

상황인식 정보 검색 기법을 이용한 하이브리드 협업 필터링 기법*

김 성 립** · 권 준 희***

A Hybrid Collaborative Filtering Method using Context-aware Information Retrieval

Kim, Sung Rim · Kwon, Joon Hee

〈Abstract〉

In ubiquitous environment, information retrieval using collaborative filtering is a popular technique for reducing information overload. Collaborative filtering systems can produce personal recommendations by computing the similarity between your preference and the one of other people. We integrate the collaboration filtering method and context-aware information retrieval method. The proposed method enables to find some relevant information to specific user's contexts. It aims to makes more effective information retrieval to the users. The proposed method is conceptually comprised of two main tasks. The first task is to tag context tags by automatic tagging technique. The second task is to recommend items for each user's contexts integrating collaborative filtering and information retrieval. We describe a new integration method algorithm and then present a u-commerce application prototype.

Key Words : Hybrid, Collaborative Filtering, Information Retrieval, Context-aware

I. 서론

인터넷과 웹의 급속한 성장으로 사용자가 이용할 수 있는 정보의 양은 폭발적으로 증가하는 환경에서 사용자에게 적합한 정보를 제공해주는 방법에 대한 연구가 활발히 진행되고 있는데 그 중에 협업 필터링 기법(Collaborative Filtering)[1-2]이 있다. 협업 필터링을 이용하여 사용자에게 정보를 제공할 때는 사용자의 취향과

유사한 다른 사용자의 취향을 참고하여 상품이나 항목에 대한 선호도를 예측하여 정보를 제공하게 된다. 즉, 사용자가 많을 때는 유사한 사용자 선호도 프로파일을 갖는 사용자들만을 따로 그룹핑을 한 후, 같은 그룹에 있는 이웃 사용자에게 의해 가장 높은 선호도를 보인 항목을 사용자에게 추천한다. 따라서 협업 필터링은 다양하고 유의한 항목들을 사용자에게 추천할 수 있는 장점이 있다.

유비쿼터스 환경에서 사용자의 상황(context) 정보를 활용하여 그에 맞게 사용자에게 정보를 제공하는 방법에 대한 연구도 진행되고 있다[3]. 여기서 상황은 사용자, 공간, 오브젝트 등의 개체와 관련된 모든 정보라고 정의된다[4].

* 본 논문은 2009년 서일대학 학술연구비 지원에 의하여 수행되었음

** 서일대학 인터넷정보과 조교수 (교신저자)

*** 경기대학교 컴퓨터과학과 부교수

이와 같이 유비쿼터스 환경에서 사용자의 상황을 고려한 정보 검색에 대한 필요성은 점점 커지고 있으나 기존의 협업 필터링 기법은 사용자의 상황을 고려하지 않고 있다. 따라서 본 논문에서는 유비쿼터스 환경에서 사용자에게 보다 효율적으로 정보를 제공하기 위해 상황인식 정보검색 기법을 이용한 하이브리드 협업 필터링 기법을 제안한다. 제안한 기법에서는 웹 2.0의 태깅 기법을 이용하여 사용자의 상황 정보를 자동으로 태깅한 후 기존의 협업 필터링 기법에 상황 정보 검색 기법을 통합한 하이브리드 협업 필터링 기법을 이용하여 각 사용자의 상황에 적합하게 정보를 제공한다.

본 논문의 구성은 다음과 같다. 2장에서는 배경 이론과 관련 연구들을 살펴보고, 3장에서 본 논문에서 제안하는 상황인식 정보 검색 기법을 이용한 하이브리드 협업 필터링 기법을 단계별로 설명한다. 그리고 4장에서 제안 기법을 구현한 모바일 환경에서의 프로토타입 어플리케이션을 보인다. 마지막으로 5장에서 결론을 맺는다.

II. 관련연구

협업 필터링 기법[5-6]은 사용자의 특성과 관련된 정보를 검색하여 사용자에게 적합한 정보를 제공하는 방법으로 많이 사용되는 기법 중의 하나이다. 즉, 협업 필터링 기법은 사용자가 선호하는 패턴과 유사한 다른 사용자의 선호도를 이용하여 사용자에게 관련된 정보를 제공하는 방법으로 크게 사용자 기반(user-based) 협업 필터링과 아이템 기반(item-based) 협업 필터링으로 나눌 수 있다[7].

사용자 기반 협업 필터링은 사용자와 사용자 간의 연관 관계를 파악하는 것이 핵심이며, 사용자 간의 유사성을 측정하여 선호도가 비슷한 다른 사용자들이 평가한 아이템을 기반으로 어떤 특정 사용자가 선호할만한 아이템을 권유하는 방식이다.

협업 필터링의 또 하나의 방법인 아이템 기반 협업 필

터링은 대부분의 사람들이 과거에 자신이 좋아했던 상품과 유사한 상품을 선호하는 경향이 있고, 반대로 선호하지 않았던 상품과 유사한 상품은 선호하지 않는 경향이 있다는 점을 바탕으로 한다. 이 방법은 아이템 간의 유사성을 측정하여 어떤 특정 사용자의 아이템 선호도를 예측하여 권유하는 방식이다.

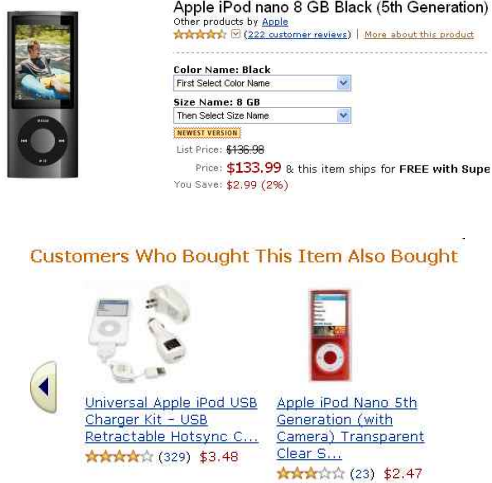
협업 필터링은 멀티미디어 정보와 문서에 대한 질을 반영할 수 있고, 기호가 유사한 다른 사용자들의 데이터를 이용하여 사용자의 잠재적 관심 분야를 찾아낼 수 있다는 장점이 있다. 하지만 많은 사용자들과 각 사용자로부터 많은 양의 데이터가 필요하다는 제약이 있다. 또한 새로운 관심 분야가 계속해서 발생하는 경우, 한 관심 분야에 대하여 다른 사용자들의 평가가 있어야 그 분야에 대한 정보를 추천할 수 있다는 문제점이 있다.

협업 필터링 기법을 이용한 응용 시스템으로는 음악 권유 시스템인 Ringo[8], 논문 검색 시스템인 Siteseer[5], 기사 권유 시스템인 GroupLens[6, 9], 비디오 권유 시스템인 Video Recommender[10], 웹에서 사용자와 관련된 정보를 찾아주는 PHOAKS[11] 등이 있다.

Last. fm[12]은 인터넷 라디오, 음악 커뮤니티 웹사이트로써 사용자 기반과 아이템 기반 협력 필터링을 모두 사용하여 사용자에게 정보를 제공한다. Last. fm은 비슷한 취향의 사람들과 비슷한 음악 검색 결과를 제공한다. 또한 Friends, group을 지정하여 취향을 공유할 수 있는 소셜 네트워크를 제공하고, 미디어플레이어와 연동되는 scrobbler, 아이팟, 아이폰 연동, 웹사이트 연동 위젯을 제공하는 등 다양한 매체를 연동한다.

Amazon[13-14]은 온라인 서점으로 시작하여 다양한 상품으로 확대한 미국 최대의 전자 상거래 사이트로써 아이템 기반의 협력 필터링 기법을 사용하여 개인화된 정보를 권유한다. 로그인 첫페이지가 개인화된 추천으로 구성되어있고, 개인 프로필 구축을 위한 상품 평가 요청 기능이 있다. 또한 로그인 후 최근 히스토리(브라우징, 구매 등)를 모두 저장한다. <그림 1>은 Amazon 사이트에서 'Apple iPod'을 검색했을 때 고객들이 Apple iPod

을 구입하면서 동시에 많이 구입한 다른 상품(Universal Apple iPod USB Charger Kit, Apple iPod Nano 5th Generation (with Camera) Transparent Clear Snap On Crystal Plastic Hard Cover Case)도 함께 검색하여 권유하는 것을 볼 수 있다.



<그림 1> Amazon(<http://www.amazon.com>)

사용자의 상황을 고려한 정보 검색에 대한 필요성은 점점 커지고 있으나 기존의 협업 필터링 기법은 사용자의 상황을 고려하지 않고 있어 유비쿼터스 환경과 자원이 제한적인 모바일 환경에서 사용자에게 정보를 효율적으로 제공하기가 어렵다는 문제점이 있다.

III. 하이브리드 협업 필터링 기법

본 절에서는 협업 필터링 기법과 정보 검색 기법을 통합한 권유 기법을 설명한다. 제안 기법은 크게 두단계로 구성되는데 첫 번째 단계는 웹 2.0의 태깅 기법을 활용하여 사용자의 상황 정보를 각 후보 정보들에 자동 태깅하는 단계로써[15]에서 제안한 방법을 적용한다. 두 번째

단계는 첫 번째 단계로부터 얻은 결과를 바탕으로 본 논문에서 제안하는 하이브리드 협업 필터링 기법을 이용하여 각 사용자의 상황에 적합한 정보를 제공하게 된다.

하이브리드 협업 필터링 기법의 목적은 유비쿼터스 환경에서 사용자들의 현재 상황에 적합한 정보를 권유하는 것이다. 기존의 전통적인 권유 기법으로 많이 사용되는 협업 필터링 기법은 각 사용자의 상황을 고려하지 않아 새로운 권유 기법이 필요하다. 이를 위해 본 논문에서는 기존의 협업 필터링 기법에 상황 정보 검색 기법을 통합한 새로운 하이브리드 협업 필터링 기법을 제안한다. 제안 기법은 4개의 주요 단계로 구성된다. 첫 번째 단계에서는 전통적인 협업 필터링 알고리즘을 수행하는데 본 논문에서는 아이템 기반 협업 필터링 알고리즘[5]을 사용한다. 이 알고리즘은 사용자가 선호도를 입력한 기존의 아이템들과 예측하고자 하는 아이템과의 유사도(similarity)를 계산하여 고객의 선호도를 예측하는 방법이다. 아이템 기반 협업 필터링 알고리즘[10]은 사용자 u 의 아이템 j 에 대해 예측 선호도 $P_{u,j}$ 를 식 (1)을 이용하여 계산한다. 식 (1)에서 $S_{i,N}$ 는 아이템 i 와 N 간의 유사도를 의미하며, $R_{u,j}$ 는 아이템 j 에 대한 사용자 u 의 선호도를 의미한다.

$$P_{u,i} = \frac{\sum_{\text{모든 유사 아이템 } N} (S_{i,N} * R_{u,N})}{\sum_{\text{모든 유사 아이템 } N} (|S_{i,N}|)} \quad (1)$$

그러나 각 아이템 정보에 대한 사용자의 선호도는 각 사용자의 현재 상황에 따라 달라질 수 있다. 즉, 첫 번째 단계는 아이템과 각 사용자 상황과의 관련성을 고려하지 않았다는 문제점이 있어서 이를 보완하기 위해 본 논문에서는 정보 검색 기법을 적용한다.

두 번째 단계에서는 상황 정보 검색 알고리즘을 수행한다. 정보 검색 기법으로 벡터 스페이스(vector space) 기반 알고리즘이 많이 사용되어 왔다. 본 논문에서는 기존의 벡터 스페이스 기반 정보 검색 알고리즘을 이용한 상황 정보 검색 알고리즘을 제안한 연구[3]를 적용한다. 이 단계에서는 각 아이템 정보에 태깅된 상황 태그의 개

수와 현재 상황 정보를 이용하여 사용자의 현재 상황과 아이템간의 유사도를 계산한다.

두 번째 단계에서는 사용자의 현재 상황과 관련된 정보를 찾기 위해 식 (2)를 이용하여 현재 사용자의 상황 C_q 와 아이템 정보 I_i 의 유사도 SC 를 계산한다. 여기서 W_{ij} 는 아이템 정보 I_i 에 대한 상황 C_q 의 가중치를 의미하며, W_{ij} 는 사용자의 현재 상황 C_q 에서의 상황 C 에 대한 가중치를 의미한다. 또한, k 는 소스 정보에 태깅된 상황 태그의 수를 의미한다.

$$SC(C_q, I_i) = \sum_{j=1}^{j \leq k} w_{ij} \times w_{ij} \quad (2)$$

세 번째 단계에서는 첫 번째와 두 번째 단계의 결과를 통합하기 위해 전통적인 협업 필터링 기법만으로는 부족했던 각 사용자의 상황과 관련된 정보를 추출할 수 있게 된다. 식(3)은 식(1)과 식(2)를 이용하여 사용자 u 와 u 의 상황 C_q 에 대한 아이템 정보 i 에 대한 통합된 예측 선호도 $IP_{u,q,i}$ 를 나타낸다. 여기서 P_{max} 는 최대 예측선호도를 의미하고, SC_{max} 는 최대 유사도를 의미한다. 또한 α 는 첫 번째와 두 번째 단계의 결과를 조정하여 통합할 수 있도록 한다.

$$IP_{u,q,i} = \alpha \times \frac{P_{u,i}}{P_{max}} + (1-\alpha) \times \frac{SC(C_q, I_i)}{SC_{max}},$$

where $0 \leq \alpha \leq 1$ (3)

마지막 네 번째 단계에서는, 세 번째 단계에서 얻은 예측 선호도 값을 내림차순으로 정렬하여 정보를 제공한다. 본 논문에서는 유비쿼터스 환경에서 널리 사용되는 모바일 환경을 고려한다. 모바일 환경은 기존의 환경과는 달리 자원이 제한된다는 제약점이 있기 때문에 사용자에게 제한된 자원에서 허용되는 양의 정보만 제공하는 것이 필요하다. 이를 위해, 마지막 단계에서는 정렬된 예측 선호도를 기반으로 상위 N개의 정보만을 추출하여 사용자에게 제공하게 된다.

IV. 프로토타입 어플리케이션

제안 기법을 서점 도메인에 적용하여 프로토타입 어플리케이션을 구현하였는데 위치 상황 생성기, 서비스 서버, 모바일 어플리케이션으로 구성된다. 위치 상황 생성기는 서점을 방문하고 있는 사용자의 상황을 시뮬레이션하기 위해 가상적으로 상황을 생성하는 프로그램으로써 C#을 사용하였다. 서비스 서버는 제안 기법을 이용하여 사용자의 상황에 적합한 정보를 서비스하는 프로그램으로 자바 언어를 사용하였다. 또한, 제안 기법과의 비교를 위하여 기존의 협업 필터링 기법에 의한 어플리케이션도 동일한 환경에서 구현하였다. 모바일 어플리케이션은 마이크로소프트 포켓 PC 에뮬레이터를 기반으로 C#을 사용한 PDA 어플리케이션을 구현하였다.

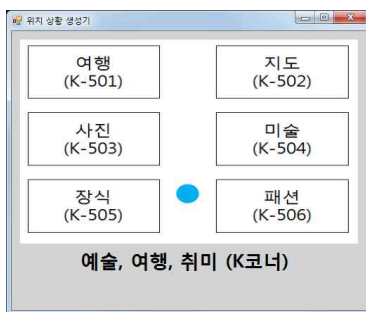
프로토타입 어플리케이션은 여름 휴가 기간에 30대인 김영희씨가 서점을 방문하는 시나리오를 기반으로 한다. 제안 기법을 이용한 어플리케이션이 설치된 PDA를 가지고 서점에 들어온 김영희씨는 현재 위치 상황에서 자신의 시간 상황과 나이 상황에 적합한 도서 정보를 PDA를 통해 자동으로 제공받아 원하는 도서를 편리하게 구매할 수 있다.

<표 1>은 서점의 K코너 위치에서의 여름 휴가와 30대라는 상황과 관련된 도서 정보를 보여준다. <표 1>에서 '판매량(나이)'은 30대 값이 태깅된 도서의 판매량이며, '판매량(시간)'은 여름 휴가 값이 태깅된 도서의 판매량을 나타낸다. 'SC'는 현재 상황 정보와 후보 정보간의 유사도를 의미하고, 'P'는 협업 필터링을 통한 사용자와 각 도서간의 예측 선호도를 의미한다. 'IP'는 상황 인식 정보 검색에 의한 SC와 협업 필터링 기법에 의한 P를 통합한 예측 선호도에 대한 계산 결과이다.

이제 김영희씨는 서점의 K코너로 이동하면, <그림 2>에서와 같이 위치 상황 생성기를 이용하여 김영희씨가 K코너로 이동한 모습을 보여준다. 김영희씨가 K코너에서 제공받을 수 있는 정보의 개수는 <표 1>로부터 11개임을 알 수 있다.

<표 1> 후보 도서 정보

도서명	판매량 (나이)	판매량 (시간)	SC	P	IP	위치
이혜영의 패션바이블	32	10	6.1	3.0	3.93	패션
홈패션 D.I.Y	20	20	5.8	3.02	3.85	패션
600가지 꽃 도감	21	10	4.5	3.0	3.45	장식
스케치 쉽게 하기	37	50	12.7	2.0	5.21	미술
제주권기여행	31	50	11.8	3.0	5.64	여행
카페 도쿄	32	80	16.3	2.5	6.64	여행
우리나라 그림 같은 여행지	29	60	13.0	2.3	5.51	여행
도로 지도 지도대사전	35	50	12.4	2.15	5.23	지도
정밀 도로지도	34	60	13.7	1.94	5.47	지도
네 멋대로 찍어라	21	20	6.0	1.5	2.85	사진
사진학강의	39	20	8.6	1.25	3.46	사진



<그림 2> 서점에서의 현재 위치

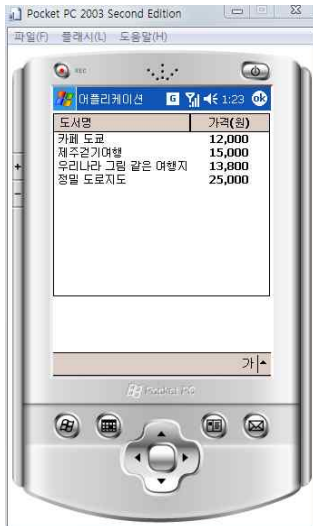
<그림 3>에서와 같이 K코너에서 기존의 협업 필터링 기법만을 사용한 어플리케이션은 협업 필터링 결과인 P 값을 정렬하여 보여주는데 다음과 같은 두가지 문제점이 있다. 첫째, 김영희씨의 '30대'라는 나이 상황과 '여름 휴가'라는 시간 상황을 고려하지 않아 현재 상황에 적합하지 않은 정보가 첫 화면에 상위로 추천된다. 둘째, 제한

된 모바일 기기(PDA)의 자원을 고려하지 않아 전체 정보를 모두 제공함으로써 김영희씨는 필요한 정보를 한번에 제공받기가 어렵다. 김영희씨의 PDA에서 한 번에 볼 수 있는 정보의 최대 개수가 4개인 현재 상황에서 기존 기법을 사용하는 경우에는 여러 번에 걸쳐 정보를 전송받아야 하는 어려움이 생긴다. 즉, 김영희씨가 현재 상황에 적합한 정보를 제공받기 위해서는 버튼을 여러번 클릭해야 하는 불편함이 있다.



<그림 3> 기존 기법을 이용한 모바일 어플리케이션

이에 비해, <그림 4>와 같이 제안 기법을 사용한 어플리케이션은 IP값을 기준으로 정렬하여 상위 4개의 정보만을 보여준다. 김영희씨에게 제공되는 정보는 김영희씨의 현재 나이 상황과 시간 상황에 적합한 정보로 구성되어 기존 기법에 비해 정보 제공에 대한 김영희씨의 만족도가 높아진다. 또한, 김영희씨의 상황에 적합한 4개의 정보만을 추출하여 제공하기 때문에 모바일이라는 제한된 환경에서 보다 효율적으로 정보를 제공할 수 있다. 이를 통해 볼 때, 제안 기법이 기존의 협업 필터링 기법만으로 정보를 추천하는 방법에 비해 좀 더 효과적으로 정보를 제공함을 알 수 있다.



<그림 4> 제안 기법을 이용한 모바일 어플리케이션

본 논문의 기여도는 다음과 같다. 첫째, 유비쿼터스 환경에서의 상황 인식 정보 검색 기법과 전통적인 협업 필터링 기법을 결합한 새로운 권유기법 알고리즘을 제안하였다. 둘째, 제안한 알고리즘을 서점 도메인에 적용한 프로토타입 모바일 어플리케이션을 개발하여 기존 기법과 본 제안 기법을 비교하여 사용자에게 정보 제공에 있어 보다 유용함을 보였다.

향후에는 사용자의 나이나 시간 상황 외에도 다른 상황들을 복합적으로 고려하여 사용자에게 필요한 정보를 제공할 수 있는 방법이 연구되어야 하고, 제안 기법의 우수성을 보이기 위해 다양한 실험이 이루어져야 할 것이다.

V. 결론

협업 필터링 기법을 이용한 권유 기법은 수많은 정보 중에서 사용자에게 유용한 정보를 추천함으로써 효과적으로 정보를 제공한다는 점에서 널리 사용되고 있다. 그러나 기존의 협업 필터링 기법은 최근의 유비쿼터스 환경에서 이용할 수 있는 각 사용자의 상황 정보를 고려하지 못한다는 문제점이 있다.

따라서 본 논문에서는 협업 필터링 기법과 상황 정보 검색 기법을 통합한 새로운 하이브리드 협업 필터링 기법을 제안하였다. 네 단계로 구성된 제안 기법의 알고리즘으로 기술하고, 모바일 어플리케이션을 프로토타이핑하여 본 제안 기법의 유용성을 보였다.

본 논문에서 제안하는 새로운 하이브리드 협업 필터링 기법은 전통적인 협업 필터링 알고리즘을 수행한 후 상황 정보 검색 기법을 이용하여 사용자의 현재 상황과 관련된 정보를 검색한다. 그리고 두 단계의 결과를 통합하여 상황 인식 컴퓨팅의 모바일 환경을 고려하여 사용자에게는 제한된 양의 정보만을 추출하여 제공한다.

참고문헌

- [1] Chen, A. and McLeod, D., "Collaborative Filtering for Information Recommendation Systems," Encyclopedia of Data Warehousing and Mining, Idea Group, 2005.
- [2] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D., "Using Collaborative Filtering to Weave an Information Tapestry," Communications of the ACM, Vol. 35, Issue 12, 1992, pp. 61-70.
- [3] Bardrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms", Proceedings of the 10th International Conference on World Wide Web, Hong Kong, 2001, pp.285-295.
- [4] Adomavicius, G., Sankaranarayanan, R., Sen, S., and Tuzhilin, A. "Incorporating Contextual Information in Recommender Systems using a Multidimensional Approach," ACM Transactions on Information Systems, Vol. 23, No. 1, 2005, pp. 103-145.

- [5] Anind K. Dey, Daniel Salber, Gregory D. Abowd, "A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications," Anchor Article of a Special Issue on Context-Aware Computing Human-Computer Interaction Journal, Vol. 16, No. 2-4, 2001, pp. 97-166.
- [6] Upendra Shardanand, and Pattie Maes, "Social Information Filtering: Algorithms for Automating 'Word of Mouth,'" In Proceedings of CHI'95 Conference on Human Factors in Computing Systems, 1995, pp. 210-217
- [7] Rucker J. and Plance, M. J, "Site-seer: Personalized Navigation for the web," Communications of the ACM, Vol. 40, No. 3, 1997, pp. 73-89.
- [8] Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "GroupLens : An Open Architecture for Collaborative Filtering of Netnews," Proceedings of the ACM Conference on Computer Supported Cooperative Work, 1994, pp. 175-186.
- [9] Kostan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J., "GroupLens: Applying Collaborative Filtering to Usnet News," Communications of the ACM, 1997, pp. 77-87.
- [10] Hill, W., L. Stead, M. Rosenstein, and G. Furnas, "Recommending and Evaluating Choices in a Virtual Community of Use," Proceedings of Conference Human Factors in Computing Systems, 1995, pp. 194-201.
- [11] Terveen, L., W. Hill, B. Amento, D. McDonald, and J. Creter, "PHOAKS : A System for Sharing Recommendations," Communications of the ACM, Vol. 40, No. 3, 1997, pp. 59-62.
- [12] <http://www.last.fm>
- [13] Linden, G., Smith, B., and York, J., "Amazon. com recommendations," IEEE Internet Computing Vol. 7, No. 1, 2003, pp. 76-80.
- [14] <http://www.amazon.com>
- [15] Sungrim Kim, Joonhee Kwon, "Folksonomy-Based Information Retrieval in Context-aware Environment," International Journal of Computer Science and Network Security, Vol. 8, No. 11, 2008, pp. 252-257.

■ 저자소개 ■



김 성 립
Kim, Sung Rim

2004년 3월~현재
서일대학 인터넷정보과 조교수
2002년 숙명여자대학교 컴퓨터학과 (이학박사)
1997년 숙명여자대학교 전산학과 (이학석사)
1994년 숙명여자대학교 전산학과 (이학사)

관심분야 : 유비쿼터스 컴퓨팅, 웹
데이터베이스, XML, 멀티미디어
의료 데이터베이스
E-mail : srkim@seoil.ac.kr



권 준 희
Kwon, Joon Hee

2003년 3월~현재
경기대학교 정보과학부
전자계산학전공 부교수
2002년 숙명여자대학교 컴퓨터학과 (이학박사)
1994년 숙명여자대학교 전산학과 (이학석사)
1992년 숙명여자대학교 전산학과 (학사)

관심분야 : 유비쿼터스 컴퓨팅 및 LBS, 공간
데이터베이스 및 GIS, 워크플로우
E-mail : kwonjh@kyonggi.ac.kr

논문접수일 : 2010년 1월 23일
수 정 일 : 2010년 2월 20일
게재확정일 : 2010년 2월 25일