
시맨틱웹 기반 개인 맞춤형 도서 추천 시스템

김진천*

Personalized Book Recommendation System based on Semantic Web

Jinchun Kim*

이 논문은 경성대학교 연구년 지원에 의하여 수행되었음.

요약

본 논문에서는 개인 맞춤 도서 추천을 위한 시맨틱웹 접근방법을 제안한다. 제안방법은 콘텐츠 기반 추천을 이용하면서도 사용자가 모든 도서 검색 시스템에 자신의 관심분야를 등록해야 하는 단점을 개선한다. 제안방법은 다양한 서지정보제공자의 도서분류 온톨로지상에서 자신의 관심분야를 등록할 수 있게 함으로써 사용자 프로파일을 공유한다. 또한 사용자 프로파일 관리 시스템은 제안방법에 의해 작성된 사용자 프로파일을 관리하고, 사용자의 관심분야와 도서분류 온톨로지상의 각 개념과의 유사성을 분석하는 기능을 제공한다. 제안방법은 사용자 프로파일의 공유를 통해 기존 키워드 검색에 비해 더 향상된 효율성을 제공한다.

ABSTRACT

In this paper, we propose a semantic web approach for personalized book recommendation. Our approach takes advantage of the content-based recommendation and improves its disadvantage that users should input their interesting fields into all book search systems they use. Our approach provides the sharing of users' profile with their interesting fields by enabling user's interesting fields to be described over each book classification ontology of various book information providers. We also provide a middleware that manages users' profiles written in RDF and analyzes similarity between user's interesting field and each concept over the book classification ontology. Our approach provide better performance than traditional keyword-based search by sharing the user's profile among book recommendation systems.

키워드

도서검색, 콘텐츠 기반 추천, 시맨틱웹, 온톨로지

Key word

Book Search, Content-based Recommendation, Semantic Web, Ontology

I. 서 론

저작도구(authoring tool)의 발달로 인해 다양한 분야에서 수많은 도서들과 문헌들이 출판되고 있어, 도서검색 기능을 제공하고 있음에도 불구하고 사용자들은 자신이 원하는 도서들을 찾기 위해 점점 더 많은 시간과 노력을 기울이고 있다[1]. 즉, 도서검색시스템의 품질을 향상시키기 위해서는 단순한 키워드 검색 결과를 제공하거나 많은 단계를 거쳐야 하는 기존 브라우징 방법에 대한 개선이 요구된다. 특히 도서관이나 인터넷 서점과 같이 출판물들을 대규모로 관리하는 시스템에서는 보다 효과적인 도서 검색방법이 필요하다. 따라서 Amazon.com의 북매처(BookMatcher)와 같이 사용자들이 관심을 가질만한 도서를 지능적으로 선별하여 추천하는 것이 필요하다.

정보검색 분야에서는 오래전부터 적합문서 선별 및 랭킹, 정보 필터링, 검색결과 클러스터링 및 분석 등 다양한 연구가 이루어지고 있으며, 최근 사용자의 요구와 관심에 따라 적합한 정보를 제공해주는 개인화서비스(Personalized Services)에 대한 많은 연구와 개발이 진행되어 왔다[1,2,3]. 대표적인 지능적 추천방법으로 콘텐츠 기반추천(Content-Based Recommendation)[4] 방식과 협업필터링(Collaborative Filtering)[5] 방식이 연구되고 있다. 콘텐츠 기반 추천시스템은 사용자 프로파일을 필요로 하며 이를 구축하기 위해 사용자의 참여를 유도하는 것이 쉽지 않아, 최근 협업필터링 및 하이브리드 추천방법에 대한 관심이 높아지고 있다.

하지만 콘텐츠 기반 추천 시스템은 이해하기 쉽고 계산과정이 간단하며, 본인이 직접 선택한 관심분야를 사용하여 추천하기 때문에 사용자 프로파일을 만드는 과정에서 사용자의 참여를 줄여준다면 효과적인 추천방법이 될 수 있다. 기존의 추천 시스템에서 사용자의 참여 유도가 어려운 이유중의 하나는 추천 서비스를 제공하는 각 서비스마다 따로 자신의 관심정보를 등록하기 때문으로, 이러한 서비스들이 사용자가 이미 등록한 다른 서비스의 관심정보를 공유하고 활용할 수 있는 방법이 필요하다.

본 논문에서는 도서검색시스템에서 콘텐츠 기반 추천 시스템의 장점을 활용하면서 사용자의 불편함을 개선할 수 있는 하나의 방법으로 시맨틱웹 접근방법을 제

안한다. 제안방법은 선행연구[6]에서 제안한 시맨틱웹 기반 푸시서비스 모델을 기반으로 설계 및 구현한다. 시맨틱웹 기반 푸시서비스 모델[6]은 서비스 온톨로지(service-specific ontology)들 상에서 콘텐츠 정보(메타데이터)와 사용자의 선호정보를 기술하고 이들 정보들 간에 강력한 상호운용성을 제공할 수 있도록 온톨로지 기반 데이터 모델을 제공한다. 또한 의미적 연관성을 바탕으로 사용자에게 적합한 정보를 추천할 수 있도록 프레임워크를 제공한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 시맨틱웹 기술에 대해 간략히 소개하고, 3장에서는 시맨틱웹 기반 도서추천시스템을 제안한다. 4장에서는 제안 시스템을 분석하고, 5장에서 결론을 맺는다.

II. 시맨틱 웹

시맨틱웹은 컴퓨터가 웹상의 자원(resource)을 의미 기반으로 표현하고 처리할 수 있도록 하는 기술로서, 기하급수적으로 증가하고 있는 웹상의 자원을 지능적으로 검색할 수 있도록 기반구조를 제공하는 차세대 웹 기술이다[7,8].

시맨틱 웹의 가장 기본이 되는 URI(Uniform Resource Identifier)는 웹 자원의 고유 식별자로 사용되며, XML Namespace 및 XMLSchema는 리소스의 표현을 구조화할 수 있도록 지원함으로써 문서의 구조적인 측면에서 상호운용성을 제공한다[8]. RDF(Resource Description Framework)와 RDF Schema는 웹상의 자원(Resource)들을 정의하고 기술하기 위한 XML 기반의 데이터 모델로서, 컴퓨터가 리소스를 의미적으로 처리할 수 있는 기반을 제공한다. Ontology는 웹 환경에서 자원에 대한 개념과 관계를 정의함으로써 웹상의 자원에 대한 의미를 정의한다. 이렇게 기술된 웹상의 자원들은 Logic, Proof, Trust 계층의 요소기술들을 통해 자원에 대한 지능화된 접근과 신뢰성을 제공받게 된다[8].

특히, RDF와 Ontology는 시맨틱웹의 핵심적인 기반 구조로서의 역할을 담당한다. RDF는 데이터를 정의하기 위해 자원간의 관계를 Triple 구조(subject-predicate-object)로 표현함으로써 컴퓨터가 의미를 처리할 수 있는 기반을 제공한다[9]. 온톨로지(Ontology)[7]는 웹 환

경에 존재하는 모든 자원을 개념(concept 또는 class)으로 정의하고 그 의미적인 처리를 위해서 그 개념간의 관계를 속성(property)으로 정의함으로써, 컴퓨터가 웹 자원들을 의미적으로 이해하고 처리할 수 있는 기반을 제공한다. 온톨로지를 정의하기 위한 여러 가지 방법이 있지만, 특히 OWL(Web Ontology Language)[10]은 W3C에서 웹 자원의 관계를 풍부하게 정의할 수 있도록 제공하는 온톨로지를 기술하기 위한 언어이다.

III. 시맨틱 웹 기반 도서 추천 시스템

3.1. 시스템의 구성

본 논문에서는 도서 검색 시스템의 품질을 향상시키기 위한 방법으로 시맨틱 웹 접근방법을 제안한다. 제안 방법은 다양한 도서들에 대한 서지정보를 정확하게 표현할 수 있도록 충분한 표현력을 제공한다. 또한 이렇게 다양한 도서 분류를 기반으로 사용자의 관심분야를 보다 정확하게 기술할 수 있으면서도 이러한 정보들 사이의 의미적인 연관성을 분석할 수 있는 방법을 제공하여 다른 시스템에서도 사용자의 관심분야를 공유할 수 있도록 한다.

제안방법은 도서정보와 사용자의 관심분야를 기술하기 위해 선행연구[6]에서 제안한 OWL(Web Ontology Language) 기반의 콘텐츠 정보 및 사용자 선호정보를 기술하기 위한 언어인 CDL(Contents Description Language)과 UPDL(User Preference Description Language)을 사용한다. CDL/UPDL은 정보를 다양한 도메인 온톨로지 상에서 표현할 수 있도록 정의된 일종의 간단명료한 언어로서, 정보를 표현함에 있어 유연성과 상호운용성을 제공하는 기반이 된다.

제안하는 개인 맞춤 도서 추천 시스템은 그림 1과 같이 사용자 프로필 관리자(User Profile Manager)와 유사성 분석기(Similarity Analyzer)로 구성된다. 먼저, 각 도서관이나 인터넷 서점과 같은 서지정보제공자는 자신들의 도서 분류 기준에 따라 기술된 도서분류 온톨로지(Book Classification Ontology)를 제공하고, 자신들의 서지정보를 이 온톨로지 상에서 CDL로 기술하여 RDF 형식으로 저장하여 제공한다.

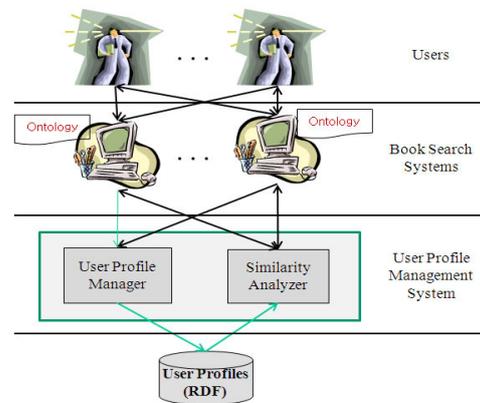


그림 1. 제안 시스템 구성 [1]
Fig. 1 Structure of the proposed system

사용자 프로필 관리자는 사용자로부터 관심분야를 입력 받아 RDF형식으로 저장하는 기능을 수행하는데, 이때 관심분야는 UPDL에 따라 다양한 도서분류 온톨로지 상에서 기술된다. 유사성 분석기는 사용자의 요청에 의해 자신의 프로필에 기술된 관심분야를 기준으로 특정한 도서분류 온톨로지 상의 개념들에 대한 평가가중치를 계산하는 기능을 수행한다. 이렇게 평가된 유사성 분석 결과는 해당 도서 검색 시스템에서 사용자에게 적합한 도서를 추천하기 위해 사용된다. 제안시스템에서 OWL과 RDF문서의 처리는 온톨로지와 RDF 문서 처리를 위한 Java API인 Jena Toolkit [11]을 사용하여 구현하였다.

3.2. 서지정보 생성

본 논문에서 도서에 관한 정보(서지정보)는 각 도서관 및 인터넷 서점과 같은 서지정보제공자가 자신의 도서분류 온톨로지를 참조하여 CDL로 기술하여 제공하는 것으로 가정한다. 본 논문에서 사용하는 도서분류 온톨로지는 일종의 도메인 온톨로지(Domain Ontology)로, 각 도서 검색 시스템에서 제공하는 도서에 대한 분류와 서지정보를 기술하기 위한 온톨로지이다. 즉, 도서분류 온톨로지는 각 도서검색시스템에서 도서의 분류를 계층적으로 정의하고 서로 다른 개념과의 관계를 프로퍼티로 정의해 줌으로써 각 서지정보에 대한 정보를 보다 정확하게 의미적으로 표현할 수 있도록 한다.

그림 2는 ACM Digital Library에서 제공하는 분류를 온톨로지 구성한 예이며, 그림 3은 이 온톨로지 상에

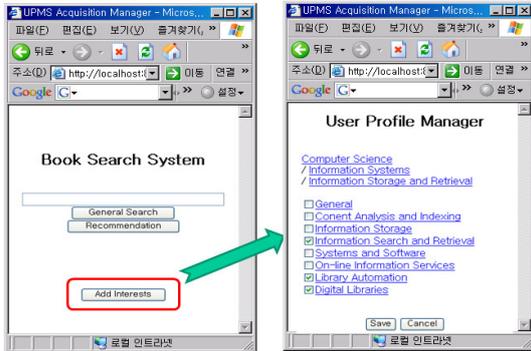


그림 5. 사용자 관심분야 등록 화면
Fig. 5 User interface for registration of a user's interesting field

그림 6은 사용자의 관심분야가 등록된 사용자 프로파일의 예를 보인 것이다. 예에서, “updl”과 “ontLibraryA”는 UPDL 온톨로지와 A대학 도서관의 도서분류 온톨로지의 네임스페이스이며, UserA는 이 도서분류상에서 “ComputerScience”분야의 가중치 10의 관심을 가지고 있다는 것을 나타내고 있다.

```

<updl:Profile rdf:ID="UserA">
  <updl:personInfo rdf:resource="#person:UserA">
  <updl:hasPreference rdf:resource="#UserA_Interest1">
</updl:Profile>
<updl:Interesting rdf:ID="UserA_Interest1">
  <updl:hasItem rdf:resource="#ontLibraryA:ComputerScience">
  <updl:hasWeight> 10 <updl:hasWeight>
</updl:Interesting>
    
```

그림 6. 사용자 프로파일 예
Fig. 6 An example of user profile

3.4. 개인 맞춤 서지정보 검색

개인 맞춤 도서 검색은 크게 개인 맞춤 브라우저와 개인 맞춤 키워드 검색으로 나눌 수 있다. 이 두 가지 검색 서비스는 프로파일 관리 시스템의 유사성 분석기로부터 획득한 도서 분류 온톨로지의 유사성 평가 결과, 즉 도서 분류 온톨로지 상의 개념(Concept)들에 대한 평가가중치(weight)를 기준으로 제공된다.

여기서 유사성을 평가하기 위한 다양한 연구들이 있으나, 본 논문에서는 콘텐츠의 분류와 사용자 선호분야와의 거리에 따라 평가가중치를 50%씩 점진적으로 감소하는 이전연구[6]의 방법을 적용하였다. 즉, 각 도서의

분류가 등록된 사용자의 관심분야와 일치하지 않을 경우, 해당 도서의 평가가중치를 계산하기 위해 그림 7과 같이 반복적으로 온톨로지 상의 상위개념과 비교하면서 동일한 개념을 찾을 때까지 평가가중치를 50%씩 감소시킨다. 예를 들어, 분류가 “Programmin Techiques”인 도서의 경우, 등록된 사용자의 관심분야와 일치하지는 않지만 상위개념과 일치하기 때문에 그림 7과 같이 단계마다 50%씩 평가가중치를 감소되어, 평가가중치는 2.5가 된다.

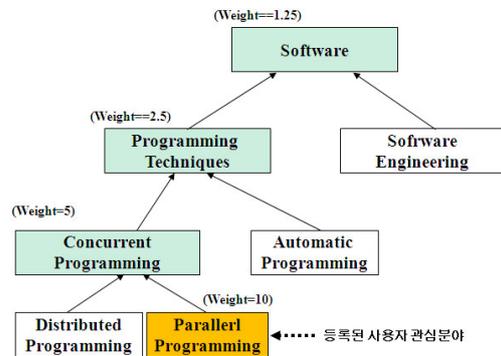


그림 7. 유사도 분석
Fig. 7 Similarity Analysis

유사성 평가를 위해 두 개념 사이의 거리에 따라 평가가중치를 감소시키는 방법은 다양하게 적용할 수 있다. 예를 들어, 거리가 늘어날 때 마다 일정한 값을 감소시킬 수도 있으며, 단계별로 1/h (h는 온톨로지의 높이)씩 감소시킬 수도 있다. 하지만 일반적으로 거리가 멀어짐에 따라 두 개념간의 유사성은 현저히 감소하기 때문에, 본 논문에서와 같이 단계별 50%씩 감소시키는 것은 직관적으로 타당한 방법으로 볼 수 있으며 구현도 다른 방법에 비해 용이하다.

그림 8은 서지정보 검색과정을 나타낸 것이다. 사용자가 서지정보 검색을 위해 서지정보제공자의 도서 검색 서비스에 접속하면, 서지정보제공자는 자신의 도서 분류 온톨로지와 함께 사용자 프로파일 관리시스템의 유사성 분석기로 리다이렉트(redirect) 시킴으로써 사용자 관심분야와의 유사성 평가를 요청한다. 사용자 프로파일 관리시스템은 사용자를 인증한 후, 유사성 평가 결과를 사용자를 통해 서지정보 제공자에게 전송한다. 서지정보제공자는 유사성 평가결과를 서지정보 브라우저

과 키워드 검색에 반영하여 개인 맞춤 서지정보 검색결과를 사용자에게 제공한다.

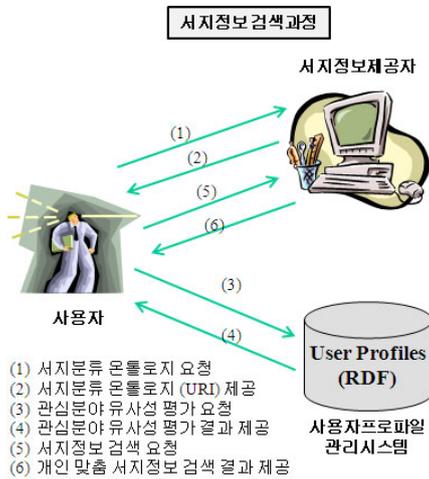


그림 8 서지정보 검색
Fig. 8 Search of Book information

IV. 제안 시스템 분석

제안 방법은 다양한 도서검색시스템이 각자의 도서분류를 사용하는 것을 허용하기 때문에 사용자의 관심분야를 풍부한 어휘를 통해 표현할 수 있다. 또한 제안 방법은 이들 관심분야 정보를 여러 시스템이 공유하여 사용할 수 있기 때문에 사용자의 중복 입력의 불편함은 감소시키면서 사용자에게 적합한 도서를 추천할 수 있다.

그림 9와 그림 10은 각각 기존 키워드 검색과 제안 방법에서의 키워드 검색의 결과를 보인 것이다. 제안 방법의 경우 기존 키워드 검색과 달리 사용자의 관심분야를 고려하여 도서를 추천함을 볼 수 있다.

제안 시스템을 평가하기 위해 ACM Digital Library, 국내 대학 도서관, 국내 인터넷 서점의 웹사이트를 참조하여 3개의 도서분류 온톨로지를 정의하였다. 이들 중 대학 도서관에서 다양한 분야의 도서에 대한 서지정보 1,000개를 해당 도서분류 온톨로지를 기반으로 CDL로 구성하였으며, 나머지 2개의 도서분류 온톨로지 상에서 사용자 관심분야 5개를 등록하였다.

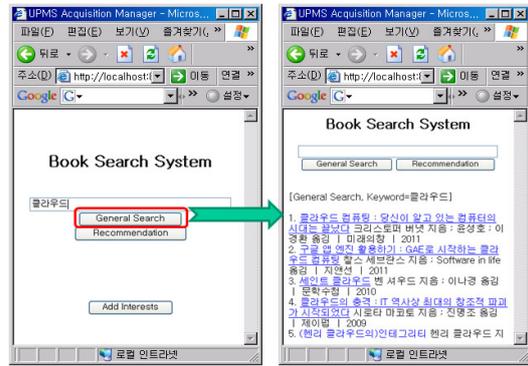


그림 9. 기존 키워드 검색
Fig. 9 Traditional keyword-based search

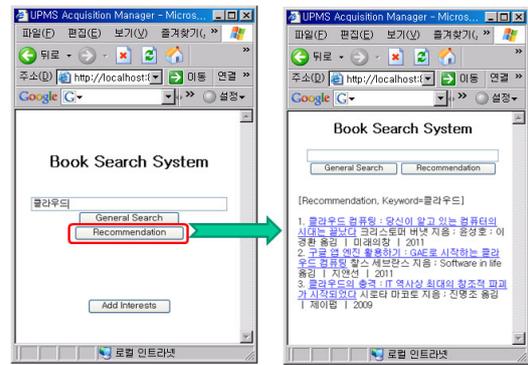


그림 10. 개인 맞춤 키워드 검색
Fig. 10 Personalized keyword-based search

검색 시스템의 효율성 평가는 일반적으로 정확률(Precision)과 재현율(Recall)로 평가될 수 있다. 정확률은 검색된 정보 가운데 적합 정보의 비율이며, 재현율은 시스템이 적합정보를 검색하는 능력을 의미한다. 검색 시스템 내의 관련있는 서지정보의 전체 수를 Td, 검색 시스템에 의해 실제 검색된 서지정보 수를 Tr, 검색된 서지정보 중 관련있는 도서의 수를 R이라고 할 때, 정확률과 재현율은 다음과 같이 계산되어진다. 이때, R 집합은 1000개의 도서 중 실험 키워드 10개에 대해 관련 있는 도서를 직접 분류하고 실험 결과와 비교하였다. 실험 키워드로는 시맨틱웹, 온톨로지, 개인화, 추천, 필터링, 모델링, 추론, 클라우드, 자바, 애플을 사용하였다.

$$\text{Precision(정확률)} = R/Tr$$

$$\text{Recall(재현율)} = R/Td$$

표 1은 10개 단어의 키워드 검색에 대한 제안시스템의 성능을 평가한 결과이며, 표 2와 표 3은 키워드 검색과 제안방법에서의 실험결과이다. 표 1에서 제안방법이 단순 키워드 검색에 비해 재현율은 다소 낮지만 정확률은 매우 높아짐을 알 수 있다.

표 1. 제안시스템의 성능 평가
Table. 1 Performance evaluation of our system

구분	키워드 검색	제안방법
정확률(%)	59	93
재현율(%)	100	96

본 논문에서의 실험을 통해 제안방법의 유용성은 제시하였으나, 제한된 수와 범위의 키워드에 대해서만 실험하였기 때문에 향후 보다 다양한 키워드에 대한 실험이 필요하다. 또한 제안방법은 도서와 사용자의 관심분야 사이의 유사도를 평가해야 하기 때문에 기존의 키워드 검색에 비해 추가적인 처리시간이 요구된다. 본 실험에서는 서지정보의 수가 많지 않아 4초 이내의 응답시간을 보였으나, 향후 충분히 많은 서지정보를 대상으로 검색시간에 대한 평가가 필요하다.

표 2. 키워드 검색에서의 실험결과
Table. 2 Experimental results in keyword search

키워드	R	Td	Tr	정확률	재현율
시맨틱웹	6	6	6	1.00	1.00
온톨로지	1	1	2	0.50	1.00
자바	5	5	9	0.55	1.00
개인화	1	1	2	0.50	1.00
애플	6	6	10	0.60	1.00
추론	4	4	6	0.66	1.00
필터링	2	2	4	0.50	1.00
추천	4	4	10	0.40	1.00
클라우드	6	6	8	0.75	1.00
모델링	4	4	10	0.40	1.00
평균	-	-	-	0.59	1.00

표 3. 제안방법에서의 실험결과
Table. 3 Experimental results in our model

키워드	R	Td	Tr	정확률	재현율
시맨틱웹	6	6	6	1.00	1.00
온톨로지	1	1	1	1.00	1.00
자바	5	5	5	1.00	1.00
개인화	1	1	1	1.00	1.00
애플	5	6	7	0.71	0.83
추론	4	4	4	1.00	1.00
필터링	2	2	2	1.00	1.00
추천	3	4	5	0.60	0.75
클라우드	6	6	6	1.00	1.00
모델링	4	4	4	1.00	1.00
평균	-	-	-	0.93	0.96

V. 결 론

본 논문에서는 기존의 도서 검색 시스템에 비해 향상된 효율성을 보이는 개인 맞춤형 도서 추천 시스템을 제안하였다. 제안 방법은 사용자의 관심분야를 각 도서 검색 시스템의 도서분류 온톨로지들 상에서 기술할 수 있도록 하고 이들을 관리하기 위한 미들웨어를 제공함으로써, 사용자의 관심분야를 서로 다른 시스템들에서도 공유하여 사용할 수 있는 방법을 제공하였다. 또한, 각 도서 검색 시스템에서는 도서들의 서지정보를 자신의 도서분류 온톨로지에 따라 기술하고, 사용자의 관심분야를 저장한 프로파일을 기반으로 평가 가중치를 계산함으로써 효율적인 도서 추천 서비스가 가능하도록 하였다. 제안방법은 키워드 검색에서 기존 검색 시스템에 비해 재현율은 비슷하면서 정확률이 매우 높아 효율적임을 확인할 수 있었다.

참고문헌

- [1] 이상기, 이병섭, 박병용, 황혜경, “나이트베이스 분류모델과 협업필터링 기반 지능형 학술논문 추천시스템 연구,” 정보관리연구, 제41권, 제4호, pp.227-249, 2010.

- [2] L.Geng, Z.Jeng, and Y.Jiang, "Research on E-Commerce personalized service based on intelligent agent technology," *IEEE Int. Conf. on Service Operations, Logistics, and Informatics*, pp. 1113-1118, 2008.
- [3] A.Saghiri and A.Bagheri, "An Adaptive Architecture for Personalized Search Engine in Ubiquitous Environment with Peer to Peer Systems," *Int. Conf. on Information and Multimedia Thechnology*, pp. 107-111, 2009.
- [4] B.Krulwich and C.Burkey, "Learning user information interests through extraction of semanticallysignificant phrases," *Proc. of the AAAI Spring Symposium on Machine Learning in Information Access*, pp. 110-112, 1996.
- [5] J.A.Konstan, B.N.Miller, D.Maltz, J.L.Herlocker, L.R.Gordon, and J.Riedl, "GroupLens: applying collaborative filtering to Usenet news," *Communications of the Association*, vol 40, no.3, pp.77-87, 1997.
- [6] 김주연, 김종우, 김진천, "개인화서비스를 위한 시맨틱웹 기반 푸시서비스 기법," *한국콘텐츠학회논문지*, 제10권, 제6호, pp.18-26, 2010.
- [7] T.Berners-Lee, *The Semantic Web*, Scientific American, 2001.
- [8] Antoniou and van Harmelen. *A Semantic web Primer*, The MIT Press, Cambridge, 2004.
- [9] F.Manola and E.Miller, *RDF Primer* (<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>), W3C (MIT, ERCIM, Keio), 2004.
- [10] M.K.Smith, C.Welty, and D.L.McGuinness, *OWL Web Ontology Language Guide* (<http://www.w3.org/TR/owl-guide/>), W3C (MIT, ERCIM, Keio), 2004.
- [11] J.J.Carroll, I.Dickinson, and C.Dollin, "Jena: Implementing the Semantic Web Recommendation," *WWW2004*, ACM, New York, USA, pp. 74-83, 2004.

저자소개



김진천(Jinchun Kim)

1983년 한양대학교 전기공학과 (공학사)

1985년 미시간주립대학교 전자 및 시스템공학과 (공학석사)

1996년 KAIST 전산학과 (공학박사)

1988년-1996년 삼성종합기술원 선임연구원

1996년-현재 경성대학교 컴퓨터학부 교수

※관심분야: HCI, 멀티미디어통신, 센서네트워크