

모바일 사용자를 위한 웹 서비스 페이지 개인화 기법

전영효^o 황인준

아주대학교 정보통신전문대학원 정보통신공학과
{jeonyh^o, ehwang}@ajou.ac.kr

WebPer: Personalizing Web service pages for mobile users

Yeonghyo Jeon^o Eenjun Hwang

The Graduate School of Information and Communication, Ajou University

요 약

인터넷 접속이 가능한 무선 단말기들의 보급이 크게 증가하면서 무선 인터넷 사용자의 수가 지속적으로 증가하고 있다. 그러나 대부분의 웹 콘텐츠와 서비스들은 데스크톱 환경에 최적화되어 무선 단말기를 통한 접근이 쉽지 않다. 이런 문제의 해결책으로 개인화를 통해 웹 콘텐츠 및 서비스를 무선 환경에 알맞게 재구성하는 방법은 웹 접근을 위한 새로운 대안으로 제시되고 있다.

본 논문에서 우리는 기존의 개인화 시스템과 달리 웹 서비스 페이지를 중점적으로 개인화하여 제공하는 기법을 제안한다. WebPer는 로그 데이터뿐만 아니라 웹 페이지 코드내 서비스 관련 특징의 평가를 통해 사용자가 자주 접속한 서비스 페이지들의 리스트를 제공한다. 사용자들은 서비스 리스트를 통해 무선 단말기의 각종 제약에 극복함과 동시에 개인화된 서비스를 제공받을 수 있다. 우리는 제안된 기법을 기반으로 프로토타입 시스템을 구현 하였으며, 다 수의 서비스 페이지를 대상으로 무선 환경에서의 서비스 개인화 정도 및 전송 데이터의 효율을 측정하였다.

1. 서 론

무선 단말기의 수요 증가와 기술 발전에 힘입어 무선 인터넷 접속이 폭발적으로 증가하고 있다. 그러나 무선 단말기를 이용한 웹 접속은 아직까지 상당히 불편하며 제한적이다. 최근 출시된 Dell AximTM X5와 Compaq iPAQ H3900과 같은 PDA는 기존 제품들보다 더욱 많은 메모리와 강력한 프로세서를 탑재하고 있다. 하지만 제한된 화면 크기와 입력 장치는 무선 단말기의 가장 큰 장점인 휴대성과 상충관계에 있기 때문에 쉽게 개선되지 못하고 있다. 이런 단점을 보완하여 제한된 무선 환경 하에서 웹 콘텐츠를 제공하기 위해 많은 연구들이 진행되어 왔다. 그러나 콘텐츠 변환 과정에서 과도한 부하가 발생하거나 변환이 수동으로 이루어지는 등 크고 작은 단점들을 안고 있으며, 특히 사용자가 꼭 필요로 하는 콘텐츠 및 서비스가 제거될 수 있는 약점을 가지고 있다.

콘텐츠 변환 과정에 개인화 프로세스를 추가함으로써 위와 같은 단점을 극복함과 동시에 사용자의 선호도에 따라 효율적으로 웹 콘텐츠 및 서비스를 요약하고 재구성할 수 있다. 기존 연구들은 로그 분석을 통해 사용자 데이터를 얻어냄으로써 자동화된 개인화 시스템을 구현하였다. 그러나 로그 분석만을 기반으로 한 개인화 시스템은 다음과 같은 문제점들을 가지고 있다. 첫째, 탐색하기 어려운 구조를 갖는다. 개인화 시스템이 재구성한 웹 콘텐츠들은 기존 콘텐츠와 다른 새로운 탐색구조를 갖기 때문에 사용자가 기존에 사용하던 콘텐츠의 구조를 일관되게 사용할 수 없다. 둘째, 단순히 접근 빈도만을 측정하기 때문에 실제로 사용자가 접근하고자 하는 페이지를 정확하게 제공할 수 없다. 예를 들어 사용자가 P₁ → P₂ → P₃ 순으로 중간의 페이지를 거쳐 목표 페이지에 접근한다면 세 개의 페이지의 접근 빈도는 모두 1이 된다. 그러나 P₁과 P₂는 실제로 사용자가 원하는 페이지가 아니다. 셋째, 개인화와 관련된 모든 데이터를 실시간으로 처리하기가 힘들다. 대용량의 로그를 전처리하여 분석하는 과정은 상당한 시간이 요구된다. 또한 콘텐츠 변환 과정에서 발생하는 과부하는 개인화 시스템의 실시간 운영에 커다란 장애 요소이다.

이런 단점을 극복하기 위해 본 논문에서는 웹 서비스 페이지

중심의 개인화 기법을 제안한다. 웹 서비스 페이지는 사용자가 원하는 서비스를 직접적으로 제공하는 페이지로써 폼과 스크립트 등 서비스와 관련된 특징을 가지고 있다. 본 논문에서는 모바일 전용 서비스와 유사하게 자동으로 생성된 서비스 페이지 리스트를 제공함으로써 사용자가 서비스에 쉽게 접근할 수 있도록 하였다. 또한 액세스 로그 분석뿐만 아니라 웹 페이지 소스의 서비스 관련된 특징을 분석함으로써 서비스 페이지를 정확히 발견할 수 있도록 하였다.

본 논문의 구조는 다음과 같다. 2장에서는 개인화 시스템에 대한 관련 연구에 대해 논의한다. 3장에서는 서비스 페이지 및 서비스 구역을 정의하고 예제를 제시한다. 4장에서는 본 논문에서 제안하는 시스템의 전체적인 구조를 설명하고, 5장에서는 실험 결과를 기술한다. 6장에서는 결론과 함께 향후 계획을 제시한다.

2. 관련 연구

지금까지의 웹 콘텐츠 및 서비스의 개인화에 대한 연구들은 주로 로그 데이터 분석에 초점을 두고 있다. PROTEUS[1]는 프락시(Proxy)나 웹 서버상의 액세스 로그 분석을 통해 사용자의 패턴과 성향에 따라 웹 페이지를 개인화해서 보여준다. 하지만 개인화 과정에 작업부하가 많아 실시간 서비스가 불가능하다. iMobile[2]은 프락시 기반의 멀티미디어 서비스를 제공한다. iMobile은 단말기 성능에 알맞은 멀티미디어 파일을 서비스하기 위해 사용자 및 단말기의 프로파일을 유지 관리한다. WebVCR[3]과 WebViews[4]는 기존 연구와는 달리 로그를 이용하지 않고 구현된 개인화 시스템이다. WebViews는 사용자가 원하는 서비스로 이동하기까지 요구되는 정보를 저장함으로써 Smart Bookmark(SMB)를 생성한다. 이를 바탕으로 콘텐츠를 가공하여 사용자 뷰(View)를 만들게 된다. 그러나 Smart Bookmark를 만드는 과정을 WebViews Recorder를 이용하여 사용자가 직접 해주어야 하는 단점이 있다.

대부분의 개인화 시스템은 무선 단말기를 위한 웹 콘텐츠 변환 시스템을 포함하고 있다. Power Browser[5]와 Digestor[6]는 웹 프락시를 이용하여 동적으로 웹 페이지의 요약된 뷰를 생성

하여 사용자에게 전달해준다. Transcoding proxy[7]에서는 각 웹 페이지마다 WYSIWYG 도구를 사용해 주석 처리를 하고 Transcoding proxy를 이용하여 무선 단말기의 작은 화면에 표시할 수 있도록 처리하였다. 결과적으로 보다 정확하게 웹 페이지를 요약할 수 있지만 서비스 제공자가 직접 주석처리를 해야 하는 단점을 가지고 있다. PageMap[8]은 웹 페이지의 구조를 분석하여 유사한 콘텐츠를 가진 요약된 구역으로 나누었다. 나뉜 구역을 바탕으로 페이지 맵을 생성하여 사용자가 작은 화면에서 손쉽게 페이지의 대부분을 검색할 수 있게 하였다.

3. 개인화를 위한 서비스 페이지 발견

현재 웹에서 제공되고 있는 대부분 서비스는 사용자의 요청에 따라 대화식으로 진행된다. 서비스를 제공하는 웹 페이지는 서버측 프로세싱을 필요로 하는 경우가 많다. 우리는 이런 웹 페이지를 서비스 페이지로 정의하였다. 서비스 페이지의 특징을 살펴보면 사용자와의 데이터 교환을 위해 폼을 가지고 있거나 웹 스크립트를 구동시키는 링크 등 서비스를 제공하기 위한 관련 요소를 가지고 있다. 또한 데이터 교환과 서버측 프로세싱은 서버의 응답 시간을 길게 만들고 전송되는 데이터양을 다른 페이지들에 비해 증가시킨다. 특히 요즘 폭발적으로 증가하는 멀티미디어 서비스는 데이터 전송량이 많은 특징이 있다. 우리는 서비스 페이지가 가지고 있는 서비스 관련 소스 코드나 사용 중에 발견할 수 있는 특징들을 페이지 내 서비스 요소로 정의하였다.

검색 폼, 미디어 재생기, 뉴스 섹션 등의 서비스 요소들은 대부분 전체 서비스 페이지 중에서 일부만을 점유하고 있다. 그러므로 사용자가 특정 서비스를 사용하고자 할 때 전체 서비스 페이지 중에서 서비스 요소들이 위치한 특정 부분만이 요구된다. 예를 들면 우리가 웹 검색을 위해 야후나 구글 등의 검색 페이지에 접근한다면, 우리는 단지 검색 키워드 입력을 위한 폼만이 필요하다. 다른 모든 서비스 요소들을 불필요하게 된다. 일반적으로 웹 페이지 저자들은 서비스 요소들은 페이지내 특정 프레임이나 셀 내에 위치시킨다. 우리는 이러한 작은 구역을 서비스 구역으로 정의하였다. 서비스 구역은 간단한 변환 과정을 거쳐 낮은 대역폭의 네트워크를 통해서 효율적으로 전달될 수 있다.

그림 1-(a)은 MSNBC 뉴스 서비스 페이지이다. 페이지의 서비스 요소를 살펴보면 하위 뉴스 카테고리들 위한 링크메뉴, 뉴스 동영상이나 하이라이트 뉴스를 볼 수 있는 웹 스크립트 링크, 로컬뉴스와 탐색 서비스를 위한 폼이 위치하고 있다. 그림 1-(b)은 그림 1-(a)에서 추출한 서비스 구역을 나타낸 것이다.

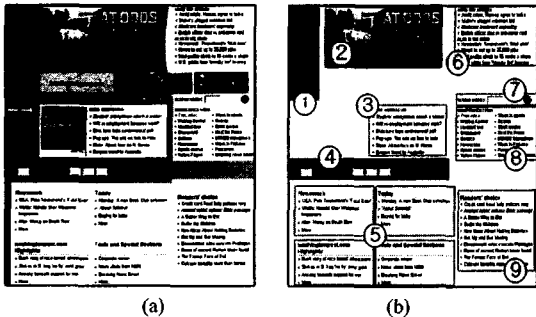


그림 1. 서비스 페이지와 서비스 구역(MSNBC)

4. WebPer 시스템 구조

WebPer 시스템의 전체 구조는 그림 2와 같다. WebPer는 다음과 같은 컴포넌트들로 구성되어진다. i) 사용자 프로파일 서

버는 사용자의 인증과 장치 프로파일을 관리하며 주기적으로 갱신되는 서비스 페이지 및 구역 리스트에 대한 정보를 제공한다. ii) 로그 마이너는 주기적으로 사용자의 사이트 접속 빈도 및 패턴을 분석하여 자주 접속하는 페이지를 발견한다. iii) 서비스 페이지 및 구역 분석 모듈은 접속 빈도와 서비스 요소를 평가하여 사이트내 서비스 페이지를 발견한다. 발견한 페이지를 서비스 구역으로 나눈 후 구역 내 링크 클릭 빈도와 서비스 요소를 평가하여 사용자가 자주 사용한 서비스 구역 정보를 사용자 프로파일 서버에 저장한다. iv) 서비스 페이지 및 구역 캐시는 사용자에게 이미 제공되었던 페이지 및 구역을 캐싱함으로써 반복적으로 사용되는 서비스에 대해 응답시간을 높인다. v) 콘텐츠 변환 모듈은 서비스 페이지 및 구역 발견 모듈이 추출한 정보를 바탕으로 서비스 페이지로부터 구역을 추출하여 간단한 변환 과정을 거쳐 사용자에게 제공한다.

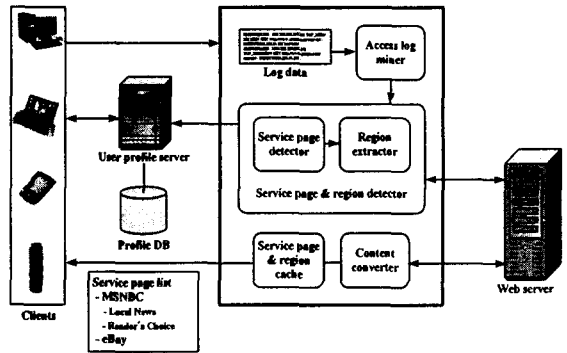


그림 2. WebPer 시스템 구조

5. 실험 결과

WebPer 시스템은 리눅스 기반의 squid web proxy[7]를 이용하여 구현되었다. Squid는 유닉스 계열에서 구동되는 오픈소스 웹 프락시 캐시로서 연구 목적으로 소스의 수정이 자유롭다. 또한 다양한 로그를 제공하며 분석 또한 용이하다. 사용자 프로파일 데이터베이스로는 eXcelon[8]을 사용하였다. eXcelon은 XML 문서를 표준 XML 형태로 저장, 검색 가능한 XML 데이터베이스이며 사용자의 인증 정보와 프로파일 관리를 위해 사용되었다. 정확한 로그를 생성하기 위하여 우리는 사용자의 모든 요청이 WebPer를 통해 이루어지도록 설정하였고, 브라우저의 캐시의 영향을 최소화하였다. 일반적으로 무선 단말기의 캐시는 영향이 미약하므로 본 실험에서는 무시하였다. 결과적으로 사용자가 접근한 모든 페이지와 링크가 로그에 기록되도록 하였다. 사용자가 서비스를 이용하기 위해 웹에 접근한다는 가정 하에 브라우저의 forward and back 버튼으로 인한 링크의 재구성은 하지 않았다. 로그를 분석하기에 앞서 우리는 ICQ, MSN 등의 메신저와 adware의 성격을 가진 프로그램이 생성한 로그 정보를 제거하였다. 이는 최대한 사용자의 의도에 의해 생성된 로그 정보만을 추출하기 위해서이다.

우리는 20명의 사용자에게 가장 자주 접속하는 웹 사이트와 그 사이트 내에서 가장 많이 이용한 서비스를 각각 10개씩 설문조사한 후 사용자들에게 PC와 무선 단말기를 이용해 7일 동안 WebPer를 거쳐 자유롭게 인터넷 서비스를 이용하게 하여 기본 로그 데이터를 수집하였다. WebPer는 사용자가 인터넷을 사용하기 시작한 순간부터 지속적으로 서비스 요소를 평가하여 서비스 페이지 및 구역을 추출하였다.

우리는 WebPer 시스템의 효율성을 검증하기 위하여 정확률(precision)과 재현율(recall)[9]을 계산하였다. 각 사용자에 대해 사용된 변수는 다음과 같다.

1) All pages : WebPer가 추출한 모든 서비스 페이지.
 2) Relevant pages: 사용자가 질문에 대답한 서비스 페이지. 단, 우리는 1)에 포함되지 않은 페이지를 제거하였다. 이러한 페이지는 사용자의 잘못된 대답이나 이전에 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 일정 기간동안 쇼핑 사이트에 접속하지 않게 된다.

3) Retrieved pages: WebPer가 추출한 사용자가 자주 접속한 서비스 페이지. 우리는 추출한 서비스 페이지의 수를 5에서 20까지 5단위로 증가시켰다. 이는 무선단말기에서 화면 스크롤 없이 한 화면상에서의 적중률을 측정하기 위한 것이다. 서비스 리스트에서 한 개 서비스 페이지는 링크를 가지고 있는 한 줄의 텍스트로 표현된다. 일반적으로 셀 폰은 5-10줄을 스마트폰은 15줄 정도를 표시할 수 있으며 PDA는 내장된 브라우저를 이용하여 약 20줄까지 표시할 수 있다. 그러므로 우리는 무선단말기의 다양한 화면 크기를 고려하여 4개의 정확도를 계산하였다.

4) Relevant and retrieved pages: 사용자의 질문 응답에도 포함되고 WebPer가 추출한 페이지들에도 포함되는 페이지.

그림 3은 20명의 사용자가 질문 대답한 서비스 페이지에 대한 WebPer가 추출한 페이지의 정확도를 나타내고 있다. 총 20개의 서비스 페이지를 추출했을 때 평균 정확률은 64.25%이고 재현율은 76.02%였다. 이 정확률은 320×240의 해상도를 가진 PDA에 적용할 수 있다.

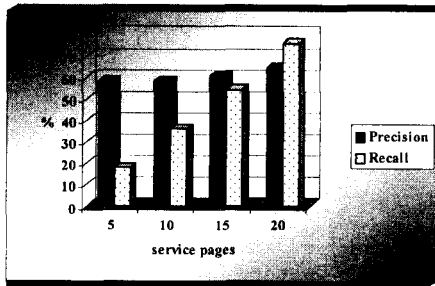


그림 3. 평균 정확도

다음으로 우리는 서비스 페이지의 용량과 구역의 용량을 비교함으로써 WebPer를 이용하여 얻을 수 있는 대역폭의 절감에 대해 측정하였다. 우리는 relevant page를 중 무작위로 6개의 서비스 페이지와 구역을 추출하였다. 표 1의 (a)는 구역이 적중(hit)함으로써 얻을 수 있는 대역폭 절감이다. 평균 데이터 전송량은 94.04%까지 감소하였다. 그러므로 구역만을 전송함으로써 무선 단말기들의 낮은 대역폭 제한을 대폭 완화할 수 있다. 표 1의 (b)는 구역의 크기를 나타내고 있다. 서비스 페이지 크기에 비해 구역은 평균 8.59%의 크기를 가짐으로써 무선 단말기의 작은 화면에도 거의 변환 없이 표시될 수 있다.

표 1. 대역폭과 디스플레이 크기 비교

(a) 대역폭

Service page (region)	Region volume(byte)	Service page volume(byte)	Saving (%)
Google (search)	1213	12578	90.36
CNN Video (politics)	2517	217404	98.84
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

(b) 디스플레이 크기

Service page(region)	Service Page display size (pixels)	Region display size (pixels)	Saving (%)
Google (search)	523×427	419×79	85.18
CNN Video (politics)	770×1024(↑)	307×331	87.11
USATODAY, sport baseball (team index)	770×1024(↑)	141×117	97.91
ebay, PDAs/Handheld PC (PDA search)	1024×1024(↑)	163×287	95.54
Amazon (product search)	770×1024(↑)	171×75	98.37
Hotmail (login)	649×637	316×205	84.33

6. 결 론

무선 인터넷 사용이 증가함에 따라 데스크톱에 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 시스템을 사용함으로써 무선 단말기의 제약을 완화함과 동시에 사용자가 관심 있는 웹 콘텐츠 및 서비스를 더욱더 쉽게 접근할 수 있다.

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

7. 참고문헌

- [1] C. R. Anderson, P. Domingos and D. S. Weld, "Personalizing Web Sites for Mobile Users," Proceedings of the 10th World Wide Web Conference, 2001.
- [2] 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.
- [3] 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.
- [4] J. Freire, B. Kumar and D. F. Lieuwen, "WebViews: accessing personalized web content and services," Proceedings of the 10th International Conference on World Wide Web, pp. 576 - 586, 2001.
- [5] O. Buyukkokten, 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.
- [6] 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.
- [7] Squid Web Proxy Cache - <http://www.squid-cache.org/>
- [8] eXcelon - <http://www.exln.com/>
- [9] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers, pp. 428-433, 2001.