
시맨틱 웹에서 개인화 프로파일을 이용한 콘텐츠 추천 검색 시스템

Contents Recommendation Search System using Personalized Profile on Semantic Web

송창우*, 김종훈*, 정경용**, 류중경***, 이정현****

인하대학교 정보공학과*, 상지대학교 컴퓨터정보공학부**, 대림대학교 컴퓨터정보계열***,
인하대학교 컴퓨터공학부****

Chang-Woo Song(ph.d.scw@hanmail.net)*, Jong-Hun Kim(jhkim@hci.inha.ac.kr)*,
Kyung-Yong Chung(dragonhci@hanmail.net)**, Joong-Kyung Ryu(jkryu@daelim.ac.kr)***,
Jung-Hyun Lee(jhlee@inha.ac.kr)****

요약

정보기술의 발전과 인터넷 사용의 증가로 이용가능한 정보들의 양이 폭발적으로 증가한다. 콘텐츠 추천 시스템은 사용자가 원하지 않는 정보를 필터링하고 유용한 정보를 추천하는 서비스를 제공한다. 기존의 추천 시스템은 데이터마이닝 기법으로 웹 접속 기록 및 유형과 사용자가 요구하는 정보를 서비스 제공자 측면에서 분석하여 콘텐츠를 제공한다. 사용자의 선호도와 생활패턴 등의 사용자 측면에서의 정보들의 표현이 어려워기 때문에 제한된 서비스를 제공할 수 밖에 없었다. 시맨틱 웹 기술은 이미지, 문서 등의 모든 객체를 대상으로 목적에 맞는 정보를 수집, 가공, 응용할 수 있도록 데이터 간에 잘 정의된 의미 있는 관계를 만들 수 있다.

본 논문에서는 시맨틱 웹 환경에서 개인화 프로파일을 동적으로 갱신하여 반영할 수 있는 콘텐츠 추천 검색 시스템을 제안한다. 개인화 프로파일은 프로파일의 특징을 담고 있는 컬렉터, 다양한 컬렉터들로부터 프로파일을 수집하는 수집기, 프로파일 특성에 기반한 고유의 프로파일 컬렉터를 해석하는 해석기로 구성된다. 개인화 모듈은 콘텐츠 추천 서버에서 개인화 프로파일과 주기적으로 동기화할 수 있도록 도와준다. 추천 콘텐츠로 음악을 선택하여 서비스 시나리오에 따라 개인화 프로파일이 콘텐츠 추천 서버에 전달되어 사용자의 선호도와 생활패턴이 반영된 추천리스트를 제공하는지 실험한다.

■ **중심어** : | 시맨틱 웹 | 프로파일 | 개인화 | 콘텐츠 | 추천 시스템 |

Abstract

With the advance of information technologies and the spread of Internet use, the volume of usable information is increasing explosively. A content recommendation system provides the services of filtering out information that users do not want and recommending useful information. Existing recommendation systems analyze the records and patterns of Web connection and information demanded by users through data mining techniques and provide contents from the service provider's viewpoint. Because it is hard to express information on the users' side such as users' preference and lifestyle, only limited services can be provided. The semantic Web technology can define meaningful relations among data so that information can be collected, processed and applied according to purpose for all objects including images and documents.

The present study proposes a content recommendation search system that can update and reflect personalized profiles dynamically in semantic Web environment. A personalized profile is composed of Collector that contains the characteristics of the profile, Aggregator that collects profile data from various collectors, and Resolver that interprets profile collectors specific to profile characteristic. The personalized module helps the content recommendation server make regular synchronization with the personalized profile. Choosing music as a recommended content, we conduct an experience on whether the personalized profile delivers the content to the content recommendation server according to a service scenario and the server provides a recommendation list reflecting the user's preference and lifestyle.

■ **keyword** : | Semantic Web | Profile | Personalization | Content | Recommendation System |

* 본 논문은 BK21 사업단에서 지원받은 연구결과입니다.

접수번호 : #071228-002

접수일자 : 2007년 12월 28일

심사완료일 : 2008년 1월 23일

교신저자 : 이정현, e-mail : jhlee@inha.ac.kr

I. 서론

정보기술과 인터넷의 발전은 정치, 경제, 사회, 문화 등 많은 부분을 변화시켰다. 다양한 매체들에 의해 사람들에게 전달되던 각종 정보들은 컴퓨터 네트워크를 통해 콘텐츠로서 빠른 속도로 전송된다. 그 결과 불필요한 정보들이 넘쳐나게 되었고 사람들이 원하는 콘텐츠를 찾을 수 있는 기능의 중요성이 부각된다. 개인 맞춤형 콘텐츠는 경쟁이 치열한 시장에서 자신을 차별화할 수 있는 하나의 수단이다.

시맨틱 웹의 자동화 기술을 이용해 좀 더 편리하게 정보를 공유하고 학습할 수 있다. 또한 이미지, 문서 등의 모든 객체를 대상으로 목적에 맞는 정보를 수집, 가공, 응용 한다[1]. 최근에는 미리 사용자가 원하는 콘텐츠만을 예측하고 추천해주는 추천 시스템이 활발히 연구되고 있다[2]. 이미 웹상에서 추천 서비스가 제공되고 있으며 실제 데이터를 보유하지 않아도 많은 사용자들은 자신의 ID만으로 상호간의 공유 및 추천한다. 국내의 벅스뮤직과 같은 음악 전문 포털사이트에서는 시대별, 연령별, 테마별 추천 서비스를 제공한다[3]. 해외에서도 많은 사용자들은 last.fm 에 가입하여 audioscrobbler 에 의해 수집된 프로파일로 개인화된 음악 방송을 즐긴다[4]. 이처럼 사용자들은 방대한 양의 데이터로부터 쉽게 콘텐츠를 추천받는다.

그러나 현재의 개인 맞춤형 콘텐츠의 서비스 제공자는 독점적으로 소유한 프로파일만을 수집하고 관리하여 사용자에게 서비스를 제공한다. 서비스 제공자들마다 독점적으로 소유한 많은 양의 프로파일들을 각각 관리하기 때문에 사용자들에게 더욱 다양한 서비스를 제공하지 못한다. 따라서 사용자의 현재 상황, 단말기의 능력, 개인 선호도 등 사용자와 관련된 프로파일 정보를 효과적으로 관리할 수 있는 방법이 요구된다[5].

본 논문에서는 시맨틱 웹에서 콘텐츠 추천을 위한 개인화 프로파일을 수집하고 갱신하여 배포하기 위한 방법을 제안한다. 개인화 프로파일은 프로파일의 특징을 담고 있는 프로파일 컬렉터, 다수의 프로파일 컬렉터로부터 정보를 수집하는 프로파일 수집기, 프로파일 특성에 기반한 고유의 프로파일을 해석하는 컬렉터 해석기

에 의해 생성된다. 이러한 개인화 프로파일을 사용하여 사용자의 선호도, 생활패턴에 따른 추천 검색 시스템을 설계한다. 개인화 프로파일을 추천 시스템에 사용하기 위해서는 선호도, 생활패턴에 대한 데이터를 구성하고 RDF로 정의한다. 또한 다수의 프로파일을 수집하여 획득한 데이터를 온톨로지 데이터베이스와 추론을 통해 개인화 프로파일을 갱신한다. 정보들은 SOAP 메시지를 통해 전송되며 콘텐츠 추천 서버에서 개인화 모듈과 추천 모듈에 의해 분석되어 추천리스트를 결정한다.

본 논문의 구성은 다음과 같다. 2장에서는 시맨틱 웹과 개인화에 대해 고찰하고, 3장에서는 본 논문에서 제안하는 콘텐츠 추천을 위한 개인화 프로파일에 대해 기술한다. 4장에서는 추천 시나리오에 따라 콘텐츠 추천 검색 시스템을 평가하고, 5장에서 결론을 맺는다.

II. 관련연구

1. 시맨틱 웹

기존의 웹에서 HTML을 이용한 표현방식은 사용자에게 문서의 내용을 그대로 보여준다. 그래서 사람이 아닌 프로그램 또는 소프트웨어 에이전트가 문서로부터 자동으로 의미를 추출하기가 어렵다. 예를 들면, 흔히 말하는 배가 음식으로서의 배인지 운송수단으로서의 배인지 구별하기가 힘든 것이다. 사용자가 원하는 내용과 무관한 많은 양의 내용으로 인해 웹의 검색을 효율적으로 사용할 수가 없었다. 따라서 표현은 같지만 뜻이 다른 데이터를 본래 의미로 해석하고 인식하며, 표준화된 정보를 제공해 줄 수 있는 시맨틱 웹이 연구되었다.

시맨틱 웹은 XML(eXtensible Markup Language)과 RDF(Resource Description Framework) 기술에 크게 의존한다. XML로 정의한 태그는 사람이 이해하기는 어렵지 않으나 기계가 자동적으로 그 의미를 이해하기는 어렵다. 특히 태그 사이의 의미 연관성을 추론하기란 매우 어렵다. 임의로 만든 태그의 사용법은 알 수 있어도 어떤 의미를 담고 있는 태그인지 파악하기는 어려운 것이 XML의 단점이다. RDF는 이런 문제를 해결하기 위해 제시된 기술이다. RDF는 특정 자원에 대한

숨은 자료를 설명하는 XML 기반의 프레임워크다.

RDF는 객체 지향적 접근의 지식표현 방식을 취하며, 세 개의 요소로 이뤄진 구조를 기본으로 한다. 즉, 개체(object)-속성(attribute)-값(value)의 구조를 갖고 있는데, A(O,V)로 표현된다. 그러나 일반적인 객체 지향적 혹은 프레임 방식과는 달리, RDF는 객체중심이 아니라 속성중심의 구조를 가진다[6].

자원의 속성이 포함된 메타데이터들 사이에는 의미(semantics), 구문(syntax), 구조(structure)의 세 가지 조건을 상호간에 정의할 수 있어야 한다. RDF는 서로 다른 의미, 구문, 구조를 가지는 메타데이터들이 공동의 협약을 이루어 상호간의 호환이 가능하다. RDF는 서로 다른 메타데이터들의 모든 의미를 정의하는 것이 아니라 각각의 메타데이터들이 제각기 필요한 데이터요소를 선언하여 사용한다. RDF는 XML을 이용하여 구문을 표현하기 때문에 SGML보다 간단한 형식으로 웹 환경에서 플랫폼과는 독립적으로 메타데이터를 구현할 수 있다. 시맨틱 웹 기술을 이용하면 문서에서 사용되는 요소의 의미와 문서 사이의 관련성 표시가 쉬워지고, 기계끼리 자동화 처리가 한결 쉬워진다[7].

2. 개인화

개인화는 이용자의 프로파일에 근거 하여 내용을 선정하고, 사용자에게 새롭거나 다른 이용자가 미처 생각하지 못했던 것을 채택하는 것이다. 즉, 사용자의 특성을 대상으로 하여 이와 관련한 다양한 통계의 분석과 비교를 통해 개별화 할 수 있는 지식 및 규칙을 찾아내는 과정이다. 이러한 개인화를 통해서 웹 사이트 운영자는 사용자에게 관한 자료를 얻고 사용자의 지속적인 이용이나 구매를 얻을 수 있으며 사용자는 자신에게 가장 알맞은 정보를 편리한 방법으로 얻을 수 있다[8].

대부분의 개인화 기술은 어느 정도 사용자의 개인 정보를 필요로 한다. 사용자 등록과정 중 인적사항이나 선호도에 관한 질문을 통해 정보를 얻는 방법도 있고 사이트 내에서 사용자의 행동을 관찰하여 정보를 얻는 방법도 있다. 이처럼 개인화 과정은 사용자의 선호도나 습관, 구매행태와 같은 정보를 수집하여 사용자에게 알맞은 정보를 제공하는 것이다. 웹 사이트 상에서의 개인화

는 웹 사이트에 들어오는 고객을 각 고객의 성향과 행태별로 세분화하여 콘텐츠를 보여주거나 서비스를 제공하는 것을 의미한다. 이는 고객의 충성도를 높여줄 뿐 아니라 데이(day) 마케팅과 원투원(one to one) 마케팅을 가능하게 해준다는 점에서의 의의가 크다. Amazon을 비롯한 전자 상거래 업체들은 사용자의 개인적 취향 및 구매기록에 따라 자신의 페이지를 구성하고 제품을 추천 받을 수 있는 기능들을 제공한다. 또한 어떤 사용자의 성향과 유사한 성향을 갖는 사용자 집단을 찾아 그 유사한 집단의 패턴정보를 근거로 하여 서비스를 제공한다[9].

최근에는 개인화에서 프로파일의 중요성이 부각되면서 다양한 프로파일을 표현할 수 있는 프레임워크의 설계가 시도되고 있다. Ubisworld[10]는 웹 기반 인터페이스를 통해 물리적 개체로 간주된 속성, 공간과 시간정보, 사용자 활동을 표현하는 기본적인 시맨틱 웹과 RDF 기반 데이터 구조를 제안한다. 그러나 이 시스템은 웹 기반 데이터 입력을 받을 뿐 상이한 자원으로부터 사용자 관련 정보의 수집은 고려되지 않았다. Haystack[11]은 개인정보의 통합관리환경의 제공을 목표로 모든 정보를 동일한 RDF로 표현한다. RDF 데이터, 온톨로지, 스키마 파일과 관계되는 시맨틱 사용자 인터페이스를 제공하여 정보 검색 기능이 향상된다.

III. 콘텐츠 추천을 위한 개인화 프로파일

많은 디지털 정보는 정적인 데이터와 동적인 데이터를 포함한다. 또한 광범위한 네트워킹, 컴퓨팅, 분산 콘텐츠 서비스와 함께 유비쿼터스 환경에서 처리된다. 개인화된 콘텐츠 추천 서비스를 사용자에게 제공하기 위해서 프로파일은 수집, 갱신, 배포되어야 한다.

1. 개인화 프로파일

1.1 구조

개인화 프로파일은 프로파일의 수집, 갱신, 배포를 위한 컬렉터, 수집기, 해석기에 의해 생성된다. 프로파일 컬렉터들은 프로파일의 특징을 담고 있으며 각 프로파일

일 컬렉터의 수집과 프로파일의 자원을 가리킨다. 프로파일 수집기는 다른 프로파일 컬렉터들로부터 다양한 프로파일을 수집하고 개인화 프로파일로 관리된다. 컬렉터 해석기는 프로파일 특성에 기반한 고유의 프로파일 컬렉터를 해석한다.

1.2 특징

네트워크를 통해 수집된 프로파일은 일반적으로 동일한 값을 가질 수 있지만 각 사용자들에 의해 다르게 해석될 수 있다. 본 논문에서 개인화 프로파일은 사용자의 상황과 선호도를 사용자의 목적에 따른 일반적인 프로파일에 의해 해석한다. 프로파일은 고도의 상관관계로 다른 프로파일과 동적으로 연결되어 있다.

사용자는 중앙집중형 데이터 저장소가 필요하다. 사용자의 홈서버가 있다고 가정하면 프로파일 수집기와 주기적으로 동기화하면서 일상생활의 유용한 데이터를 이용할 수 있다. 본 논문에서는 사용자 지식의 관리를 지원하고 사용자 정보와 히스토리를 제공한다.

수집된 프로파일은 정보공유 외에도 사용자들 사이에 상황 인식 서비스를 목적으로 생성하며 결정에 도움을 준다. 프로파일 수집기들의 통합은 새로운 서비스를 제공하는데 도움을 줄 것이다.

1.3 요구사항

개인화 프로파일은 실제 유비쿼터스 컴퓨팅 응용프로그램에서 효과적이며 다음과 같은 사항들이 요구된다.

프로파일 요소는 정적인 개인정보뿐만 아니라 동적인 사용자의 활동까지 공유하며 이를 표현하기 위해 개인화 프로파일은 개발자 요구에 따라 추가될 수 있다. 그러므로 개인화 프로파일을 위해서 기존 데이터 모델 활용과 새로운 특성의 추가 확장이 용이해야 한다.

개인화 프로파일은 다양한 프로파일 자원으로부터 수집에 의해 구축된다. 따라서 수집된 프로파일 정보는 잘 구성된 구조로 표현되어야 한다. 몇몇 표준 어휘는 응용프로그램 개발의 유비쿼터스 컴퓨팅 안에서 상호운용성을 유지하는데 필요하다. RDF는 어플리케이션 개발자에게 도움을 주며 상황인식 개인화 서비스를 제공할 수 있다.

개인화 프로파일은 기계에 의해 해석될 수 있으며 응용프로그램 소프트웨어는 공유하고 분석하여 사용자 프로파일 정보를 교환할 수 있다. 개인화 프로파일은 동적으로 변하는 사용자 활동 정보에 따라 다르게 구성된다.

1.4 RDF 개인화 프로파일

RDF는 상황정보 모델링을 위한 요소들이 확장된 지식 표현 언어이다. 그러므로 RDF는 본 논문에서 개인화 프로파일의 시맨틱 표현을 위해 사용된다. RDF의 기본 구조는 주어, 동사, 목적어에 해당하는 세 개의 정보(Subject, Predicate, Object)를 한 쌍으로 가진다. 이를 문장(statement)이라 하고 사람이나 웹 문서 등 특정 대상이 특정 속성에 대해 특정 값을 가지는 것을 표현한다.

개인화 프로파일에서 동사로 표현되는 속성의 타입, 각 속성 값의 범위와 RDF 스키마에서 대표적으로 도메인(domain)과 범위(range)라 불리는 주어진 속성의 자원 타입을 포함한 트리플 구조 기반 템플릿들의 쌍을 정의하였다.

RDF는 정적인 묘사를 위해 디자인되어 있으므로 동적으로 갱신을 할 수 없다. 따라서 본 논문에서는 프로파일 수집기에 프로파일의 동적인 갱신을 위한 모듈을 포함한다.

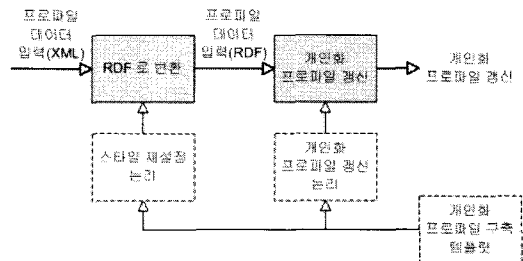


그림 1. 프로파일 수집기 안에서 개인화 프로파일의 갱신 모듈

개인화 프로파일의 갱신과 설계를 위한 절차는 [그림 1]과 같다. XML 형식으로 입력 받은 프로파일은 스타일 재설정 논리를 경유하여 RDF로 변환된다. 그리고 개인화 프로파일은 개인화 프로파일 갱신 논리를 경유하여 개인화 프로파일로 갱신된다. 스타일 재설정 논리

는 XSL로 설계한다. 개인화 프로파일 갱신 모듈은 트리플 추가, 정확한 값 갱신 그리고 트리플간 동일 표현의 통합 기능이 있다. 개인화 프로파일 갱신 모듈은 Jena[12] 기반으로 구현되었다.

RDF로 표현된 프로파일을 처리하기 위해 RDF 데이터 처리 언어(RDQL)[5]를 사용한다. RDQL은 응용프로그램 개발자에게 주어진 조건 하에서 처리해야만 하는 선언 부분 상태를 쓸 수 있는 방법을 제공한다. 그러므로 RDQL을 사용하여 타겟 상태 및 개인화 프로파일로부터 값을 처리한다.

만약 다중 응용프로그램이 실행중이며 프로파일 수집기에서 RDF 상태 접속을 한다고 가정한다. 프로파일 데이터 안의 변화는 서비스 호출의 트리거를 검출하기 위한 조건을 만족해야 한다. 그 후 상황기반 서비스 제공자는 프로파일 수집기로부터 트리거 정보를 받아 상황 정보에 접속한다.

2. 추천 검색을 위한 개인화 모듈

개인화 모듈의 구조는 [그림 2] 와 같다. 사용자가 쿼리 중재자에게 쿼리를 넘기면 바이너리로 변환되어 더 정확한 재정의의 위해서 개인화 관리자에게 파싱된 쿼리가 전달된다. 사용자 선호도 등의 정보는 사용자 프로파일 안에 저장된다. 프로파일에 의해 갱신된 쿼리는 평가를 위해 쿼리 중재자에게 반환된다. 갱신 쿼리는 개인화 관리자에 의해 재평가되고 통합되어 개인화 결과로써 사용자에게 전달된다.

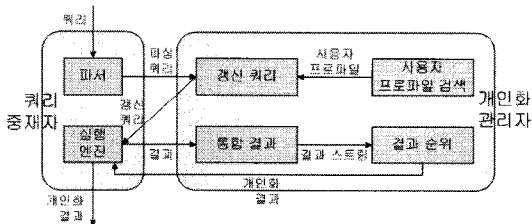


그림 2. 개인화 모듈

개인화 모듈의 시퀀스 다이어그램은 [그림 3] 과 같다. 최상위 시퀀스는 키워드 기반의 간단한 쿼리로 사용한다. 사용자는 쿼리 중재자에게 쿼리 메소드를 통해 쿼리를 넘긴다.

리를 넘긴다. 쿼리 중재자는 이 쿼리를 평가하고 개인화 관리자에게 개인화 쿼리 요청 메소드를 통해 사용자 프로파일에 따라 재순위화되고 검색된 결과를 호출한다.

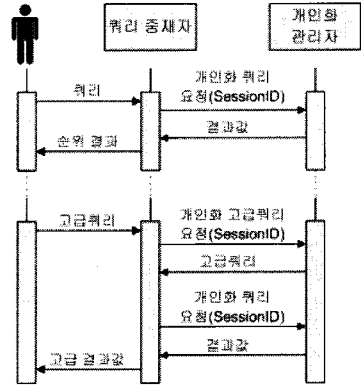


그림 3. 개인화 모듈 시퀀스 다이어그램

고급쿼리의 시퀀스 다이어그램은 더 복잡하다. 쿼리 중재자는 개인화 고급쿼리 요청 메소드를 사용하여 개인화 관리자로부터 개인화된 쿼리를 전달받는다. 이 메소드는 사용자 선호도와 최적화를 통해 쿼리를 개인화한다. 그 후 쿼리 중재자는 개인화된 쿼리를 평가하고 개인화 쿼리 요청 메소드를 호출하여 개인화 관리자로부터 결과를 전달받는다. 이 메소드는 사용자 프로파일 에 따라 결과를 순위화한다. 순위 결과는 최종적으로 사용자에게 반환된다.

IV. 개인화된 콘텐츠 추천 시스템

1. 콘텐츠 추천 시스템

본 논문에서 제안한 시스템은 선호도뿐 아니라 다양한 생활패턴이 저장되어 있는 개인화 프로파일을 생성하고 내용기반 및 협력적 필터링 기법을 사용하여 추천 리스트를 구성한다. 추천에 사용되는 내용기반 필터링에서는 먼저 콘텐츠 정보 DB에서 사용자의 취향과 사용자 정보에 해당하는 콘텐츠로만 검색하여 리스트를 구성하고 이 리스트로 초기 프로파일을 만든다. 초기 프로파일은 사용자가 콘텐츠 추천 서비스를 이용함에 따

라 이전의 히스토리와 현재 이용 중인 콘텐츠를 토대로 업데이트 된다. 이렇게 구성된 프로파일은 협력적 필터링의 통계적인 방법으로 분석되어 사용자의 콘텐츠 추천 서비스 요청 시 개인화 프로파일에 해당하는 콘텐츠 추천 리스트를 콘텐츠 정보 DB에서 추천 한다.

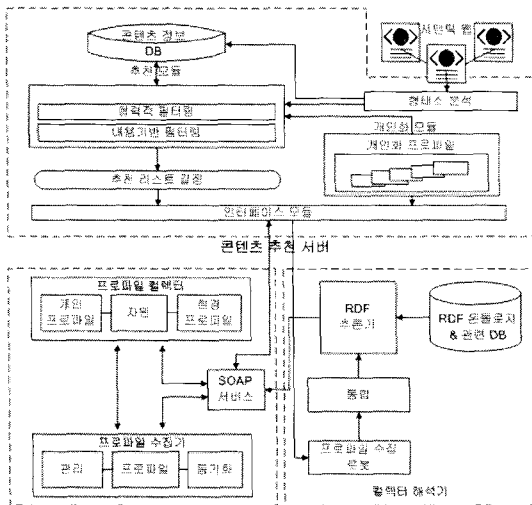


그림 4. 콘텐츠 추천 시스템

[그림 4]는 전체 시스템의 구성도이다. 본 논문에서 제안한 추천 검색 시스템은 시맨틱 웹으로부터 전달받은 여러 가지 데이터를 해석하고 추론한 뒤 정보화하여 확보된 사용자 프로파일을 콘텐츠 정보 DB와 함께 필터링을 거쳐 콘텐츠를 추천해주는 역할을 수행한다. 이를 위해 콘텐츠 추천 시스템은 크게 프로파일 컬렉터, 프로파일 수집기, 컬렉터 해석기, 콘텐츠 추천 서버로 구성된다.

콘텐츠 추천 서버는 이벤트가 발생했을 때 프로파일 컬렉터에 의해 생성된 정보들을 분석한 후 통합하여 컬렉터 해석기에 전달한다. 컬렉터 해석기는 분석된 데이터를 RDF 온톨로지 객체 DB들과 함께 RDF 추론기를 통해 컴퓨터가 이해할 수 있게 정보화시켜 프로파일 수집기에 넘긴다.

그 후 프로파일 수집기는 서버의 개인화 모듈과 동기화하며 개인화 프로파일을 생성한다. 개인화 모듈은 이벤트 발생시마다 사용자 개인화 프로파일을 갱신하고

추천 모듈에 전달한다. 추천 모듈에서는 생성된 프로파일을 형태소 분석하고 콘텐츠 정보 DB에서 검색할 때 내용기반 필터링, 협력적 필터링을 통해 최적의 콘텐츠 추천리스트를 결정하는 역할을 한다. 구성된 리스트는 콘텐츠 추천 결정에 의해 콘텐츠의 우선순위가 결정되고 새롭게 리스트가 구성되어 인터페이스 모듈에 전달된다. 전달된 리스트는 SOAP 서비스에 의해 응용프로그램에 전달된다. 또한 사용자가 선택한 콘텐츠 정보를 사용자 프로파일에 업데이트하는 기능도 수행한다.

2. 개인화 프로파일의 예

2.1 선호 콘텐츠 프로파일

사용자는 관심 있는 콘텐츠의 자세한 정보를 주식과 같이 프로파일로 저장된 것을 볼 수 있다. 사용자는 검색하고 있는 범위 안에서 콘텐츠의 자세한 정보를 봄으로써 추천리스트를 생성할 수 있다. 추천리스트는 선호 콘텐츠 프로파일로써 저장된다. 선호 콘텐츠 클래스와 주식 클래스는 프로파일 클래스의 서브 클래스로 정의한다.

선호 콘텐츠 프로파일의 기본 모델은 [그림 5]와 같다. 자원들은 선호 콘텐츠 프로파일에서 묘사되고 통합될 수 있다. 서로 다른 두 URI 들의 통합은 속성 값, ID, 콘텐츠 프로파일이 장착된 선호 콘텐츠 프로파일을 가져올 수 있다. 콘텐츠 추천 서버는 객체와 그 속성 정보를 선호 콘텐츠 프로파일이 추가될 때 시간과 위치 프로파일로써 빈도수를 저장할 수 있다. 관련 주식 정보는 선호 콘텐츠에 추가된다.

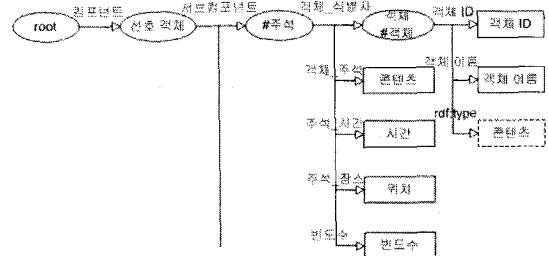


그림 5. 선호 콘텐츠 프로파일 모델

선호 객체는 화면에 표시되고 매 시간 사용자 레지스

터는 선호 객체, RDF 트리플 승인 또한 추가된다. 이 프로파일에서 객체 정보는 사용자 주석에 연결되어 시간-공간 정보 안에서 추가된다. 프로파일 사이의 링크는 다수의 프로파일 자원으로부터 수집되어 특정 조건에서 자세한 정보 처리를 가능하게 한다.

2.2 이용한 콘텐츠 프로파일

이 프로파일은 사용자의 일상생활 정보를 가리킨다. 콘텐츠 추천 서비스 시나리오에서 정보는 사용자 프로파일 관리자로부터 출판되고 교환된다. 승인 트리플은 이용한 콘텐츠 프로파일에 추가되고 시간과 위치 프로파일은 정보가 입력받은 프로파일로서 처음 받았을 때 추가된다. 이용한 콘텐츠와 일상생활 클래스는 프로파일 클래스의 서브클래스로 정의한다.

자원들은 이용한 콘텐츠 프로파일과 콘텐츠 프로파일 안의 자원들로 통합될 수 있다. 이 두 URI 의 통합은 ID 속성 값이 있을 경우 이용한 콘텐츠 프로파일의 콘텐츠 프로파일 안에 장착된다. 결과적으로 선호 콘텐츠와 이용한 콘텐츠는 ID를 통해서 링크 될 것이다.

2.3 선호 환경 프로파일

이 프로파일은 사용자의 관심 및 사용자의 위치 정보를 가리킨다. 이 프로파일은 요구되는 정보 데이터베이스 등에 접속 방법과 같은 사용자 관련 정보를 가리킨다. 선호 환경 클래스는 프로파일 클래스의 서브클래스와 현재위치 클래스로 정의한다.

3. 실험

3.1 서비스 시나리오

사용자는 콘텐츠의 추천리스트를 미리 편집하고 콘텐츠 추천 서버에 저장할 수 있다. 일반적으로 콘텐츠는 이전 콘텐츠 히스토리로부터 추론하여 자동으로 추천리스트에 추가된다. 사용자가 콘텐츠를 이용할 때 개인화 프로파일에 기반하여 추천된다. 이 추천 정보는 사용자의 스케줄에 저장되어 이용 가능한 시간에 화면에 표시된다. 이용중인 콘텐츠의 정보는 콘텐츠 추천 서버에 저장되고 사용자 선호도에 따라 정보를 얻게 된다. 그리고 사용자는 쉽게 접할 수 있는 잠재적인 콘텐츠를 찾을 때 참고할 수 있다. 콘텐츠 추천 서버는 콘텐츠를 이용

할 때 자동적으로 추천리스트를 확인한다. 콘텐츠 추천 서버는 계획했던 콘텐츠의 이용을 하지 않고 그만두려고 한다면 사용자에게 경고 메시지를 보낸다.

3.2 실험결과

본 논문에서 제안한 콘텐츠 추천 검색 시스템을 평가하기 위해서 서비스 시나리오에 따라 실험한다. 이때 사용자의 개인화 프로파일은 콘텐츠 추천 서버에 전달되어 콘텐츠 추천리스트를 제공하는지 확인한다. 시맨틱 웹에서 RDF로 정의된 프로파일을 사용하였으며 콘텐츠 추천 서버는 데스크톱 PC에 구현한다. 콘텐츠 정보 DB는 가요 500곡으로 구성한다.

아이디가 User 라는 사용자가 PC에 앉아 콘텐츠 추천 검색을 요청한다. PC는 홈 서버를 경유하여 선호도와 생활패턴이 기록되어 있는 프로파일로부터 현재 정보를 가져와 개인화 모듈에 의해 개인화 프로파일로 바뀐다. 개인화 프로파일은 프로파일 수집기를 통해 콘텐츠 추천 서버에 전달된다.

[그림 6]은 사용자의 선호도, 생활패턴에 따른 콘텐츠 추천리스트다. 초기의 프로파일들은 사용자가 콘텐츠 정보 DB에서 선호도, 생활패턴에 따라 선택한 콘텐츠로 따로 구성하고 쿼리를 요청시 선택된 콘텐츠 리스트로 프로파일을 추가로 구성한다. 이때 프로파일은 선호도, 생활패턴이 유사할 때 사용자가 선택한 콘텐츠 리스트가 된다. User는 20대이며 선호 장르가 발라드인 것을 playlist로 추천받는다.



그림 6. 개인화 프로파일을 이용한 콘텐츠 추천리스트

추천 검색 방법으로는 프로파일을 형태소 분석하여 파도란 단어의 출현 빈도를 계산한다. 그 다음 빈도가 가장 높은 단어와 콘텐츠 정보 DB에서 핵심어가 파도와 일치하는 콘텐츠를 검색하고 검색 결과를 리스트로 구성한다. 이와 같은 작업을 프로파일 별로 행한다. 또한 리스트의 순위는 요청된 쿼리의 콘텐츠 리스트 중 빈도가 가장 높은 단어와 콘텐츠 정보 DB의 핵심어가 일치하는 콘텐츠가 가장 우선순위가 높다. 그리고 연령, 선호도, 생활패턴 프로파일 순으로 제한하여 리스트를 구성한다. 리스트 중 중복되는 콘텐츠는 가장순위가 높은 콘텐츠만 남기고 모두 삭제된다. 또한 사용자에게 의해 선택된 콘텐츠는 개인화 프로파일에 기록되어 선택이 반복될수록 사용자에게 적합한 콘텐츠 리스트를 서비스할 수 있게 한다.

추천리스트의 상위에는 본 상황에서 사용자가 선택했던 콘텐츠들의 대표 단어와 일치하는 핵심어를 가지는 콘텐츠들이 리스트 되고 중간에는 개인화 프로파일과 관련된 콘텐츠들이 리스트로 구성된다. 실험결과 선호도와 다양한 생활패턴의 정보가 기록된 개인화 프로파일이 반영된 추천리스트가 출력됨을 확인할 수 있다.

V. 결론

기존의 추천 시스템은 단순 질의어로 음악을 검색하거나 사용자의 취향을 수동적으로 입력 받아 그 데이터를 분석해 추천해주는 시스템이다. 이러한 시스템은 데이터마이닝 기법으로 웹 접속 기록 및 유형과 사용자가 요구하는 정보만을 분석하여 콘텐츠 추천시 중요한 사용자의 생활패턴에 대해 고려하지 못했다. 또한 시스템과의 상호교류가 빈번해야 하므로 서비스를 이용하는 사용자에게 많은 불편을 가져왔다.

본 논문에서는 사용자의 프로파일 정보를 능동적으로 획득, 인식하고 이 정보를 콘텐츠 추천 및 검색에 사용하여 개인화된 콘텐츠를 서비스할 수 있는 시스템을 설계하였다. 상이한 프로파일 정보를 사용하기 위해 프로파일을 RDF로 정의하고 통계적 방법의 내용기반 필터링과 협력적 필터링을 응용하였다. 데이터 마이닝을 사

용함에 따라 지식의 재사용이 가능하며 정보들 간에 관계나 패턴을 유기적으로 도출할 수 있었다.

RDF의 사용은 프로파일 공유를 통한 서비스를 위해 중요한 요구조건이다. 그러나 RDF만으로는 조직적으로 정의된 어휘를 위해 많은 수집 데이터 모델을 결정하기에는 어렵다. 따라서 본 논문은 콘텐츠 추천 검색 시스템을 설계하여 콘텐츠 추천 시나리오에 따라 RDF로 표현된 개인화 프로파일을 이용하면서 유용한 모델을 제시했다. 또한 몇 가지 프로파일은 관련 프로파일의 자원으로부터 프로파일 수집 사이에 링크로 연결되어 있다는 것을 알 수 있었다.

그러나 제안한 프레임워크는 다양한 상황인식 서비스와 실례 분석 및 피드백을 반드시 요구하는 프로토타입을 통해서 평가되어야 한다. 프로파일을 이용한 추천 시스템들은 많은 이점에도 불구하고 사용자들과 프로파일의 수집 사이에서 발생하는 것에 대한 프라이버시 문제를 야기시킬 수 있다. 따라서 프로파일 정보의 사용은 반드시 콘텐츠 추천 서비스를 이용하기 위한 서비스 제공자와 함께 교환이 이루어져야 하며 권위있는 협회를 통해 표준안이 마련되어야 할 것이다. 또한 연구[13]에서 필요성을 제시했듯이 사용자가 프라이버시의 제어를 가능하게 하는 툴, 정책 기반 매칭 엔진 개발이 이루어져야 한다.

참고 문헌

- [1] Tjoa, A.M., Andjomshooa, A., Shayeganfar, F., and Wagner, R., "Semantic Web challenges and new requirements," Database and Expert Systems Applications, 2005.
- [2] Kobayashi, I., and Saito, M., "A Study on Information Recommendation System that Provides Topical Information Related to User's Inquiry for Information Retrieval," Web Intelligence and International Agent Technology Workshops, pp.385-388, 2006.
- [3] <http://www.bugs.co.kr/>

[4] <http://www.last.fm/>

[5] Xujuan Zhou, Sheng-Tang Wu, Yuefeng Li, Yue Xu, Raymond Y.K. Lau, and Peter D. Bruza, "Utilizing Search Intent in Topic Ontology-based User Profile for Web Mining," Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, 2006.

[6] D. Beckett, "RDF/XML Syntax Specification," W3C, 2004.

[7] Lin, Lv., Yu-Shu, and Liu., "Research of English Text Classification Methods Based on Semantic Meaning," Information and Communications Technology : ITI 3rd International Conference, 2005.

[8] J. Pitkow, H. Schutze, T. Cass, R. Cooley, D. TurnBull, A. Edmonds, E. Adar, and T. Breuel, "Personalized Search," Communications of the ACM, Vol.45, No.9, pp.50-55, 2002.

[9] <http://www.amazon.com/>

[10] <http://www.ubisworld.org/>

[11] D. Huynh, D. R. Karger and D. Quan, "Haystack: A Platform for Creating, Organizing and Visualizing Information Using RDF," In Proc. Of Semantic Web Workshop, 2002.

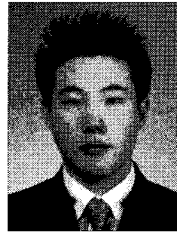
[12] <http://jena.sourceforge.net/>

[13] Kolari, P., Li Ding, Shashidhara, G., Joshi, A., Finin, T., and Kagal, L., "Enhancing Web privacy protection through declarative policies," Policies for Distributed Systems and Networks, Sixth IEEE International Workshop, pp.57-66, 2005.

저자 소개

송창우(Chang-Woo Song)

준회원

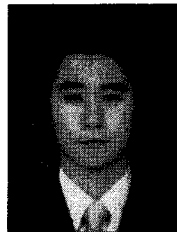


- 2004년 8월 : 한국교육개발원 학점은행제 컴퓨터공학 전공(공학사)
- 2007년 2월 : 인하대학교 컴퓨터·정보공학과(공학석사)
- 2007년 3월 ~ 현재 : 인하대학교 정보공학과 박사과정

<관심분야> : 임베디드/유비쿼터스 시스템, 상황인식, 데이터마이닝

김종훈(Jong-Hun Kim)

준회원

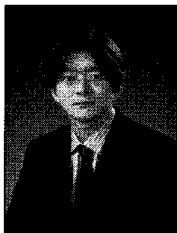


- 2001년 2월 : 인천대학교 물리학과 (학사)
- 2003년 2월 : 인하대학교 전자계산공학과(공학석사)
- 2007년 8월 : 인하대학교 컴퓨터 정보공학과 박사수료

<관심분야> : 음성인식, 임베디드/유비쿼터스 시스템, 데이터마이닝, 인공지능

정경용(Kyung-Yong Chung)

정회원



- 2000년 2월 : 인하대학교 전자계산공학과(공학사)
- 2002년 2월 : 인하대학교 컴퓨터 정보공학과(공학석사)
- 2005년 8월 : 인하대학교 컴퓨터 정보공학과(공학박사)

- 2005년 9월 ~ 2006년 2월 : 한세대학교 IT학부 교수
- 2006년 3월 ~ 현재 : 상지대학교 컴퓨터정보공학부 교수

<관심분야> : 데이터마이닝, 지능시스템, 인공지능

류 중 경(Joong-Kyung Ryu)

정회원



- 1988년 2월 : 한국방송통신대학교 전자계산학과(이학사)
- 1991년 2월 : 인하대학교 산업대학원 정보공학과(공학석사)
- 1983년 ~ 1991년 : 대림산업 정보시스템실 대리

- 2003년 2월 : 인하대학교 컴퓨터정보공학과 박사수료
- 1992년 3월 ~ 현재 : 대림대학 컴퓨터정보계열 부교수
<관심분야> : 소프트웨어공학, HCI, ERP, CRM, 유비쿼터스

이 정 현(Jung-Hyun Lee)

정회원



- 1977년 2월 : 인하대학교 전자과(공학사)
- 1980년 9월 : 인하대학교 전자공학과(공학석사)
- 1988년 2월 : 인하대학교 전자공학과(공학박사)

- 1979년 ~ 1981년 : 한국전자기술 연구소 시스템 연구원
- 1984년 ~ 1989년 : 경기대학교 전자계산학과 교수
- 1989년 1월 ~ 현재 : 인하대학교 컴퓨터공학부 교수
<관심분야> : 자연어처리, HCI, 음성인식, 정보검색, 고성능 컴퓨터구조