

유비쿼터스 환경을 위한 RSS 뉴스 채널 콘텐츠의 개인화 모바일 서비스 기법

한 승 현[†] · 류 동 엽^{††} · 임 영 환^{†††}

요 약

무선단말기는 유비쿼터스 환경을 위한 최적의 장치지만 무선단말기를 이용한 인터넷 서비스는 데스크 탑 환경에 비해 제한적이다. 따라서 본 연구에서는 콘텐츠 기반의 개인화를 이용한 무선 인터넷 서비스 방법에 대해 제안한다. 기존의 웹 사이트의 뉴스기사 검색에서 제공하는 RSS와 연계된 웹 콘텐츠에서 빠르게 데이터를 획득하고, 개인화 기법을 적용하여 수많은 뉴스 채널에서 빠르고 쉽게 선호하는 콘텐츠를 이용할 수 있다. 제안 방법을 사용함으로써 콘텐츠 제작비용을 줄일 수 있으며 사용자들의 선호도에 대응하는 개인화된 모바일 웹 뉴스 콘텐츠를 제공할 수 있다.

키워드 : 콘텐츠, 모바일, 개인화, 실시간, 트랜스코딩, 유비쿼터스

A Personalized Mobile Service Method of RSS News Channel Contents for Ubiquitous Environment

Seung-Hyun Han[†] · Dong-Yeop-Ryu^{††} · Young-Hwan Lim^{†††}

ABSTRACT

Although wireless devices are the most suitable device for ubiquitous environment, they have restrictive capacities when using internet services than desktop environments. Therefore this research proposes a wireless internet service method that uses contents-based personalization. The existing websites can easily and promptly access desired news articles and other data through RSS-linked web contents and by the personalization method. The proposed method will make using wireless internet easier while lowering contents production costs. Moreover, personalized mobile web news contents that satisfy the preferences of users can be serviced.

Key Words : Contents, Mobile, Personalization, Realtime, Transcoding, Ubiquitous

1. 서 론

인터넷 서비스의 가장 큰 비중을 차지하고 있는 것은 웹 서비스이다. 대다수의 인터넷 사용자들은 자신이 원하는 정보를 이용하기 위해 수많은 페이지들을 직접 방문하고 검색함으로써 원하는 정보를 이용한다. 하지만 무선단말기에 탑재된 브라우저의 제약과 낮은 하드웨어 성능, 불편한 인터페이스를 지닌 무선단말기에서 위와 같은 접근 방식은 이용의 불편함은 물론 비용과 시간적 측면에서도 비효율적이다.

디지털 컨버전스 환경의 OSMU(One Source Multi-Use)를 위해 기존 웹 페이지를 모바일 웹 페이지로 변환하는 다양한 방법들이 연구되었다[1, 2]. 이러한 방법들은 비교적 신

속하게 페이지를 생성 할 수 있지만 다양한 단말기에 최적화된 서비스나 사용자 취향에 대한 고려가 부족하고 많은 수의 페이지가 생성된다는 단점을 지닌다. 따라서 이러한 단점을 보완하기 위해서는 다음과 같은 문제를 고려해야한다.

첫째, 사용자의 특성과 선호도에 대한 고려가 필요하다. 즉, 사용자의 선호도를 고려한 무선 웹 서비스가 제공되어야 비용의 낭비를 막을 수 있으며 사용자의 만족도를 높일 수 있다.

둘째, 다양한 무선 단말기에 최적화된 무선페이지를 서비스 할 수 있어야한다. 크기와 성능이 다른 무선단말기에 적합한 무선페이지를 서비스해야 특정 단말기에 종속적이지 않은 서비스가 가능하다.

셋째, 원본 웹 콘텐츠의 내용이 변경되었을 때, 무선 단말기에서도 변경된 내용이 반영되어야한다. 수시로 변경되는 웹 콘텐츠의 내용을 콘텐츠 제공자의 수작업에 의지한다면 시간과 비용면에서 비효율적이다.

※ 본 연구는 서울시 산학연 협력사업(10581cooperateOrg93112) 지원으로 수행되었음.

† 준 회원 : 송실대학교 대학원 컴퓨터학과 박사과정

†† 정 회원 : 송실대학교 대학원 컴퓨터학과 공학박사

††† 종신회원 : 송실대학교 미디어학부 교수

논문접수 : 2007년 3월 19일, 심사완료 : 2007년 5월 29일

본 연구는 기존 웹 콘텐츠의 정보를 이용해 무선 웹 페이지로 서비스할 때 사용자의 선호키워드 및 연관키워드들을 분류하여 모바일 웹 콘텐츠를 생성한다. 그리고 생성된 콘텐츠의 적합성, 콘텐츠 선호도를 조사하여 최종적으로 개인화 웹 콘텐츠를 생성하는 방법을 제안한다. 또한 능동적 채널 생성이 가능한 RSS(RDF Site Summary)를 이용함으로써 쉽고 빠르게 콘텐츠를 생성할 수 있다. RSS는 웹 사이트를 통한 정보 교환 커뮤니티의 표준으로 자리 잡고 있으며 성공적인 XML서비스 모델로 대표되고 있다. ABC, BBC, NewYork Times, Yahoo, CNET, MSDN 등을 비롯한 국내 인터넷 포털, 신문사, 방송국 등이 대부분 RSS 서비스를 제공하고 있어 그 확장범위는 더욱 넓어지고 있다.

본 논문의 구성은 다음과 같다. 2장에서는 모바일 웹콘텐츠 변환 및 개인화기법에 관련된 연구들을 소개한다. 3장에서는 키워드 분석을 통한 개인화된 콘텐츠의 생성 방법과 콘텐츠 적합성 등 개인화를 위한 적용방법을 설명한다. 4장에서는 실험 결과 및 성능 평가에 대해 설명하고 마지막으로 5장에서는 결론을 기술한다.

2. 관련연구

PC 환경보다 제약이 많은 모바일 환경을 위한 콘텐츠의 생성과 개인화 기법에 대한 연구들이 활발히 진행되어 왔다. Power Browser[3]는 기존의 웹 콘텐츠 구조를 보다 단순하게 요약하고 재구성하는 방법을 제공한다. 특히, 특정 페이지를 링크한 페이지의 수가 많은 경우 상위의 콘텐츠로 배치하는 방법을 적용하고 있다.

웹 콘텐츠를 모바일용으로 수동 변환하여 제공하는 경우 [4, 5]는 변환에 소요되는 오버헤드가 높은 비중을 차지하기 때문에 실시간 정보전달이 어렵다. 위에 열거된 방식들[3, 4, 5] 모두 콘텐츠 변환에 있어 개별 사용자의 성향을 고려하지 않는다.

모바일 사용자들 위한 웹 페이지의 개인화 기법[6]의 경우 페이지를 웹 페이지 접근 빈도와 로그로부터 분석된 서비스 요소들 및 가중치를 기준으로 서비스 구역으로 나누어 자주 사용되는 페이지들을 개인화하고 구역추출을 이용한 무선 단말기의 작은 화면의 단점을 보완하고자 하였다. MyNews[7]는 사용자가 의견을 반영할 수 있는 어노테이션(annotation)을 기준으로 웹 페이지를 선택, 그룹, 삭제의 부분들로 분리하고 선택, 그룹 부분에 대해서만 사용자의 선호도 설정을 반영하여 변환한다. 위 연구[6, 7]의 경우 웹 페이지를 구역화해서 변환하는 방법으로 주제별 분류에는 용이하나 키워드와 같은 사용자의 직접적인 관심요소를 기준으로 웹 페이지 내용을 변환하기엔 적절하지 못하다.

개인화 기법에 사용되는 기본 데이터는 사용자가 제공한 이름, 주소, 나이, 선호도, 취미 등의 명시적(explicit)데이터와 사용자와 웹 서버간의 상호 흐름을 분석하여 얻은 묵시적(implicit)데이터를 사용할 수 있으며 이러한 데이터들의 분석은 주로 데이터베이스나 로그 파일을 통해 이루어지고 있다.

데이터들의 분석 방법은 연관규칙(Association Rule)을 이용한 추론[8]이나 사례기반 추론(Case based Reasoning, CBR)[9] 등의 방법이 있으며 후자는 특정 사용자와 상관없는 다른 사용자들의 사례를 근거로 한 방법이며, 전자는 사례와 일치하는 규칙을 발견하여 추론하는 방법이지만 규칙을 발견하지 못하는 경우 추론이 불가능하다는 단점을 지닌다.

Agrawal et al.이 1993년 처음 소개한 항목집합으로 표현된 트랜잭션에서 각 항목간의 연관성을 반영하는 규칙[8]은 데이터베이스에서 사용자가 지정하는 정도의 통계적 수치를 만족하는 규칙을 발견하는데 사용된다. 이는 지지도(support)와 신뢰도(confidence)라는 통계적 척도를 적절하게 이용하여 데이터 사이의 연관성을 발견한다. 그러나 이러한 추론 시스템을 도입하는 것은 사례가 충분히 쌓인 경우에만 가능하며 실시간으로 콘텐츠를 제공해야 하는 시스템의 경우 새로운 등장 요소에 대한 규칙을 발견하고 분석하여 추론을 적용하는데 소모되는 시간비용은 커다란 위험요소이다. 또한 단순 로그 분석을 통한 방법은 다음과 같은 문제를 가진다[6].

첫째, 원본 웹 콘텐츠와 개인화된 콘텐츠의 경우 서로 상이한 탐색구조를 가질 수 있다. 둘째, 페이지 접근에 대한 빈도수를 주로 측정하기 때문에 실제 사용자가 원하는 콘텐츠인지 정확히 파악하기 힘들다. 셋째, 개인화와 관련하여 축적된 대용량 로그 데이터를 복잡한 전처리 과정을 거쳐 분석하게 되면 실시간 처리 시스템에 도입하기 어렵다. 위와 같은 제약사항을 해결위해 본 연구에서 제안하는 방법은 다음과 같다.

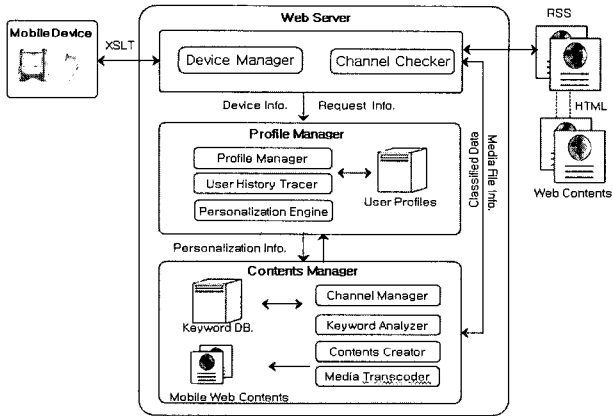
첫째, 원본 웹 콘텐츠와 변환 생성된 개인화된 모바일 웹 콘텐츠는 동일한 구조로 탐색이 가능하며 새로운 키워드 접근을 위해 접근 단계를 축소한 탐색 방법이 제공된다. 하지만 그로 인해 콘텐츠가 속한 분류 자체에 영향을 미치지 않도록 설계하였다. 둘째, 콘텐츠의 생성과 선호도 계산에 단순 페이지 접근 빈도수가 아닌 콘텐츠 내의 키워드 접근 빈도와 연관 키워드 분석을 통한 방법을 제공한다. 셋째, 대용량 로그를 분석하지 않고 키워드 분석을 통한 빠른 전처리 과정을 도입하여 실시간 콘텐츠 생성에 적용하였다.

3. 시스템 구성

3.1 시스템의 처리구조

본 연구에서 제안하는 시스템의 구조는 (그림 1)과 같다. 무선단말기가 서비스를 요청하면 Device Manager는 프로토콜 헤더정보를 분석해 접속한 브라우저 종류, 단말기 종류, 지원되는 이미지와 사운드 형식 등의 기기정보를 파악한다. 그리고 이 정보는 Profile Manager를 통해 사용자 프로파일에 기록된다.

Profile Manager는 사용자의 브라우징 과정에서 요청 키워드, 연관 키워드, 선호 장르에 대한 데이터를 획득하여 사용자 프로파일에 추가한다. 그리고 이 데이터들은 Personalization

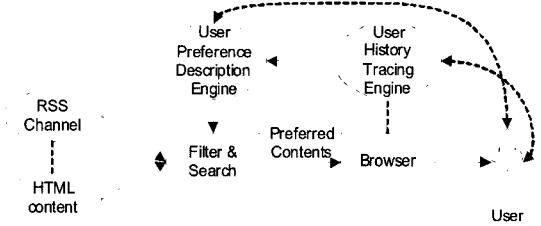


(그림 1) 시스템 구조도

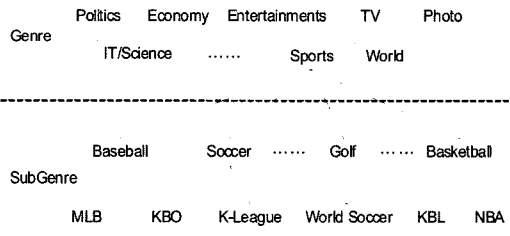
Engine을 통해 개인화적용에 사용된다. Contents Manager의 Contents Creator는 RSS채널의 각 아이템에 기록된 데이터에서 필요한 부분을 획득한다. 그리고 링크된 HTML 페이지로부터 풀 텍스트 피드(Full-text feeds)를 추출해서 무선 웹 페이지 생성에 사용할 XML데이터를 생성한다. 또한 풀 텍스트 피드와 관련된 이미지를 Media Transcoder가 처리할 수 있도록 미디어 파일의 정보 및 원본 데이터를 넘겨준다. Media Transcoder는 웹에서 사용하는 이미지 형식들(BMP, JPG, GIF, PNG)을 모바일 기기에 적합한 사이즈, 퀄리티, 포맷 등에 대해 변환하는 모듈이다. 마지막으로 XSLT를 통해 모바일 기기에 적합하게 정의된 XML+XSL이 XHTML과 WML2.0형식으로 서비스된다.

3.2 개인화 기법의 적용

본 연구에서는 페이지 접근 빈도수 측정이 아닌 보다 상세한 범위의 관심 대상에 대한 정보를 획득한다. 키워드 획득에 의한 채널을 생성함으로써 보다 세밀한 개인화 영역에 대한 무선페이지를 생성할 수 있다. 또한 사용자 프로파일에 기록된 사용자의 선호 장르 및 키워드 등은 지속적으로 변화가 생길 수 있기 때문에 사용자의 브라우징 히스토리의 추적을 통해 추가로 데이터 획득하여 재처리하는 과정을 거치도록 하였다. (그림 2)는 사용자의 선호도 데이터와 브라우징 히스토리 추적을 통한 사용자 프로파일로의 반영 과정이다.



(그림 2) 사용자 선호도 반영의 흐름



(그림 3) 장르에 대한 구분

사용자 프로파일의 선호도 정보의 기술은 <표 1>과 같은 구조의 정보들로 분류된 데이터들을 기준으로 분류하였다. 또한 이러한 분류 정보는 개인화 처리기법의 적용에 대한 사용자 만족도 측정에도 사용하였다.

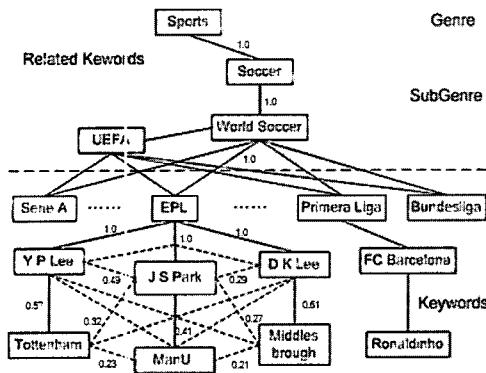
본 연구에서는 뉴스 기사 중심의 RSS를 이용하였기 때문에 (그림 3)처럼 장르를 분류하였다. 이러한 분류를 통해 선호하는 장르 및 서브장르가 사용자 프로파일에 구성된다.

(그림 4)는 장르 및 서브 장르명과 각 장르에 등장하는 키워드간의 연관성을 보여주고 있다. 그림 상에 표현된 키워드간의 연관계수는 <표-2>에 묘사된 아래 수식(2)를 통해 구해진 연관계수의 누적평균값이다.

<표 2>는 사용자의 임의의 키워드의 요청으로 생성된 채널에서 임의로 추출된 20개의 샘플 콘텐츠 내의 풀 텍스트 피드에서 각 키워드의 출현 빈도수, 키워드가 등장한 콘텐츠의 수, 사용자의 선호 키워드에 대한 정보, 항목 그룹간의 상관관계를 의미하는 CORREL (correlation), 연관계수를 구하기 위해 사용한 수식(2)로 얻은 값, 채널 그룹이 여러 개인 경우 수식(2)값의 누적 평균, 수식(4)로 계산된 콘텐츠 선호도를 보여주고 있다. 그룹 항목간의 상관관계 계산에 사용되는 correlation 계산식은 다음 수식(1)과 같다.

<표 1> 선호도 정보 기술을 위한 기본 분류

Channel_maintenance_Preferences	Concerned Preferences		
	Keyword_Preferences	Classification_Preferences	View_Preferences
Publication_Source Channel_Creation_Keyword Channel_Connection_Info. Channel_Update_Info. Renewal_Items Deleted_Items	Subject_Keyword Relational_Keyword Content_Exactness Content_Preferences Date_Period Image_Existence	1 st _Genre 2 nd _Genre 1 st _SubGenre 2 nd _SubGenre	Personalization_appliance Image_Existence Feeds_Length



(그림 4) 키워드 분류와 키워드 간 연관계수

상관관계가 높으면 1에 가깝고 상관관계가 반대의 경향을 띠면 -1에 가깝게 나타난다.

$$\rho_{x,y} = \frac{Cov(X, Y)}{\sigma_x \sigma_y} \quad (1)$$

상관관계식을 적용한 결과를 보면 두 그룹 속성 간 표준 편차가 0인 경우는 부정(不定)형 에러가 발생하며 또한 음수값의 상관계수가 등장하고 또한 수치 편차가 높지 않아 연관성 계산을 위한 수식에 적용하기 어렵다. 따라서 본 연구에서는 키워드간의 연관성 조사 및 생성된 콘텐츠의 적합성, 생성된 콘텐츠의 선호도 조사를 위해 다음과 같은 수식과 알고리즘을 사용한다.

수식(2)은 협업적 필터링(Collaborative Filtering) [10]에서 자주 사용하는 항목과 항목간의 관련성에 기반한 항목기반 추천기법 계산식 중 하나인 Conditional Probability based Similarity 수식[11]을 본 연구의 키워드 항목간의 관련

성 계산을 위하여 변형하여 사용하였다.

임의의 키워드_(m)와 채널 상의 다른 키워드_(r) 사이의 연관 계수를 구하기 위한 추정은 수식(2)을 통해 계산하였다. 수식(2)을 통한 키워드 간의 연관계수 계산은 생성된 채널 그룹들의 누적 평균값이 비교적 정확도가 높게 나타났다.

$$RC(k_m, k_r) = \frac{(freq(k_m, \sum_i FTF_i)^2 + freq(k_r, \sum_i FTF_i)) \times upper(k_r, \sum_i FTF_i)}{freq(k_m, \sum_i FTF_i) \times \sum_n UPk_n \times \sum_i FTF_i} \quad (2)$$

where $(0 \leq RC(k_m, k_r) \leq RC(k_m, k_m))$

$RC(k_m, k_r)$: 키워드_(m)과 키워드_(r)의 연관 계수

$freq(k_m, \sum_i FTF_i)$: 모든 풀 텍스트 피드 내의 키워드_(m)의 빈도수

$freq(k_r, \sum_i FTF_i)$: 모든 풀 텍스트 피드 내의 키워드_(r)의 빈도수

$upper(k_r, \sum_i FTF_i)$: 모든 풀 텍스트 피드 내의 키워드_(r)의 등장 횟수

$\sum_n UPk_n$: 채널 그룹 내 사용자 선호 키워드의 개수

$\sum_i FTF_i$: 채널 그룹 내 총 풀 텍스트 피드의 개수

<표 2>를 통해서 값이 1보다 큰 경우는 키워드_(m)이 자기 자신이거나 키워드_(m)가 속한 소속관계임을 판단할 수 있다. 장르와 그 내부의 키워드 간의 연관성은 사용자에게 제공할 콘텐츠를 개인화하는데 있어 중요한 요소이다. 연관성이 부족한 내용이 관심 콘텐츠로 반영되면 적용한 개인화

<표 2> 생성된 채널 그룹 내 임의의 키워드들에 대한 연관성 분석

키워드	빈도수	등장컨텐츠 수	선호 키워드	CORREL	수식 (2)	수식(2)의 채널그룹 누적평균	컨텐츠 선호도
제리 베벌	1	1		-0.25	0.19	0.09	0.25
킵스	0	0		DIV/0Error	0.00	0.06	0.00
드로그바	5	1	P	0.30	0.19	0.15	1.23
레들리 킹	0	0		DIV/0Error	0.00	0.06	0.00
레딩	3	2	P	-0.03	0.38	0.56	0.74
.....
맨유	26	7	P	-0.15	1.35	0.87	6.40
맨체스터 유나이티드	8	8	P	0.40	1.52	1.00	1.97
미들즈브러	1	1	P	0.44	0.19	0.34	0.25
박지성	31	9	P	1.00	1.75	1.31	7.63
설기현	14	3	P	-0.24	0.57	0.68	3.45
이동국	5	2	P	0.00	0.38	0.44	1.23
이영표	7	3	P	0.34	0.57	0.72	1.72
첼시	19	5	P	-0.06	0.96	0.67	4.68
토트넘	1	1	P	0.30	0.19	0.40	0.25
퍼거슨 감독	5	3	P	-0.18	0.57	0.34	1.23
프리미어리그	1	1		-0.25	0.19	0.21	0.25
프리미어리그	22	8	P	0.60	1.54	1.20	5.42
호날두	6	2		0.25	0.38	0.25	1.48
.....
UEFA컵	1	1		0.30	0.19	0.15	0.25

기법의 오류를 의미하기 때문이다.

요청한 키워드에 대해 생성된 콘텐츠 내용의 적합성 검사는 다음 수식(3)을 사용하였다. 콘텐츠가 제대로 생성되었는지 여부는 해당 키워드가 생성된 콘텐츠에 등장하는 빈도수에 영향을 받는다.

$$CSC(k_m) = \frac{freq(k_m, SJ) + freq(k_m, FTF) + \sum_j FTF(HRk_j)}{freq(k_m, FTF)} \quad (3)$$

where ($CSC(k_m) > 1$)

$CSC(k_m)$: 키워드(k_m)에 의해 생성된 콘텐츠에 대한 적합성 체크 값

$freq(k_m, SJ)$: 하나의 제목내의 채널 생성 키워드(k_m)의 빈도수

$freq(k_m, FTF)$: 하나의 풀 텍스트 피드 내의 채널 생성 키워드(k_m)의 빈도수

$\sum_j FTF(HRk_j)$: 하나의 풀 텍스트 피드 내의 연관계수가 높은 연관키워드 수의 합 ($HR(k) \geq 0.3$)

CSC(Contents Suitability Check)값은 제목내의 키워드 빈도수, 풀 텍스트 피드 내의 키워드 빈도수 그리고 풀 텍스트 피드 내의 연관성이 높은 키워드 빈도수의 합이 풀 텍스트 피드 내의 채널 생성 키워드(k_m)의 빈도수로 나눈 결과값으로 적어도 1보다 커야 한다. 이는 사용자가 질의 키워드(k_m)로 RSS 뉴스 채널을 검색하였을 때, 뉴스 기사의 제목에 질의 키워드가 나타나지 않는 경우와 풀 텍스트 피드 내에서 질의 키워드(k_m)와 연관성이 높은 키워드들도 나타나지 않는 경우의 연계성을 조사하여 풀 텍스트 피드의 내용이 전혀 상이한 경우를 적합성 기준에 의해 필터링한다.

생성된 채널의 콘텐츠에 대한 각 키워드의 콘텐츠 선호도는 <표 2>의 결과와 같으며 다음과 같은 수식(4)에 의해 계산하였다. 이에 따라 모든 풀 텍스트 피드내의 키워드(k_m)의 빈도수, 생성된 채널의 수와 사용자가 최근에 빈번히 요청한 키워드, 사용자가 선택한 선호 키워드의 수에 따라 그 채널이 사용자가 선호하는 내용이 많이 담겨 있는지 측정하였다.

$$UPC(k_m) = \frac{freq(k_m, \sum_j FTF_j) (\sum_g CCk_g + \sum_r RRk_r)}{\sum_r RRk_r \times \sum_n UPk_n} \quad (4)$$

$UPC(k_m)$: 키워드(k_m)에 의해 생성된 채널의 사용자 콘텐츠 선호도

$freq(k_m, \sum_j FTF_j)$: 생성된 하나의 채널 상의 모든 풀 텍스트 피드 내의 키워드(k_m)의 빈도수

$\sum_g CCk_g$: 질의 키워드에 의해 생성된 전체 채널 개수

$\sum_r RRk_r$: 사용자가 최근 빈번히 요청한 키워드의 개수

$\sum_n UPk_n$: 사용자가 선택한 선호 키워드의 개수

다음 [Algorithm 1]은 위 수식 (2), (3), (4)를 사용하여 계산된 사용자 선호도를 고려한 관심채널 설정과정을 나타낸다.

[Algorithm 1] 사용자 관심 채널생성과 개인화의 단계
(1) 질의(request) 키워드를 서브장르 내 키워드 분류에서 사용자가 선택하거나 분류에 존재하지 않으면 새로 입력한다.
(2) 새로 입력된 키워드는 해당 장르 내 연관키워드로 추가된다.
(3) 동일한 서브 장르에 대해 다른 사용자들이 질의한 키워드 또한 연관 키워드로 분류된다.
(4) 생성된 채널의 모든 Full Text Feeds를 탐색하며 수식(2)을 적용하여 키워드 간의 연관계수를 조사하여 콘텐츠 필터링의 기반 데이터로 이용한다.
(5) 연관 키워드로 설정된 모든 키워드들은 동일한 장르상에서 생성된 모든 채널들에 대하여 수식(3)에 의해 콘텐츠 적합성을 검증에 적용한다.
(6) 위 단계(5)에 의해 콘텐츠 적합성이 통과된 채널상의 모든 콘텐츠는 수식(4)에 의해 키워드에 대한 콘텐츠 선호도로 기록되며 질의 키워드들에 대한 콘텐츠 선호도가 사용자 프로파일에 기록된다. 이는 서브장르 내 키워드 리스트 뷰를 위한 그룹화(그림8-c)에 적용된다.
(7) 콘텐츠 선호도가 높은 키워드를 가지는 해당 서브장르 및 상위 장르는 개인화 단계에서 최상위의 뉴스 장르 배치에도 우선순위를 갖도록 반영하며 이 정보 역시 사용자 프로파일에 기록된다.

3.3 실시간 정보의 획득

사용자가 요청하는 키워드에 대해서 생성되는 채널의 콘텐츠는 실시간 정보 제공을 유지하기 위해 일정시간이 지나면 폐기되고 새롭게 생성되어야 한다. 따라서 최신의 정보를 전달 받을 수 있도록 사용자의 Request 키워드에 대한 정보 갱신 여부를 [Algorithm 2]와 같이 확인하고 처리한다. 이 과정을 통해 사용자는 항상 웹 사이트의 최신 정보를 모바일 웹 페이지를 통해서 서비스 받을 수 있다. 또한 원본 콘텐츠 소스로 RSS를 이용하기 때문에 다른 콘텐츠 소스보다 정보의 업데이트 유무를 빠르게 처리할 수 있다.

사용자의 질의 키워드를 입력받으면 이미 해당 키워드로 생성된 채널이 존재하는지 여부를 확인한다. 만약 채널이 생성되지 않았다면 RSS 원본 채널로부터 데이터를 획득하여 무선 웹용 채널 콘텐츠를 생성한다. 이때 콘텐츠 필터링 과정을 거치며 수식(2)의 적용에 의해 채널상의 부적합 콘텐츠를 걸러낸다. 이미 생성된 채널이 존재한다면 원본 채널 상에 추가된 아이템이 존재하는 여부를 조사하여 추가된 아이템이 있으면 무선 웹 용 채널 콘텐츠를 재생성하고 추가된 아이템이 없다면 사용자 뷰를 위한 개인화 과정을 거친 무선 웹 용 채널 콘텐츠를 서버의 응답으로 보낸다.

```

[Algorithm 2] 채널 콘텐츠의 생성 및 갱신
// Input: User Request Keyword
// Output: Personalized Pages
if Request Keyword::Channel Existence=True
  if Channel Contents Created Time>1Hours
    Connect to RSS Source Channel
    Confirm Additional Item
    if additional Item Existence=True
      Channel Renewal() {
        Channel Creation() & Contents Filtering() }
    end if
  end if
else
  Connect to RSS Source Channel
  Channel Creation() {
    Input a Request Keyword in Keyword List
    Crawling Pages through RSS Channel &
    Link Contents pages with the keyword }
  Contents Filtering() {
    Relational Keyword &
    Content Suitability Check &
    User Preferred Contents Operation }
end if
Personalization()
Send content pages to User
    
```

4. 실험 결과

4.1 실험 환경

실험 환경 구성은 다음과 같다. 운영체제는 Windows 2003 Standard Edition, 웹 서버 IIS 6.0, CPU는 펜티엄-4 3.0GHz, 메모리 1GB, 개발 도구는 Visual Studio.NET과 ASP를 사용했다. 클라이언트 접속 테스트는 OpenWave V7 Simulator와 LG-SD910 모바일 폰을 사용했다.

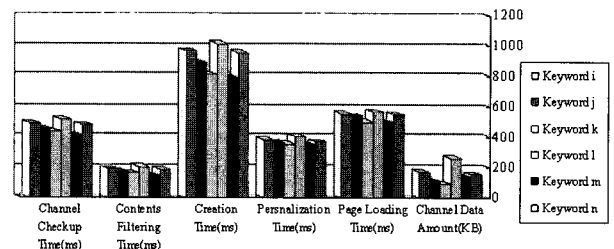
4.2 실시간 콘텐츠 생성 및 테스트

(그림 5)는 사용자가 선호 장르를 거쳐 선택한 질의 키워드 20개중 추출한 샘플 키워드 6개에 대한 원본 웹의 채널 변화 검증, 콘텐츠 필터링, 채널 생성, 개인화 기법의 적용,

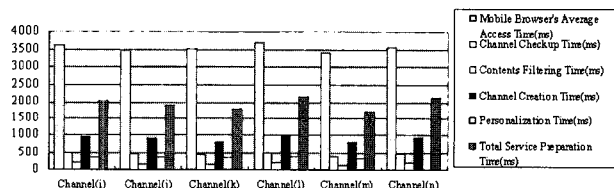
페이지 로딩 시간과 각 채널을 구성하는 데이터의 용량을 보여주는 차트이다. 또한 이미지 트랜스코딩 과정을 거쳐 용량이 줄어든 이미지들과 최적화된 모바일 웹 페이지들로 인해 하나의 채널 상의 모든 콘텐츠에 대한 데이터 용량임에도 차지하는 용량이 크지 않음을 알 수 있다.

(그림 6)은 샘플키워드들(i~n)에 의해 생성된 콘텐츠들(i~n)에 대한 모바일 브라우저의 평균 액세스 타임과 토털 서비스 준비 시간들을 보여준다. 토털 서비스 준비 시간이란 RSS 채널의 콘텐츠 변화를 체크하고 콘텐츠 필터링을 거쳐 모바일 서비스를 채널을 생성하고 개인화 과정을 거치는 시간들의 총합을 의미한다. 실험 결과 각 채널 콘텐츠에 대한 모바일 브라우저의 평균 접근 시간보다 전체 서비스 준비 시간이 현저히 작기 때문에 무선 인터넷을 통한 실시간 서비스가 가능하였다.

다음의 <표 3>은 임의의 서브장르(Sports -Soccer -EPL)내의 키워드에 대한 사용자 콘텐츠 선호도에 따른 개인화를 통해 그룹화가 어떻게 적용을 되었는지 보여준다.



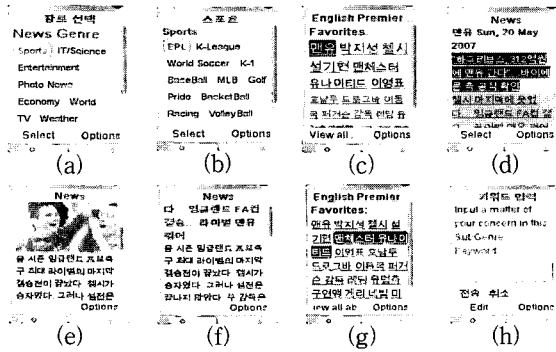
(그림 5) 사용자의 관심 키워드에 의한 채널 생성의 결과



(그림 6) 생성된 콘텐츠에 대한 모바일 브라우저의 평균 접근시간과 전체 서비스 준비시간의 비교

<표 3> 생성된 채널 내 키워드의 선호도에 따른 그룹화 적용

뉴스 장르 (Sample)	질의 키워드	전체 사용자의 요청 빈도	사용자의 콘텐츠선호도평균	선호도에 따른 그룹화	폰트 크기	배경 색상
스포츠-축구-프리미어리그 (Sports-Soccer-EPL)	맨유	162	6.62	group 1	15	#FFFF00
	박지성	134	7.10	group 1	15	#FFFF00
Meta Data 표현	첼시	115	5.56	group 1	15	#FFFF00
	설기현	97	4.45	group 1	15	#FFFF00
<meta name="1st_Genre" content="Sports" /> <meta name="2nd_Genre" content="Soccer" /> <meta name="1st_SubGenre" content="World Soccer" /> <meta name="2nd_SubGenre" content="EPL" />	맨체스터	31	2.21	group 2	13	#CCFFCC
	_유나이티드					
	이영표	28	2.97	group 2	13	#CCFFCC
	호날두	33	1.83	group 3	11	#CCCC66
	드로그바	22	1.15	group 3	11	#CCCC66
	이동국	26	1.92	group 3	11	#CCCC66
	퍼거슨 감독	19	1.32	group 3	11	#CCCC66
	레딩	18	0.91	group 4	10	#FFCC66
	유럽축구연맹	7	0.52	group 4	10	#FFCC66
	게리 네빌	4	0.38	group 5	9	#FFCC66
미들즈브러	6	0.29	group 5	9	#FFCC66	
토틸넘	5	0.41	group 5	9	#FFCC66	
.....						



(그림 7) (a)장르의 선택, (b)서브장르의 선택, (c)개인화가 적용된 채널 키워드의 선택, (d)선택한 채널 키워드에 대한 콘텐츠, (e)이미지가 포함된 풀 텍스트 피드의 탐색, (f)이미지 선호 옵션을 제거한 경우, (g)개인화가 제거된 채널 키워드의 선택, (h)추가적인 채널 키워드의 입력 화면

임의의 사용자(i)가 요청했던 키워드들에 의해 생성된 콘텐츠에 대한 각 키워드 선호도의 누적 평균을 기준으로 서브 장르내의 사용자(i)가 선호하는 키워드를 그룹화한다.

동일 서브 장르내의 임의의 질의 키워드에 의해 생성된 모든 하위 콘텐츠들은 동일한 장르 분류를 갖고 있기 때문에 해당 장르에 적용한 XSL문서들의 메타 데이터 안에 장르분류가 나타나도록 설정하였다. 전체 사용자들의 각 키워드에 대한 요청 빈도수가 특정 개인의 각 키워드에 대한 선호도와 부합되지 않는 것을 알 수 있다. 사용자 선호도의 계산에 의해 일정 수치 이상의 값을 갖는 키워드들에 대한 그룹화를 적용하였으며 폰트 사이즈 및 배경 색상이 다르게 나타나도록 설정하였다. 사용자 선호도에 의한 그룹화 적용의 모습은 (그림 7)의 (c)에 나타나 있다. (그림 7)은 본 연구에서 제안한 방법을 통해 생성된 모바일 웹 콘텐츠를 브라우저하는 모습이다.

사용자가 개인화를 적용한 콘텐츠를 브라우저하는 단계는 (그림 7)의 (a)~(e)까지이며 (f)~(g)는 부가적인 옵션 설정에 의한 화면이며 (h)는 서브장르 내에서 사용자가 원하는 키워드 채널이 구성되어 있지 않은 경우 원하는 질의 키워드를 입력하기 위한 화면이다.

(그림7-h)의 화면을 통해 새로 입력된 질의 키워드는 해당 장르에서의 사용자 선호 키워드와 서브장르의 연관 키워드로 사용된다.

5. 결론

보다 빠른 속도의 무선 접속을 지원하는 단말기의 보편화와 더불어 언제, 어디서나 인터넷을 이용할 수 있다는 장점이 제공되고 있지만 탑재된 브라우저의 성능 제약, 기존 인터넷 페이지의 큰 데이터 량, 모바일 웹 접속의 상대적 인고비용, 유용한 콘텐츠의 부족은 모바일 웹 이용의 장애물들이었다. 본 연구는 독창적인 개인화 기법을 적용하여 모바일 웹 접근의 제약을 완화시키고 사용자 선호도가 반영된

실시간 콘텐츠 접근이 가능한 방법을 제시하였다. 또한 이를 RSS를 통해 서비스함으로써 능동적 정보 업데이트가 가능하도록 했다. 제한한 콘텐츠 생성과정과 관심 키워드에 의한 채널 생성, 연관키워드 분석을 통해 사용자는 단순 빈도수 계산 및 복잡한 추론에 의한 개인화 과정을 거치지 않고 효과적인 결과물을 얻을 수 있었다. 제안방법은 사용자와 무선단말기에 적합한 개인화가 가능하며 RSS 뉴스 서비스를 제공하는 모든 콘텐츠에 적용할 수 있다는 장점을 갖는다.

참고 문헌

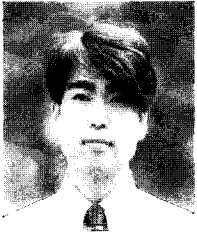
- [1] W.W.Lu :Compact multidimensional broadband wireless: the convergence of wireless mobile and access, (*Journal*) *IEEE Communications Magazine*, Vol.38, No.11, pp.119-123, 2000.
- [2] V. Kwan, F. Lau, C. Wang, "Functionality Adaptation: A Context Aware Service Code Adaptation for Pervasive Computing Environments", *IEEE/WIC International Conference on Web Intelligence*, Halifax, Canada, October 13-17, 2003.
- [3] 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.
- [4] <http://www.avantgo.com>
- [5] <http://www.earthlink.net>
- [6] 전영효, 황인준, "모바일 사용자를 위한 웹 서비스 페이지 개인화 기법", *정보과학회논문지, 컴퓨팅의 실제* 제11권 제1호, pp69-80, 2005.
- [7] 송특섭 외 6인, "MyNews: 모바일 환경에서 사용자 관심사를 고려한 XML문서 트랜스코딩", *정보처리학회논문지*, 제12-B권 제2호, pp.181-190, 2005.
- [8] T. Agrawal, T. Imielinski, and A. Swami, "Mining associations between sets of Items in Massive Databases", *Proceedings of the ACM SIGMOD International conference on Management of Data*, Washington D.C., May 1993, pp. 207-216.
- [9] P. Domingos, "Unifying Instance Based and Rule Based Induction, *Journal of Machine Learning*, Vol.24, No.2, pp.141-168, 1996.
- [10] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms", *Proc. of the Tenth International World Wide Web Conference on World Wide Web*, pp.285-295, 2001.
- [11] 김영지, 문현정, 옥수호, 우용태, "사례기반추론 기법을 이용한 개인화된 추천시스템 설계 및 구현", *정보처리학회논문지*, 제9-D권 제6호, pp1009-1016, 2002.



한 승 현

e-mail : power5v1@naver.com
2002년 숭실대학교 대학원 컴퓨터학과
공학 석사
2003년~현재 숭실대학교 대학원 컴
퓨터학과 박사과정

관심분야: 멀티미디어, 모바일 시스템, 인공 지능



류 동 엽

e-mail : aceryu@naver.com
2003년 숭실대학교 대학원 컴퓨터학과
공학 석사
2007년 숭실대학교 대학원 컴퓨터학과
공학 박사

관심분야: 영상처리, 멀티미디어, 모바일 시스템



임 영 환

e-mail : yhlim@ssu.ac.kr
1977년 경북대학교 수학과 이학사
1979년 한국과학기술원 전산학과이학
석사
1979~1996년 한국 전자 통신 연구소
책임 연구원
1985년 Northwestern University(이학박사)

1996년~현재 숭실대학교 미디어학부 교수
관심분야: 멀티미디어 공학, 모바일 시스템