동적 프로파일을 이용한 시맨틱 웹 기반의 맞춤 뉴스 서비스

추은하, 장은실, 이용규
동국대학교 컴퓨터공학과
e-mail:esjang@dongguk.edu

A Semantic Web-based Personalized News Service Using Dynamic Profiles

Eun Ha Choo, Eun Sill Jang, Yong Kyu Lee
Dept. of Computer Engineering, Dongguk University

요 약

기존의 뉴스 서비스는 정적 프로파일을 사용하여 고정된 관심분야 만을 서비스하기 때문에 관심이 바뀌었을 경우에는 이를 쉽게 반영하지 못하는 문제점이 있다. 따라서 본 논문에서는 이러한 문제점을 해결하기 위하여 프로파일을 동적으로 관리하여 개인의 관심이 바뀌어가는 것을 바로 반영할 수 있도록 하고, 정보 간의 의미를 파악하여 관련 정보를 쉽게 찾을 수 있도록 도와주는 시맨틱 기술을 적용한 맞춤 뉴스 서비스 시스템을 설계 및 구현한다. 그 결과, 사용자의 변화된 관심에 따른 맞춤 뉴스 서비스를 제공할 수 있다.

1. 서론

기존 웹에서의 개인별 맞춤 뉴스 서비스는 개인이 직접 관심사항을 선택하여 그에 해당하는 분야의 뉴스만을 서비스하였다. 이러한 관심사항은 시간이 지남에 따라 변할 수 있고, 처음 선택할 때 적당히 선택할 수 있어 실제의 관심사항과는 멀어지는 경우도 있다. 그러므로 최근에 개인이 관심있어 하는 뉴스 정보를 제공하기 어려워지는 문제가 있다.

한편, 시맨틱 웹은 웹상의 정보에 잘 정의된 의미를 부여함으로써 컴퓨터도 문서의 의미를 쉽게 이해할 수 있도록 하여 의미론적인 정보 검색을 가능하게 한다[3]. 이러한 의미를 지닐 수 있도록 하는 언어로는 온톨로지를 사용한다.

본 논문에서는 시맨틱 웹 기반에서의 동적 프로파일을 이용한 맞춤 뉴스 서비스 시스템을 설계 및 구현한다.

즉, 시맨틱 웹을 기반으로 하여 개인의 관심이 변하게 되는 경향을 사용자 프로파일에 동적으로 반영할 수 있도록 서비스를 설계하여 기존의 맞춤 서비스의 문제점을 해결하고 나아가 보다 더 편리한 맞춤 서비스가 가능한 시스템을 제공하고자 한다.

2. 관련 연구

본 절에서는 시맨틱 웹 관련 기술과 관심 뉴스를 위한 추천 방법, 웹 기반 뉴스 정보 제공 서비스에 관하여 소개한다.

2.1 시맨틱 웹 관련 기술

시맨틱 웹은 웹의 정보에 의미를 부여하기 때문에 컴퓨터도 정보를 쉽게 이해하여 사용자가 원하는 정보를 보다 정확하게 얻을 수 있도록 한다. 이렇게 정보에 의미를 담을 수 있도록 하는 주요 기술로서 온톨로지를 사용하며, 본 논문에서는 OWL(Web Ontology Language)이라는 웹 온톨로지를 사용한다[2].

OWL은 시맨틱 마크업 언어로서, 체계적인 온톨로지 구축을 지원할 수 있으며, XML이나 RDF, RDF-S 보다도 더 많은 의미 표현 수단을 사용하여 컨텐츠를 작성할 수 있어 이들보다 뛰어나다고 할 수 있다[5].

온톨로지의 의미를 파악하기 위해서는 추론을 해야 하는데, 본 논문에서는 추론을 위하여 OWLJessKB(Knowledge Base)를 사용한다. OWLJessKB는 OWL로 이루어진 파일을 Jess(Java Expert System

Shell)를 사용하여 시맨틱 추론을 하는 도구이다[6].

본 논문에서는 OWLJessKB를 사용자 검색 뉴스의 키워드 정보를 가지고 OWL로 이루어진 온톨로지에서 유사 어휘 정보를 찾기 위한 추론 과정에 사용하고자 한다.

2.2 관심 뉴스를 위한 추천 방법

추천 방법에는 사용자의 나이, 성별 등과 같은 인구통계학적인 정보에 의한 추천 방법과, 사용자의 프로파일과 문서 간의 단어 정보를 이용하여 유사도를 구하는 내용기반 추천 방법[1], 여러 사용자들의 선호도를 조사하고 특정 사용자와 유사한 정보를 갖는 다른 사용자들의 정보를 추천하는 협업기반 추천 방법[4] 등이 있다. 본 논문에서는 사용자의 관심 뉴스 추천을 위해 협업기반 추천 방법을 적용한다.

2.3 웹 기반 뉴스 정보 제공 서비스

현재 웹에서 이메일을 통하여 개인별로 제공하는 뉴스 서비스로는 드림인터페이스에서 국내외 주요기사를 관심분야별로 제공하는 매일매일 웹진서비스[7]나 삼성 경제 연구소에서 제공하는 SERIZINE[8] 등이 있다. 이러한 서비스들은 처음 사이트에 가입 시 프로파일의 관심항목을 선택하고, 선택된 관심항목의 뉴스만을 서비스하게 되는데 개인의 관심이 바뀌었을 경우나 처음의 선택을 허위로 작성하였을 경우에는 직접 다시 변경해야 하는 번거로움을 거쳐야 한다.

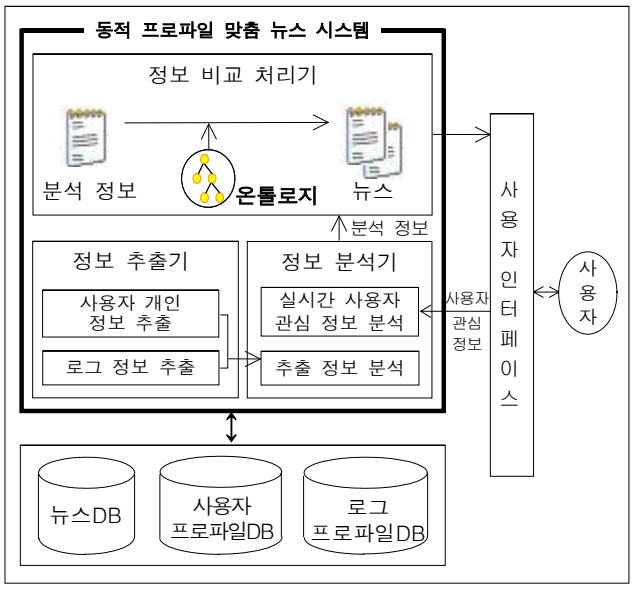
3. 맞춤 뉴스 서비스 설계

본 절에서는 맞춤 뉴스 서비스의 개요를 살펴보고, 맞춤 뉴스에 필요한 동적 프로파일을 설계한다. 시맨틱 웹을 구현하기 위하여 ITNews 온톨로지를 기술하고, 온톨로지의 추론과정을 살펴본다.

3.1 맞춤 뉴스 서비스 개요

맞춤 뉴스를 제공하기 위하여 프로파일을 동적으로 설계하며, 시맨틱 웹의 기술을 적용하여 사용자가 관심있는 뉴스를 보다 정확하게 찾을 수 있도록 한다.

다음 [그림 1]은 동적 프로파일 맞춤 뉴스 시스템의 구성도이다. 동적 프로파일 맞춤 뉴스 시스템에서는 정보 추출기, 정보 분석기, 그리고 정보 비교 처리기로 구성된다. 정보 추출기에서는 사용자 및 로그 프로파일 데이터베이스로부터 필요한 정보를 추출하며, 정보 분석기에서는 추출된 정보와 사용자가 웹 브라우저에서 입력한 정보를 모니터링하여 이를 분석하고 뉴스를 검색하는데 필요한 정보를 수집한다. 정보 비교 처리기에서는 정보 분석기에서 모아진 분석정보를 뉴스 온톨로지와 비교하여 관련된 어휘를 찾은 후 관련 뉴스를 검색한다.



[그림 1] 동적 프로파일 맞춤 뉴스 시스템 구성도

3.2 동적 프로파일

사용자 프로파일은 사용자의 기본 정보와 관심 정보를 담고 있다. 동적 프로파일은 한번 기재된 사용자 프로파일에서 시간의 흐름에 따라 바뀔 수 있는 사용자의 관심 정보를 관리해 주는 기능을 한다.

본 논문에 적용된 사용자 프로파일의 구성은 이름, 아이디, 비밀번호, 관심 뉴스 목록, 추천 뉴스 목록으로 이루어졌으며, 관심 뉴스 목록과 추천 뉴스 목록의 세부 항목으로는 모바일, 인터넷, 보안, 교육, 게임, 기타 등으로 이루어졌다. 관심 뉴스 목록은 사용자가 관심있는 분야를 선택할 수 있도록 하였으며, 관심 뉴스 목록과 추천 뉴스 목록은 시스템에서 동적으로 관리할 수 있다.

3.2.1 사용자 정보 분석

프로파일을 동적으로 관리하기 위해서는 사용자의 관심 정보를 분석하여야 한다. 먼저 프로파일의 관심 뉴스 정보를 사용자 프로파일 데이터베이스에서 추출하고, 추출한 정보와 사용자가 실시간으로 검색하는 뉴스에서 키워드 정보를 추출한다. 키워드 정보는 온톨로지를 통하여 유사한 어휘를 분석하고, 분석된 어휘와 키워드가 일치하는 뉴스 문서를 검색한다. 검색하여 나온 뉴스 문서는 사용자에게 추천 문서로 제공한다.

[그림 2]는 일정기간동안 사용자가 검색한 뉴스 문서를 분석하고 관심 뉴스 목록과 추천 뉴스 목록을 처리하는 과정을 나타낸 알고리즘이다. 일정기간 동안 사용자가 관심을 가지고 검색하는 뉴스 분야를 분석하여 사용자의 관심 정도를 정한다. 관심의 정도에 따라 프로파일에 정의되어 있는 관심 뉴스 목록은 추천 뉴스 목록으로, 추천 뉴스 목록은 일반 뉴스 목록이나 관심 뉴스 목록으로 변경한다.

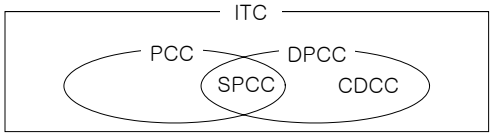
```
일정기간정보 분석 () {
  while (day < 일정기간) {    /*일정기간동안 */
    switch (클릭문서) {      /* 사용자가 클릭한 문서에 따라서 */
      case '인터넷':         /* 해당 클릭 문서의 목록을 */
        인터넷카운트++;  /* 카운트한다 */
      case '모바일':
        모바일카운트++;
          :
      default:
        break;
    }

   인터넷카운트평균값 = 인터넷카운트/일정기간 * 100;
   목록조정(인터넷카운트평균값, 목록범위);
   모바일카운트평균값 = 모바일카운트/일정기간 * 100;
   목록조정(모바일카운트평균값, 목록범위);
          :
   기타카운트평균값 = 기타카운트/일정기간 * 100;
   목록조정(기타카운트평균값, 목록범위);

   목록조정(평균값, 목록범위) {
   if (평균값 > 일정기준1){   //목록의 상향조정을 위한 기준을 정함
       if (목록범위 == 관심목록) 목록범위 유지;
       else if (목록범위 == 추천목록) 목록범위 = 관심목록;
       else if (목록범위 == 일반목록) 목록범위 = 추천목록;   }
   else if (평균값 < 일정기준2){ //목록의 하향조정을 위한 기준을 정함
       if (목록범위 == 관심목록) 목록범위 = 추천목록;
       else if (목록범위 == 추천목록) 목록범위 = 일반목록;   }
  }
 }
}
```

[그림 2] 관심 뉴스 목록 처리과정

[그림 3]은 협업 기반에서 추천 뉴스 목록의 선택을 위한 관계를 나타낸다.



- ITC(ITNews Category) : ITNews의 하부목록
- PCC(Personal Concern Category) : 개인의 관심 뉴스 목록
- DPCC(Different Personal Concern Category) : 비슷한 관심을 갖는 다른 사람들의 관심 뉴스 목록
- SPCC(Similarity Personal Concern Category) : DPCC와 PCC의 공통 관심 뉴스 목록
- CDCC(Common Different Concern Category) : DPCC들 간의 SPCC를 제외한 공통 관심 뉴스 목록

[그림 3] 협업 기반의 추천 뉴스 목록 선택 관계

그림에서 보이는 CDCC는 추천 뉴스 목록의 대상이 되어 일정기간동안 사용자가 관심을 보이는 정도에 따라 관심목록이나 일반목록으로 변경이 이루어진다. 이렇게 협업기반의 추천 방법을 사용하면 사용자의 관심의 범위를 넓힐 수 있는 장점이 있다.

3.3 ITNews 온톨로지 정의 및 추론

시맨틱 웹을 구현하기 위해서는 중요한 기술로 온톨로지가 있다. 온톨로지는 어휘마다 의미를 부여하여 어휘간의 관계를 컴퓨터도 파악할 수 있도록 한다. 본 논문에서는 IT 관련 뉴스에 해당하는 어휘만을 모아서 ITNews 온톨로지를 작성하였다.

[그림 4]는 ITNews 온톨로지의 일부이다.

```
<!-- ITNews 클래스 선언 -->
<owl:Thing rdf:ID="ITNews"/>

<!--ITNews의 하위 클래스로 Mobile 클래스 선언-->
<owl:Class rdf:ID="Mobile">
      <rdfs:subClassOf rdf:resource="#ITNews"/>
      <owl:oneOf rdf:parseType="MCollection">
            <owl:Thing rdf:about="#service"/>
            <owl:Thing rdf:about="#telecom"/>
            <owl:Thing rdf:about="#mobilephone"/>
       </owl:oneOf>
 </owl:class>

<owl:Class rdf:ID="cellularphone">
       <owl:equivalentClass
        rdf:resource="&ITNews;mobilephone"/>
<owl:Class>
```

[그림 4] ITNews 온톨로지의 일부

ITNews의 클래스를 선언하였으며, ITNews의 하위 클래스로 Mobile이 있으며, Mobile 내에 MCollection이라는 이름으로 service, telecom 그리고, mobilephone이 요소로 있으며, cellularphone과 mobilephone은 같은 의미의 클래스임을 나타내고 있다.

이러한 온톨로지의 의미를 파악하기 위해서는 추론을 해야 하며, [그림 5]는 추론 질의 예이다.

ITNews의 mobile 범위 안에서 제시어인 telecom과 유사한 의미의 클래스를 찾는 질의에서 telecomcompany와 telecom은 유사하다는 결론을 얻는다.

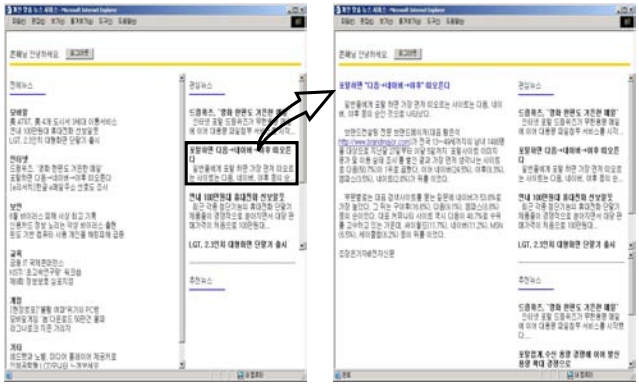
```
(defquery sim_word
     (PropertyValue &ITNews:mobile ?a(=(simility(class
      ?a)) telecom)))
=>
   telecomcompany와 telecom은 유사하다
```

[그림 5] 추론 질의 예

4. 구현 및 실험 평가

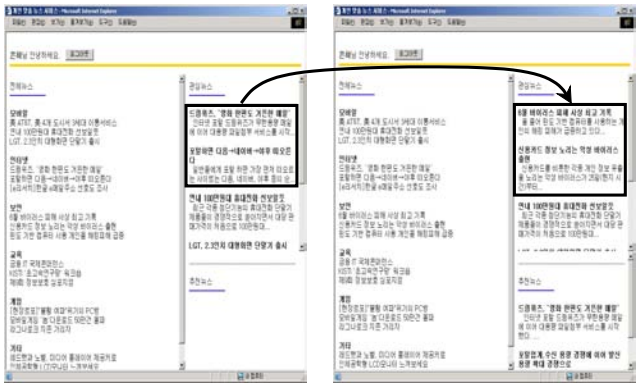
구현 환경으로 운영체제는 Windows 2000 Server, 데이터베이스 서버는 MS-SQL Server 2000, 개발언어는 Java, XML, JSP, OWL 등을 사용하였다.

사용자 뉴스 검색 인터페이스는 [그림 6]과 같다. 왼쪽 그림은 사용자가 처음 접하는 화면이며, 기사를 클릭하면, 오른쪽 그림처럼 화면의 왼편에 기사 전문을 보여주는 형식이다. 또한 검색된 기사의 관련 뉴스로 화면의 오른편 하단에 추천 뉴스를 보여주고 있다.



[그림 6] 사용자 뉴스 및 추천 뉴스 제공 인터페이스

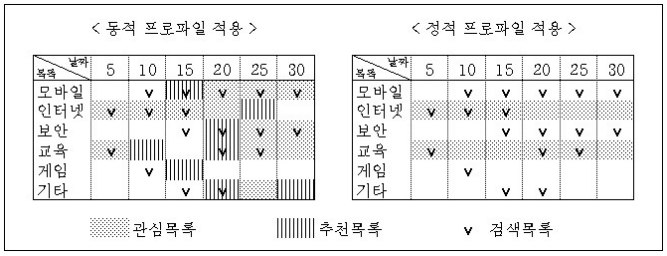
[그림 7]은 동적 프로파일을 이용하여 사용자의 관심 목록이 변하여 반영된 인터페이스이다.



[그림 7] 변경된 관심 목록이 반영된 인터페이스

왼쪽 그림의 사용자 관심 뉴스인 인터넷 관련 뉴스가 오른쪽 그림의 보안 관련 뉴스로 바뀌어 있음을 볼 수 있다. 또한 오른쪽 그림의 하단에는 기존의 관심 뉴스였던 인터넷 관련 뉴스들이 추천 뉴스로 바뀌어 있음을 볼 수 있다. 이렇게 동적 프로파일을 이용하여 관심 뉴스와 추천 뉴스를 사용자의 관심에 맞추어 변경을 하고, 한 화면에서 바로 보여주게 됨으로써 사용자에게 편리함을 제공한다.

한편, [그림 8]은 동적 프로파일과 정적 프로파일을 각각 적용한 뉴스 서비스를 비교하여 보여주고 있다.



[그림 8] 동적 프로파일과 정적 프로파일을 적용한 관심목록 비교

일정기간을 5일로 설정하여 5일동안 사용자가 검색한 목록을 모니터링하여 5일째에는 해당 기간동안 사용자가 관심을 가지고 검색한 목록을 기록한다. 'v'는 사용자가 관심을 가지고 검색한 목록을 표시한 것이며, '▒'는 사용자에게 제공한 관심목록을, '|||'는 사용자에게 제공한 추천목록을 표시한 것이다.

동적 프로파일을 적용한 사용자 프로파일의 관심목록은 처음에는 사용자가 정의한데로 '인터넷'과 '교육'만이었으나, 점차 사용자의 관심사항이 변하게 되고, 이러한 변화를 적용하여 관심을 가졌던 목록을 추천하게 되고 지속적으로 관심을 보인 목록을 관심목록으로 변경하여 마지막에는 관심목록이 '모바일'과 '보안,' '교육'으로 변화된 것을 볼 수 있다. 정적 프로파일을 적용한 사용자 프로파일은 사용자가 관심을 가지고 검색한 목록과는 상관없이 처음 관심목록으로 정하였던 '인터넷'과 '교육'만을 지속적으로 관심목록으로 제공하고 있음을 볼 수 있다.

5. 결론 및 향후 연구

본 논문에서는 시맨틱 웹 기반의 동적 프로파일을 이용하여 맞춤 뉴스 서비스 시스템을 설계 및 구현하였다.

동적 프로파일을 이용하여 처음에 선택된 관심사항만을 반영하는 기존의 뉴스 서비스에 비해 관심사항을 유동적으로 반영함으로써 폭넓은 분야의 기사를 편리하게 접할 수 있게 하고, 관심 뉴스 목록과 추천 뉴스 목록을 구분하여 서비스하게 됨으로써 관심분야를 좀더 빠르게 접할 수 있게 되었다. 또한, 시맨틱 웹에서의 온톨로지를 통한 각 요소의 관계정의와 추론 규칙을 세우고, 이를 이용하여 유사한 정보를 좀더 정확하게 찾을 수 있게 되었다.

향후 연구로는 다양한 분야의 온톨로지를 정의함으로써 검색 가능한 분야를 늘릴 수 있도록 하겠다.

참고문헌

[1] C. Basu, H. Hirsh, and W. Cohen, "Recommendation As Classification: Using Social And Content-Based Information," Proc. of the 14th International Conference on Artificial Intelligence, pp. 714- 720, Madison, Wisconsin, USA, 1998.
[2] S. Decker, P. Mitra, and S. Melnlk, "Framework for the Semantic Web: an RDF tutorial," IEEE Internet Computing, Vol. 4, No. 6, pp. 68-73, 2000.
[3] T. B. Lee, J. Hendler, and O. Lassila, The Semantic Web, Science American, 2001.
[4] P. Melville, R. J. Mooney, and R. Nagrajan. "Content-Boosted Collaborative Filtering," Proc. of the SIGIR-2001 Workshop on Recommender Systems, pp. 187-192, Edmonton, Canada, 2002.
[5] OWL, http://www.w3.org/TR/owl-features/, 2004.
[6] OWLJessKB, http://edge.cs.drexel.edu/assemblies/owljsskb/, 2004.
[7] 드림인터페이스, "매일매일웹진," http://www.mailmail.co.kr, 2004.
[8] 삼성경제연구소, "SERIZINE," http://www.seri.org, 2004.