

온톨로지 기반 웹 콘텐츠 추천 기법

김제민^o, 박영택
숭실대학교 컴퓨터학과

kimjemins@hotmail.com^o, park@computing.ssu.ac.kr

Web Contents Recommendation based on Ontology

Je-Min Kim^o Young-Tack Park

Dept of Computer Science, Soongsil University

요 약

추천 시스템은 사용자 프로파일을 기반으로 개인 취향에 맞는 정보나 제품에 대한 이용성을 향상 시킨다. 본 논문에서는 시맨틱 환경 내에서 사용자 개개인에 맞는 웹 콘텐츠를 제공하기 위한 온톨로지 기반의 웹 콘텐츠 추천 방법론을 제안한다. 이를 위해서 2가지에 초점을 두었다. 첫 번째, 사용자 프로파일의 쓰임새를 향상시키기 위해 온톨로지 모델을 적용한다. 이는 비슷한 서비스를 제공하는 여러 웹 서비스 사이트에서 사용자의 기호 정보를 공유할 수 있다는 이점을 갖는다. 또한 온톨로지를 기반으로 생성된 사용자 프로파일은 콘텐츠 추천 점수 계산을 위한 정확한 입력 데이터를 제공한다. 두 번째로 각각의 웹 콘텐츠들의 추천 점수를 계산하는 함수를 정의한다. 제안하고자 하는 함수는 각 웹 콘텐츠의 계층구조와 웹 콘텐츠를 구성하는 속성들의 관계를 명시한 온톨로지를 기반으로, 사용자 프로파일의 내용과 웹 콘텐츠의 개념 유사도(Concept Similarity)와 관계 유사도(Relation Similarity) 구한다. 따라서 본 논문에서는 전체 유사도(Concept Similarity+Relation Similarity)를 추천 점수로 적용한다.

1. 서 론

현재 HTML에 기반 된 웹 정보가 방대해진 상황에서 키워드 입력 기반의 웹 검색 엔진은 주로 단어의 빈도수나 어휘 정보를 이용하여 문서의 유사도를 측정하고 랭킹을 매기기 때문에 사용자의 의도와는 관계없이 많은 문서를 결과로 출력한다. 이로 인하여 사용자는 불필요한 정보를 걸러 내느라 시간을 낭비하게 된다. 시맨틱 웹은 현재의 웹을 확장하여 웹에 올라오는 정보에 정형화된 의미를 부여함으로써 컴퓨터와 사람이 협동적으로 작업을 수행할 수 있도록 프레임워크를 제공한다[1]. 따라서 현재 웹 서비스 시스템에 시맨틱 웹 기술을 적용하면, 사용자가 특정 정보를 검색할 때 더욱 정확한 결과를 가져올 뿐만 아니라 서로 다른 이형질 소스의 정보의 통합 및 관리가 가능하다.

시맨틱 웹의 중추적인 역할을 담당하는 온톨로지는 특정 개념에 대해 공유가 가능하도록 정형화된 형식으로 명시한 명세 사항이다[2]. 즉, 어느 특정 도메인에 관련된 단어들을 계층적으로 표현하고, 추가적으로 이를 확장할 수 있는 추론규칙(Inference Rule)을 포함한다[3]. 온톨로지의 중요한 특징은 같은 개념에 대해서 서로 다른 단어를 사용하더라도 이를 해결할 수 있으며, 각각의 개념에 한정 조건(Restriction)을 명시함으로써 개념들 간의 계층 구조를 추론할 수 있다.

본 논문의 근본을 이루는 개념은 온톨로지에 기반을 둔 메타데이터다. 메타데이터는 데이터를 설명하는 데이터로서 어떤 특정 객체나 자원(Resource)에 대한 서술적인 정보를 나타낸다[4]. 시맨틱 웹에서는 온톨로지를 기

반으로 웹 문서를 구성하고 있는 여러 가지 메타 정보를 메타데이터로 저장함으로써 웹 문서의 내용을 컴퓨터가 이해하고 컴퓨터 간의 정보 공유가 가능하도록 하고 있다. 따라서 메타데이터에 대한 개념을 바탕으로 본 논문에서는 시맨틱 환경 내에서 사용자 개개인에 맞는 웹 콘텐츠를 제공하기 위한 온톨로지 기반의 웹 콘텐츠 추천 방법론을 제안한다.

효과적인 웹 콘텐츠 추천을 위해 본 논문은 크게 두 가지에 초점을 맞추고 있다. 먼저, 사용자 프로파일의 쓰임새를 향상시키기 위해 온톨로지 모델을 적용한 사용자 프로파일 구조 정의다. 이는 비슷한 서비스를 제공하는 여러 웹 서비스 사이트에서 사용자의 기호 정보를 공유할 수 있다는 이점을 갖는다. 또한 온톨로지를 기반으로 생성된 사용자 프로파일은 콘텐츠 추천 점수 계산을 위한 정확한 입력 데이터를 제공한다. 두 번째, 사용자 프로파일을 기반으로 각각의 웹 콘텐츠들의 추천 점수를 계산하는 함수를 정의한다. 제안하고자 하는 함수는 각 웹 콘텐츠의 계층구조와 웹 콘텐츠를 구성하는 속성들의 관계를 명시한 온톨로지를 기반으로, 사용자 프로파일의 내용과 웹 콘텐츠의 유사도(Similarity) 구한다. 따라서 계산된 유사도는 웹 콘텐츠 추천 점수로 적용된다.

2. 온톨로지 기반 웹 콘텐츠 추천

본 논문은 메타데이터에 대한 개념을 바탕으로 시맨틱 환경 내에서 사용자 개개인에 맞는 웹 콘텐츠를 제공하기 위한 온톨로지 기반의 웹 콘텐츠 추천 방법론을 제안하고 있다. 따라서 본 논문에서는 두 가지 특징의 존재를 가정하고 연구를 진행하였다.

첫 번째는 같은 계열의 웹 서비스를 아우르는 통합 온

올로지(Universal Ontology)의 존재이다. 온톨로지의 특징 중의 하나는 여러 시스템간의 자료 공유이다. 만약 각각의 웹 서비스 시스템이 서로 다른 온톨로지를 사용할 수 있기 때문에 이들을 전체로 아우르는 통합 온톨로지가 없다면 자료 공유는 그만큼 힘들어진다. 두 번째는 개념들 간의 관계를 추론하기 위한 추론 엔진(Reasoner)이다. 온톨로지는 추론규칙을 통해서 각 개념을 표현하는 단어들의 의미와 관계를 확장하기 때문에, 이러한 확장관계들을 명시하기위한 시맨틱 웹 추론엔진이 필요하다. 따라서 본 논문은 온톨로지 형식으로 작성된 사용자 프로파일을 여러 시스템이 공유하고 이해하며 처리할 수 있도록 각 서비스 시스템의 온톨로지들을 모두 아우르는 통합 온톨로지를 고려하고 적합한 추론엔진을 적용하였다. 본 논문에서는 시맨틱 추론엔진 중 하나인 RacerPro[5]를 채택하였다. 본 시스템에서 RacerPro의 역할은 다른 웹 서비스 시스템이 구축한 사용자 프로파일에 기록된 사용자 기호 정보의 개념과 자사가 사용하는 온톨로지의 개념들 간의 계층 관계를 추론하여, 이미 구축되어진 타 시스템의 사용자 프로파일을 사용할 수 있게 하는 것이다.

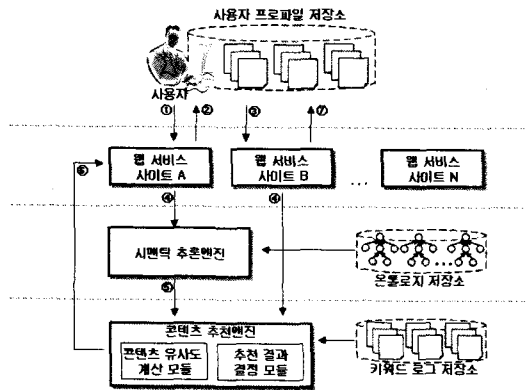


그림 1 온톨로지 기반 웹 콘텐츠 추천 시스템

[그림 1]은 본 논문에서 제안하는 온톨로지 기반 웹 콘텐츠 추천 시스템의 전체적인 프레임워크를 보여준다. 먼저 사용자가 특정 웹 서비스 사이트에 접속하여 서비스를 요청하면 (콘텐츠 검색), 웹 서비스 시스템은 사용자의 질의를 분석하여 서비스 검색결과를 저장한 후, 사용자에게 맞춤형 콘텐츠를 제공하기 위한 첫 단계로서 사용자 프로파일을 요구하게 된다. 이때 사용자의 컴퓨터 에이전트는 3가지 메시지중 하나를 웹 서비스 시스템에 전송한다. 즉, 자신이 접속한 웹 서비스 사이트가 생성한 프로파일이 있는 경우 매치되는 프로파일 전송하거나, 웹 서비스 사이트가 생성한 프로파일이 없는 경우 비슷한 서비스를 제공하는 웹 서비스 사이트가 생성한 프로파일을 전송하거나, 해당 서비스에 대한 프로파일이 존재하지 않는다는 메시지를 전송한다. 웹 서비스 시스템은 자사에서 제공한 프로파일을 전달받았을 경우, 콘텐츠 추천엔진에 프로파일을 바로 전송하여 콘텐츠 추천을 실행하지만,

타사(비슷한 서비스 제공)에서 구축한 프로파일을 제공받았을 경우, 시맨틱 추론 엔진에 프로파일을 전송하여 통합 온톨로지 기반의 시맨틱 추론을 통해서 자사의 콘텐츠 추천 엔진이 처리할 수 있는 형식으로 정제한다. 마지막 단계로서 콘텐츠 추천엔진은 제안하고자 하는 유사도 함수를 사용하여, 사용자 프로파일과 일차적인 콘텐츠 검색 결과(웹 페이지)들간의 유사도를 계산한 후, 추천 점수로 적용한다. 따라서 웹 서비스 시스템은 일정한 임계값 이상의 추천 점수를 갖는 콘텐츠를 사용자에게 추천할 수 있다.

3. 사용자 프로파일의 구조

웹 서비스 시스템에서의 사용자 프로파일은 사용자의 기호(User Preference) 정보를 문자, 숫자, 기호로 표현한 것이다. 따라서 사용자에게 맞춤형 콘텐츠를 제공하는 시스템은 사용자의 기호를 분석하여 정기적으로 사용자 프로파일을 구축 및 수정하고, 이를 통해서 사용자에게 적합한 서비스를 제공한다. 사용자 프로파일은 보통 웹 서비스 시스템에 따라 서로 다른 형식으로 저장되기 때문에 웹 서비스 시스템간의 사용자 프로파일 공유가 거의 불가능하다. 본 논문에서의 사용자 프로파일은 온톨로지 기반 메타데이터 구조로 표현하는 것을 제안한다. 이는 시맨틱 웹 환경에서 비슷한 서비스를 제공하는 웹 서비스 시스템간의 사용자 프로파일 공유를 가능하게 한다. 따라서 본 논문에서는 사용자 프로파일을 웹 온톨로지 언어 중 하나인 OWL로서 표현하였다.

웹 온톨로지 언어인 OWL은 정보의 내용을 직접 처리할 수 있는 소프트웨어 에이전트를 구현하는데 활용될 수 있도록 설계된 언어이다. OWL은 풍부한 어휘(vocabulary)와 형식적 의미론(formal semantics)을 포함하고 있기 때문에 기계 해석이 가능한 웹 콘텐츠를 제작하는데 있어 XML, RDF 및 RDF 스키마(RDF-S)보다 뛰어나다. OWL은 표현력이 서로 다른 세 개의 하위 언어 (OWL Lite, OWL DL, OWL Full) 로 구성되어 있다. [6] 이는 표현력이 높으면 효율적인 추론이 어렵기 때문인데, 실제 추론 수준에 따라 적절한 수준의 OWL을 사용해야 한다.

```
<MovieNews rdf:ID="aboutMovieProfileFirst">
  <aboutMovie rdf:resource="#KingKong"/>
</MovieNews>
<MovieNews rdf:ID="aboutMovieProfileSecond">
  <aboutMovie rdf:resource="#Kings_Man"/>
</MovieNews>
```

그림 2 OWL로 표현된 사용자 프로파일

[그림 2]는 OWL을 이용하여 영화 뉴스에 대한 기호도를 표현한 사용자 프로파일의 간단한 예를 보여준다. 즉 이 사용자 프로파일의 주체는 현재 "KingKong"과 "Kings_Man"이라는 영화에 무척 관심이 높다는 것을 알 수 있다.

보통 사용자 프로파일을 구축하려면 사용자와 웹 서비스 시스템 간에 상당한 인터랙션 정보가 필요하다. 때문에 특정 웹서비스 사이트는 자주 방문하지 않는 사용자에 대해 정확한 맞춤형 서비스를 제공하기 어렵다. 따라서 사용자 프로파일을 온톨로지 형식으로 작성하여 사용

하면, 비슷한 서비스를 제공하는 여러 웹 서비스 시스템에서 공유도 가능하겠지만, 자체적인 프로파일 구축 없이도 사용자에게 보다 질 높은 맞춤형 서비스를 제공할 수 있는 가능성이 높아진다.

3.1 사용자 프로파일과 시멘틱 추론 엔진

시멘틱 추론 엔진은 크게 3 가지 기능을 제공한다. 첫 번째는 온톨로지 일관성 검사(Consistency Check)다. 이는 온톨로지 디자이너가 온톨로지를 정확하게 구축하고 있는가에 대한 검사로서, 대용량 온톨로지를 결점 없이 구축하는 것을 도와준다. 두 번째는 각각의 클래스들에 대한 계층 관계(Class Subsumption Relation)를 추론이다. 클래스는 한정조건(Restriction)을 가지고 있기 때문에 추론 엔진은 이것을 바탕으로 명시된 계층 구조 이외에 새로운 계층 구조 및 관계를 추론하여 에이전트들이 이를 이해하고 처리할 수 있도록 도와준다. 세 번째는 개체(instance)와 클래스의 관계 추론이다. 명시된 클래스와 개체의 관계뿐 아니라 각 클래스의 한정조건을 기반으로 새로운 클래스와 개체의 관계를 추론한다.

시멘틱 웹 기술이 적용된 서비스 시스템은 각각의 서비스 에이전트들이 특정 정보를 이해하고 처리할 수 있도록 새로운 내용을 자동 추론할 수 있는 추론 기술이 필요하다. 본 논문에서의 시멘틱 추론엔진은 각 웹 서비스 에이전트들이 다른 웹 서비스 시스템이 작성한 사용자 프로파일에 대한 정보를 이해하고, 이를 처리할 수 있는 용도로 사용된다. 따라서 각각의 클래스들에 대한 계층 관계와 개체-클래스 관계를 추론한다.



그림 3 추론을 위해 적용된 개념간의 한정 조건

본 장에서는 시멘틱 추론을 통한 사용자 프로파일의 공유를 설명하기 위해서, 먼저 문화정보에 관한 통합 온톨로지를 기반으로 문화정보 관련 웹 서비스를 제공하는 시스템들이 존재한다는 가정 하에, 사용자 프로파일의 공유를 위한 단계를 설명한다.

[그림 3]은 온톨로지를 구성하는 개념들에 적용된 한정 조건을 보여준다. [그림 3]에서 보는 것과 같이 문화정보 통합 온톨로지의 MovieNews 클래스는 Movie Article 클래스의 하위 클래스이며, 항상 informationType이라는 속성 값으로 MovieInformation이라는 개체를 갖는 익명 클래스(Anonymous Class)와 동등(Equivalent)하다는 것을 보여준다. Culture_News 클래스는 informationType이라는 속성 값으로 Information 클래스에 속하는 개체 중 하나를 갖는 익명 클래스의 하부 클래스다. 온톨로지 구조에서 MovieInformation은 Information의 개체이기 때문에 informationType이라는 속성 값으로 MovieInformation이라는 개체를 갖는 익명 클래스는 informationType이라는 속성 값으로 Information 클래스에 속하는 개체 중 아무거나 갖는

Culture_News의 하부 클래스다. 그런데 이 익명 클래스는 MovieNews 클래스와 동등하기 때문에 결론적으로 MovieNews는 Culture_News의 하부 클래스라는 것을 추론할 수 있다. [그림 4]는 [그림 3]에서 보여주는 각 개념에 대한 한정 조건을 이용하여, 시멘틱 추론 엔진이 새로운 계층 관계를 시멘틱 추론한 통합 온톨로지의 계층 구조를 보여준다.

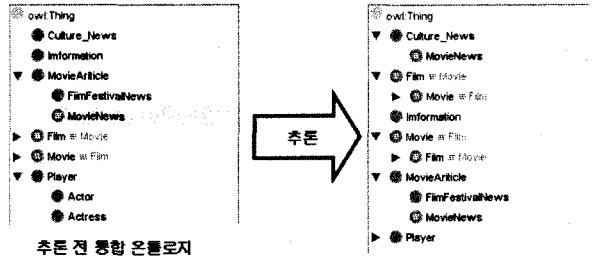


그림 4 계층 관계 추론 과정을 거친 통합 온톨로지

이렇게 개념의 계층 관계 추론은 사용자 프로파일을 공유하기 위한 기반을 제공한다. 예를 들어 영화 뉴스를 제공하는 서비스 시스템 A는 [그림 2]과 같은 형태로 사용자 프로파일을 제공한다. 즉 사용자 프로파일 개체를 제공하는 클래스는 MovieNews 클래스다. 반면 문화정보를 제공하는 서비스 시스템 B는 [그림 5]의 왼쪽편에 보이는 것과 같은 형태로 사용자 프로파일을 제공한다. 즉 사용자 프로파일 개체를 제공하는 클래스는 Culture_News 클래스다. 따라서 계층 관계를 기반으로 객체와 클래스 관계를 추론함으로써 [그림 5]의 오른쪽에 보이는 것과 같이 시스템 B의 사용자 프로파일을 시스템 A가 이해할 수 있는 형태로 변환할 수 있다. 즉, 시스템 B가 제공하는 사용자 프로파일 중 영화 정보 사용자 프로파일은 Culture_news 클래스에 위치하며 informationType 속성 값으로 MovieInformation을 갖는다. MovieNews 클래스는 informationType 속성 값으로 MovieInformation을 갖고, 계층 관계 추론을 통해서 Culture_News의 하부 클래스로 명시되었기 때문에 객체와 클래스의 관계 추론을 통해서 시스템 B가 제공하는 영화 정보 사용자 프로파일은 MovieNews 클래스에 위치하게 된다. 따라서 시스템 A는 사용자가 자사에서 제공한 사용자 프로파일을 소유하지 않고 비슷한 서비스를 제공하는 시스템 B의 사용자 프로파일을 갖고 있다고 해도, 이 두 시스템이 통합 온톨로지를 기반으로 사용자 프로파일을 제공하기 때문에, 사용자가 시스템 B가 제공한 사용자 프로파일을 소유하고 있다면 이를 기반으로 맞춤형 서비스를 제공할 수 있다.

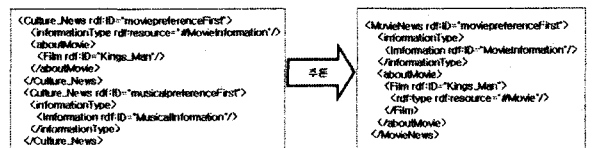


그림 5 객체와 클래스의 관계 추론을 거친 사용자 프로파일

4. 추천 점수 계산을 위한 함수

일반적인 웹 서비스 시스템에서 사용자에게 콘텐츠를 제공하는 방법은 사용자가 입력한 검색 키워드와 웹 페이지 간의 유사도를 측정된 후, 유사도가 높은 웹 콘텐츠 순으로 출력하는 알고리즘을 사용하는 웹 페이지 순위 적용(Ranking)이다. 본 논문에서 제안하는 추천 점수 계산 기법 역시 이와 비슷한 단계로 진행되는데, 먼저 사용자 프로파일의 내용과 시맨틱 웹 기반의 웹 콘텐츠들의 유사도를 측정하기 위해서, 각 문서의 계층구조와 문서의 속성 관계를 명시한 온톨로지를 기반으로 개념 유사도(Class Similarity)와 관계 유사도(Relation Similarity) 구한다.

온톨로지는 상위 개념과 상위 개념의 속성들을 상속 받는 하위개념으로 구성되는데, 제안하고자 하는 개념 유사도와 관계 유사도는 온톨로지의 이러한 특징을 이용한 유사도 계산방법이다. 즉 서로 다른 하위 개념들이 같은 상위 개념에 속한다면 개념적으로 유사한 부분을 공유하게 되는 것이며, 이들의 특정 속성 값의 범위가 같은 개념이라면 역시 어느 정도의 유사성을 공유하게 되는 것이다. 본 논문에서 제안하는 개념 유사도와 관계 유사도의 정의는 다음과 같다.

- **개념 유사도** - 사용자 프로파일과 웹 콘텐츠의 개념들이 위치하는 개념간의 거리를 계산
- **관계 유사도** - 사용자 프로파일과 웹 콘텐츠의 속성 값들이 위치하는 개념간의 전체 유사도의 합

4.1 온톨로지와 메타데이터 모델 정의

온톨로지 기반의 사용자 프로파일과 웹 콘텐츠는 OWL로 표현된 메타데이터 형식을 갖는다. 따라서 본 장에서는 온톨로지 기반의 메타데이터(Metadata)의 유사도를 구하기 위해서 가장 기본이 되는 온톨로지와 메타데이터에 대한 모델을 정의한다. 온톨로지는 일종의 스키마로서 “공유가 가능하도록 정형적이고 명시적으로 표현된 개념”으로 정의할 수 있다. 보다 간단하게 표현하면 특정 도메인을 대표하는 단어들(개념)을 정의하고, 이들이 가지는 속성을 정의하며 이들 간의 관계를 설정한다. 다음 식은 온톨로지의 구성에 대한 모델이다.

$$O := (C, P, A, H^c, prop) \tag{식 1}$$

C와 P는 개념과 개념들 간의 관계(속성)를 나타내며 H^c는 개념간의 계층구조를 표시한다. H^c는 C₁이 C₂의 하위개념이라는 상세 정보를 H^c=(C₁, C₂)로 나타낼 수 있으며, P 역시 C₁과 C₂가 P라는 관계를 맺고 있다는 상세 정보를 prop(P) = (C₁, C₂)로 나타낸다. 즉, prop(P) = (C₁, C₂)는 P의 도메인이 C₁이고 범위(Range)가 C₂이라는 의미를 갖는데, 이는 domain:P→C₁, range:P→C₂로 나타낸다. 단, 범위(range)가 일반적인 데이터 값(string, integer,...)을 가질 경우 range(A):=STRING으로 나타낸다.

$$MD := (O, I, L, inst, instr, instl) \tag{식 2}$$

위의 식은 메타데이터의 구성에 대한 모델이다. O와 I는 특정 온톨로지와 온톨로지의 메타데이터 집합(instance set)을 나타내며, L은 인스턴스가 가질 수 있는 리터럴들의 집합을 의미한다. 특정 개념(Class)에 속하는 인스턴스는 inst(I) = C로 표현되고, 인스턴스 간의 관계는 P(I₁, I₂)로 표현되는데 관계를 형성하는 범위(Range)가 특정 개념(Class)이라면, P → instr | x | 로 나타내며, 범위(Range)가 일반적인 데이터 값일 경우 P → instl | x | 로 나타낸다.

4.2 개념 유사도

개념 유사도는 사용자 프로파일과 웹 콘텐츠의 개념들이 위치하는 개념(Class)간의 거리를 의미한다. 온톨로지의 개념계층(Class Hierarchy)에서는 각 개념들은 하위 개념을 가지며, 이들 하위 개념들은 또 다른 하위 개념을 가질 수 있다. 이러한 개념계층을 바탕으로 두 문서간의 개념관계를 설명하기 위한 거리(Distance)를 측정함으로써 두 문서가 얼마나 유사한 개념을 공유하는지 알 수 있다. 이 거리를 “개념 유사도”라고 부른다. 다음 식은 두 문서간의 개념 유사도를 구하는 방법을 정의하고 있다. (C, H^c)는 개념 C에 속한 문서의 개념계층 H^c를 의미한다.

$$CM(C_1, C_2) := \frac{|(C_1, H^c) \cap (C_2, H^c)|}{|(C_1, H^c) \cup (C_2, H^c)|} \tag{식 3}$$

4.3 관계 유사도

관계 유사도는 사용자 프로파일과 웹 콘텐츠가 갖는 속성의 속성 값들이 위치하는 개념간의 전체 유사도를 의미한다. 온톨로지의 개념들은 각 개념에 속하는 개체(instance)의 특징을 설명해주는 속성(Property)을 가지며, 이 중 객체 속성(Object Property)은 속성 값으로 다른 개념에 속하는 개체를 택할 수 있다. 이러한 객체 속성을 바탕으로 사용자 프로파일과 웹 콘텐츠가 갖는 속성들의 유사성을 설명하기 위해 속성 값들의 유사도를 측정함으로써 사용자 프로파일과 웹 콘텐츠가 얼마나 유사한지 알 수 있다. 따라서 관계 유사도는 사용자 프로파일과 웹 콘텐츠의 속성 값들의 개념간의 거리가 근간을 이루는데, 이는 사용자 프로파일과 웹 콘텐츠의 속성 값들의 전체 유사도로서 정의된다. 전체 유사도는 4.4에서 설명하도록 한다. 다음 식(4)는 사용자 프로파일과 웹 콘텐츠가 공통된 객체 프로퍼티를 가지지 않는 경우를 정의한 것이다. 이는, 서로 다른 개념 내의 개체들이 더 이상 다른 개체들과 특정한 관계를 갖지 않음을 의미하며 이때의 속성 값들의 전체 유사도(Min_Similarity)는 0이 된다. 반면에 식(5)은 사용자 프로파일과 웹 콘텐츠가 공통된 객체 프로퍼티로 공통된 개체를 가지는 경우를 정의한 것이다. 이는, 서로 다른 개념 내의 개체들이 같은 개체와 관계를 갖는 것을 의미하며, 이때의 속성 값들의 전체 유사도(Mex_Similarity)는 1이 된다.

$$\text{if } A_s(P, I_1) = 0 \vee A_s = (P, I_2) \\ \text{then } \text{MinSim}_{(P)} := 0 \quad (\text{식 4})$$

$$\text{if } A_s(P, I_1) = A_s(P, I_2) \\ \text{then } \text{MaxSim}_{(P)} := 1 \quad (\text{식 5})$$

다음 식은 사용자 프로파일과 웹 콘텐츠간의 관계 유사도를 구하는 방법을 정의하고 있다. $\text{Sim}(A_s(P, I_1), A_s(P, I_2))$ 는 객체 I_1 (사용자 프로파일)에 대한 프로퍼티 P의 속성 값과 객체 I_2 (웹 콘텐츠)에 대한 프로퍼티 P의 속성 값의 전체 유사도를 의미한다.

$$RS(I_1, I_2) := \frac{\sum_{R \in \text{Rown-1}} \text{Sim}(A_s(R, I_1), A_s(R, I_2))}{Rn} \quad (\text{식 6})$$

4.4 추천 점수

추천 점수는 사용자 프로파일과 웹 콘텐츠 대한 전체 유사도가 적용되며, 전체 유사도에는 개념 유사도와 관계 유사도가 적용된다. 따라서 일정한 임계값 이상의 유사도를 갖는 웹 콘텐츠는 사용자에게 유사도가 큰 순으로 추천된다. 본 논문에서는 사용자 프로파일과 웹 콘텐츠간의 유사도를 구할 때 가중치를 고려하지 않고 개념 유사도와 관계 유사도를 절반씩 적용하였다. 다음 식은 두 문서간의 유사도를 구하는 방법을 정의하고 있다. $\text{Sim}(I_1, I_2)$ 는 사용자 프로파일 I_1 , 웹 콘텐츠 I_2 의 추천 점수를 의미한다.

$$\text{sim}(I_1, I_2) = \frac{CS(I_1, I_2) + RS(I_1, I_2)}{2} \quad (\text{식 7})$$

5. 관련 연구

일반적인 웹 환경에서 웹 콘텐츠 추천을 위해 가장 많이 사용되고 있는 방법은 협력적 여과 시스템(Collaborate Filtering System)이다. 협력적 여과 시스템은 콘텐츠를 이용하는 사용자와 가장 유사한 기호를 나타내는 선호도에 따라서 사용자에게 필요한 콘텐츠를 제공하는 시스템이다[7]. 협력적 여과 시스템은 콘텐츠의 내용을 직접 분석할 필요 없이 비슷한 기호를 갖는 사용자들의 관계를 이용하며, 추천의 범위를 넓힘으로써 여러 종류의 콘텐츠를 추천할 수 있다[8]. 또한, 콘텐츠의 내용뿐만 아니라 콘텐츠의 질적인 우수성에 따라 콘텐츠를 추천할 수 있다는 장점을 갖는다[9]. Ringo나 GroupLens와 같은 협력적 여과 시스템은 피어슨 상관을 이용하여 사용자들의 콘텐츠에 대한 선호도를 비교하여 상관관계를 계산하고, 이를 기반으로 유사한 사용자를 찾는다[10].

협력적 여과 시스템을 사용하는 대표적인 추천 시스템인 Amazon은 사용자가 특정 도서에 관심을 가지고 있을 때, 관심을 가지는 도서를 구입한 다른 사람이 동시

에 구입한 도서를 사용자에게 추천하고 있다[11,12]. 이는 사용자의 현재 관심을 기반으로 하는 단기간의 추천이며, 일반적인 경향 분석에 대한 추천으로 사용자의 세부적인 관심을 고려할 수 없으며, 모든 사람에게 동일한 추천을 수행할 수밖에 없다. 따라서 본 논문에서 제안한 방법은 주기적인 학습을 통해 생성된 온톨로지 기반 사용자 프로파일의 공유를 통해서 개개인에 맞는 콘텐츠를 추천하기 때문에 위 사항과 같은 단점을 극복할 수 있다. 현재 Amazon 이외에도 BarnesAndNoble, ZDNet, MyLaanch, PersonaLogic과 같이 인터넷 쇼핑물을 검색하여, 사용자에게 맞는 결과를 추천해주는 시스템이 있다.

6. 결론 및 향후 연구

본 논문에서 메타데이터에 대한 개념을 바탕으로 시맨틱 환경 내에서 사용자 개개인에 맞는 웹 콘텐츠를 제공하기 위한 온톨로지 기반의 웹 콘텐츠 추천 방법론을 제안하였다. 시맨틱 웹은 온톨로지를 기반으로 웹 문서를 구성하고 있는 여러 가지 메타 정보를 메타데이터로 저장함으로써 웹 문서의 내용을 컴퓨터가 이해하고 컴퓨터간의 정보 공유가 가능하도록 하고 있다. 또한 온톨로지를 기반의 메타데이터는 콘텐츠 추천 점수 계산을 위한 정확한 입력 데이터를 제공한다. 따라서 메타데이터에 대한 개념을 바탕으로 크게 2 가지에 초점을 맞추어 연구를 진행하였다. 먼저, 사용자 프로파일의 쓰임새를 향상시키기 위해 온톨로지 모델을 적용한 사용자 프로파일 구조 정의한 후, 두 번째로 사용자 프로파일을 기반으로 각각의 웹 콘텐츠들의 추천 점수를 계산하는 함수를 정의하였다. 본 논문에서 제안한 방법론은 시맨틱 웹 환경에서 사용자 프로파일을 공유함으로써 사용자에게 보다 질 높은 맞춤형 서비스를 제공하기를 원하는 시스템에 적용이 가능하다.

향후 진행 될 연구는 크게 2 가지다. 본 논문에서 정의한 구조로 사용자 프로파일을 구축하기 위해 온톨로지 기반 사용자 웹 콘텐츠 선호도 학습에 관한 연구와 추천 점수를 계산하는 함수에 개념 유사도와 관계 유사도간의 가중치 조절을 적용하기 위한 연구가 진행될 것이다. 이와 별도로 본 연구가 실제 웹 서비스 시스템에 적용되기 위해서 악의적인 의도를 갖는 타 시스템이 사용자에게 사용자 프로파일을 요청할 경우, 이를 판단하고 처리하는 보안 문제에 관한 연구가 필요하다.

참고 문헌

- [1] Berners-Lee, T., Hendler, J. and Lassila, O., "The Semantic Web," Scientific American, 2001
- [2] Gruber, T., "A translation approach to portable ontologies," Knowledge Acquisition, Vol. 5, No. 2, pp.199~220, 1993
- [3] 최중민, "시맨틱 웹의 개요와 연구 동향," 정보과학회지, Vol. 21, No. 03, pp. 4~10, 2003.03
- [4] Lassila, O., "Web metadata: a matter of semantics," IEEE

- Internet Computing, Vol. 2, No. 4, pp.30~37, 1998
- [5] <http://www.racer-systems.com>. RacerPro User's Guide Version 1.9. Racer Systems GmbH & Co. KG, 2005
- [6] <http://www.w3.org/TR/owl-features/>, "OWL Web Ontology Language Overview", W3C Recommendation, 2000
- [7] W. S. Lee, "Collaborative learning for recommender systems." In Proceedings of the Conference on Machine Learning, 1997
- [8] J. Delgado and N. Ishii, "Formal Models for Learning of User Preferences, a Preliminary Reprint," In proceedings of International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 1999
- [9] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedel, "Analysis of Recommendation Algorithms for E-Commerce," Proc. of the ACM E-Commerce 2000, 2000
- [10] A. Kohrs and B. Merialdo, "Using Category Based Collaborative Filtering in The Active Webmuseum," Proceedings of the IEEE International Conference on Multimedia and Expo-Vol. 1, 2000
- [11] Lawrence Holder and Meral Ozsoyoglu, "Data mining, Data Warehousing & OLAP and ECommerce Breakout Group Report," NSF IDM Workshop, 2001
- [12] Yezdi Lashkari, Max Metral and Pattie Maes, "Collaborative Interface Agents," Conference of the American Association for Artificial Intelligence, Seattle, August, 1994