페이지 수에 따른 재현율-정확율 곡선을 나타내고 있다.

다음으로 우리는 서비스 페이지 전체의 용량과 구역의 용량을 비교함으로써 제안된 시스템을 이용하여 얻을 수 있는 대역폭의 절감에 대해 측정하였다. 우리는 제안된 시스템에 의해 추출된 서비스 페이지 중 페이지

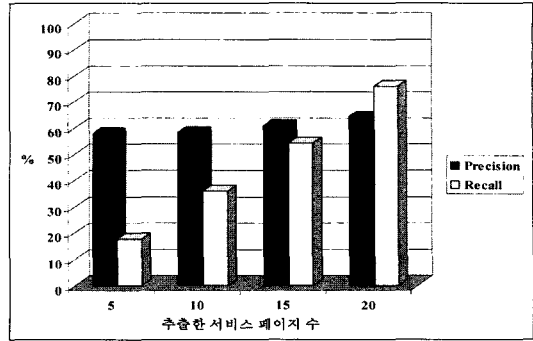


그림 9 추출 페이지 수 증가에 따른 정확율

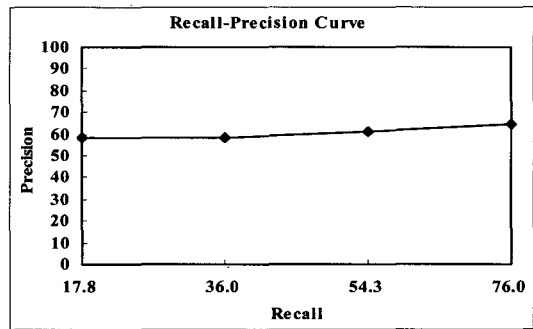


그림 10 재현율-정확율 곡선

특성을 고려하여 10개의 서비스 페이지와 구역을 추출하였다. 표 5의 (a)는 구역이 추출됨으로써 얻을 수 있는 대역폭 절감을 보여준다. 구역만을 전송했을 때 서비스 페이지 전체 크기 대비 평균 데이터 전송량을 약 90%까지 감소시킬 수 있었다. 즉 서비스 페이지 전체를 모두 전송하지 않고 사용자가 자주 사용하는 구역만을 전송함으로써 무선 단말기의 낮은 대역폭 제한을 대폭 완화할 수 있었다. 표 5의 (b)는 추출된 서비스 구역의 크기를 나타내고 있다. 데스크탑에 최적화된 서비스 페이지 크기에 비해 구역은 전체 페이지 대비 약 11%의 크기를 가짐으로써 일반적으로 128*128에서 240*320의 화면크기를 갖는 무선 단말기의 작은 화면에도 최소한의 스크롤을 통해 표시될 수 있다.

6. 결론

무선 인터넷 사용의 증가에 따라 데스크탑에 비해 상대적으로 기능이 제약적인 무선 단말기를 통한 웹 콘텐츠 접근에 대한 대안이 요구되는 실정이다. 또한 방대한 정보의 양 때문에 개인에게 맞춤형 정보를 제공하기 위한 웹 개인화가 제안되었다. 개인화 시스템은 무선 단말기의 하드웨어적 측면과 소프트웨어적 측면을 통합적으로 개인화할 수 있다. 개인화 시스템을 사용함으로써

표 5 대역폭과 디스플레이 크기 비교
(a) 대역폭

페이지 특성	서비스페이지(구역)	구역 용량 (byte)(A)	서비스페이지 용량(byte)(B)	절약율 (%) ((B-A)/B)×100
단순한 폼 입력 또는 링크 클릭 기반 서비스	WhatIs.com (search)	571	169676	99.66
	Google (search)	1213	12578	90.36
	USATODAY, sport baseball (team index)	3633	155313	97.66
	ebay, PDAs/Handheld PC (PDA search)	2863	332547	99.14
	Amazon (product search)	2797	121339	97.69
	Hotmail (login)	3974	20438	80.56
복잡한 폼 입력 또는 멀티미디어 서비스	CNN Video (politics)	2517	217404	98.84
	MTV.com (music video)	22057	217778	89.87
	National Geographic Online (Photo of the day)	36399	101097	64.00
	Travelocity (Find me the best priced trip)	17627	149512	88.21
평균 대역폭 절약율				90.60

(b) 디스플레이 크기

페이지 특성	서비스페이지(구역)	서비스 페이지 디스플레이 크기 (pixels)(A)	구역 디스 플레이크기 (pixels)(B)	서비스페이지 대비구역크기(%) (B/A)×100
단순한 폼 입력 또는 링크 클릭 기반 서비스	WhatIs.com (search)	800×1024(↑)	332×23	0.93
	Google (search)	523×427	419×79	14.82
	USATODAY, sport baseball (team index)	770×1024(↑)	141×117	2.09
	ebay, PDAs/Handheld PC (PDA search)	1024×1024(↑)	163×287	4.46
	Amazon (product search)	770×1024(↑)	171×75	1.63
	Hotmail (login)	649×637	316×205	15.67
복잡한 폼 입력 또는 멀티미디어 서비스	CNN Video (politics)	770×1024(↑)	307×331	12.89
	MTV.com (music video)	764×1024(↑)	376×456	21.92
	National Geographic Online (Photo of the day)	728×1024(↑)	475×310	19.75
	Travelocity (Find me the best priced trip)	778×1024(↑)	478 ×326	19.56
서비스페이지 대비 평균 구역크기				11.37

무선 단말기의 제약을 완화함과 동시에 사용자가 관심 있는 웹 콘텐츠 및 서비스를 더욱 쉽게 접근할 수 있다.

본 논문에서 우리는 서비스 페이지에 초점을 둔 새로운 개인화 시스템을 제안하였다. 제안된 시스템은 로그 뿐만 아니라 페이지내 서비스 요소들을 평가하여 자주 사용되는 서비스 페이지들을 개인화한다. 결과적으로 사용자들은 모바일 전용 서비스와 유사한 형태로 재구성된 웹 서비스를 제공받을 수 있다. 또한 콘텐츠 변환의 부하를 줄일 수 있고 구역추출을 이용하여 무선 단말기의 작은 화면의 단점을 극복할 수 있다.

참 고 문 헌

[1] comScore Networks, Inc. - <http://www.comscore.com>
 [2] Computer Industry Almanac Inc. - <http://www.c-i-a.com>
 [3] O. Buyukkocuten, H. Garcia Molina, A. Paepcke and T. Winograd, "Power Browser : Efficient Web Browsing for PDAs," Proceedings of the Conference on Human Factors in Computing Systems, pp. 430-437, 2000.

[4] T. Bickmore and B. Schilit, "Digester: Device-independent Access to the World Wide Web," Proceedings of the 6th International World Wide Web Conference, pp. 655-663, 1999.
 [5] AvantGo, Inc. - <http://www.avantgo.com>
 [6] Earthlink, Inc. - <http://www.earthlink.net>
 [7] C. Sharma, Wireless Internet Enterprise Applications A Willey Tech Brief, John Wiley & Sons, Inc, pp.91-93, 2001.
 [8] C. R. Anderson, P. Domingos and D. S. Weld, "Personalizing Web Sites for Mobile Users," Proceedings of the 10th World Wide Web Conference, 2001.
 [9] C. R. Anderson, P. Domingos and D. S. Weld, "Adaptive web navigation for wireless devices," Proceedings of the 17th International Joint Conference on Artificial Intelligence, 2001.
 [10] D. Billsus, M. J. Pazzani and J. Chen, "A learning agent for wireless news access," Proceedings of the 2000 International Conference on Intelligent User Interfaces, pp. 33-36, 2000.
 [11] C. Yih-Farn, H. Huale, R. Jana, S. John, S. Jora, A. Reibman and W. Bin, "Personalized multimedia

services using a mobile service platform," Wireless Communications and Networking Conference, 2002.

- [12] <http://www.w3.org/Mobile/CCPP/> - CC/PP (Composite Capabilities /Preference Profiles).
- [13] J. Freire, B. Kumar and D. F. Lieuwen, "Web-Views : accessing personalized web content and services," Proceedings of the 10th International Conference on World Wide Web, pp. 576-586, 2001.
- [14] V. Anupam, J. Freire, B. Kumar and D. Lieuwen, "Automating Web navigation with the WebVCR," Computer Networks, Vol 33(1-6), pp. 503-517, 2000.
- [15] J. Steinberg and J. Pasquale, "A web middleware architecture for dynamic customization of content for wireless clients," Proceedings of the 11th International Conference on World Wide Web, Honolulu, Hawaii, USA, 2002.
- [16] M. Hori, G. Kondoh, K. Ono, S. Hirose and S. Singhal, "Annotation-based web content transcoding," Proceeding of 9th International World Wide Web Conference, pp. 197-211, 2000.
- [17] D. Song and E. Hwang. "PageMap: Summarizing Web Pages for Small Display Devices," Proceeding of International Conference on Internet Computing, Vol. 2, pp. 506-512, June 2002.
- [18] J. Srivastava, R. Cooley, M. Deshpande and Tan, P.-N. "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data," SIGKDD Explorations, Vol. 1(2), pp. 12-23, January 2000.
- [19] B. Mobasher, R. Cooley and J. Srivastava. "Automatic Personalization Based on Web Usage Mining," Communications of the ACM, Vol. 43(8), pp. 142-151, August 2000.
- [20] K.D. Fenstermacher and M. Ginsburg. "Mining client-side activity for personalization," Advanced Issues of E-Commerce and Web-Based Information Systems (WECWIS 2002), Proceedings of 4th IEEE International Workshop, pp. 205-212, 2002.
- [21] S. Dua, E. Cho and S.S. Iyengar. "Discovery of Web frequent patterns and user characteristics from Web access logs: a framework for dynamic Web personalization," Application-Specific Systems and Software Engineering Technology, 2000. Proceedings of 3rd IEEE Symposium, pp. 3-8, 2000.
- [22] W3C (World Wide Web Consortium)-<http://www.w3.org>
- [23] <http://www.w3.org/RDF/> - RDF
- [24] Squid Web Proxy Cache-<http://www.squid-cache.org/>
- [25] eXcelon - <http://www.exln.com/>
- [26] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, pp. 428-433, 2001.



전 영 효

1994년~2001년 아주대학교 정보통신대학 정보 및 컴퓨터 공학 학사. 2002년~2004년 아주대학교 정보통신전문대학원 정보 및 컴퓨터 공학 석사(멀티미디어 데이터베이스). 2004년~현재 삼성전자 무선사업부 근무. 관심분야는 모바일 어플리케이션, 모바일 콘텐츠 가공, 멀티미디어 검색



황 인 준

1988년 서울대학교 컴퓨터공학과(학사)
1990년 서울대학교 컴퓨터공학과(석사)
1998년 Univ. of Maryland at College Park 전산학과(박사). 1998년 8월~1999년 8월 Bowie State Univ., Assistant Professor. 1999년 9월~2003년 2월 아주대학교 정보통신전문대학원 조교수. 2003년 3월~2004년 8월 아주대학교 정보통신 전문대학원 부교수. 2004년 9월~현재 고려대학교 전자컴퓨터 공학과 조교수. 관심분야는 데이터베이스, 멀티미디어 시스템, 정보 통합, 전자 상거래, XML 응용, 유비쿼터스 컴퓨팅