

온톨로지 기반의 지능형 시맨틱 블로그 모델링

양경아⁰, 김기현¹, 양재동¹, 최 완²
전북대학교 전자정보 공학부

{kayang⁰, khkim¹, jdyang¹}@chonbuk.ac.kr, wchoi²@etri.re.kr

Intelligent Semantic Blog Modeling Based on Ontology

Kyung-Ah Yang⁰, Ki-Heon Kim¹, Jae-Dong Yang¹ and Wan Choi²

^{0,1}Division of Electronics and Information Engineering, Chonbuk National University

²Electronics and Telecommunications Research Institute

요 약

본 논문에서는 시맨틱 웹 환경의 핵심기술인 온톨로지 기술을 블로그에 접목하기 위한 블로그 모델링 방식을 제안한다. 블로그에 사용되는 개념과 개념 간의 관계를 온톨로지로 표현하여 공유된 지식으로 이용함으로써 블로그 상에서 보다 세부적이고 정확한 검색 및 추론을 가능하게 하였다. 블로그에 시맨틱 웹 기술인 온톨로지를 적용시킴으로써 사용자의 특성에 맞는 개별화된 블로그 서비스를 제공할 수 있으며 사용자에게 관심 블로그 커뮤니티에 대한 링크 추천이나 사용자 관점에 따른 커뮤니티를 동적으로 구성 할 수 있다. 또한 사용자가 중요하게 생각하는 관점에 따라 원하는 블로그 콘텐츠 검색이 가능하다.

1. 서 론

“ Web+ Blog ” 의 합성어로서 웹에 기록하는 일지라는 의미를 지닌 블로그는 개인이 스스로 작성한 칼럼, 기사, 일기 등을 웹에 올려 여론을 형성하는 사이트이다. 1998년 소수에 지나지 않았던 블로그는 2004년에는 4백만 개 이상으로 증가하면서 콘텐츠 중심으로 강력한 영향력을 발휘하는 커뮤니케이션으로 자리잡고 있다. 단기간동안 급속도로 성장한 블로그는 그 배경이나 관점에 따른 구조화가 되어 있지 않은 채 인터넷 상에 흩어져 있다. 이처럼 블로그 콘텐츠의 양이 방대해지면서 효율적 검색 능력이 저하되고 이는 블로그 성장에 장애적 요소가 되고 있다. 이 문제를 해결하기 위해 HP는 블로그에 시맨틱 구조를 적용하여 다른 블로그 채널에서도 콘텐츠를 활용할 수 있도록 한 “ SWAD-Semantic Blogging ” [1] 프로젝트를 수행하였다. 웹사이트 간 정보를 공유할 수 있는 표준인 RDF를 이용하여 블로그 엔트리(entry)를 기술하고 이를 검색에 활용하였으며 Semantic View, Semantic Navigation, Semantic Query 세 분야로 블로그 콘텐츠를 접근할 수 있도록 하였다. K-Collector[2]는 소집단의 비즈니스 그룹 구성원들이 만든 블로그만을 범위로 국한하여 what, where, who, when 과 같은 4가지 관점으로 블로그 콘텐츠를 분류하고 검색할 수 있게 하였다. SWAD-Semantic Blogging 프로젝트의 경우 블로그 콘텐츠를 기술하는데 15가지 Dublin Core 메타데이터를 활용하였으며 K-Collector는 What, Where, Who, When 4가지 주요 관점만을 제공하고 있다. 이러한 연구들은 형식화된 구조를 블로그에 접목시키기 위한 시도를 하였으나 범위가 매우 제한적인 메타데이터만을 사용한 탓에 블로그 사용자의 다양한 요구를 충족시키기에는 한계가

있다. 또한 블로그 콘텐츠의 표현 언어로 RDF를 사용하고 있어 RDF의 제한된 기능만을 활용하므로 블로그의 구조적 표현이나 세부적인 관계 설정, 다양한 추론 기능을 제공하기에는 아직 미흡한 수준이다. 본 연구에서는 시맨틱 웹 기술을 적용하기 위하여 블로그를 모델링하고 온톨로지를 적용함으로써 블로그의 구조적 지식을 공유한다. 이를 기반으로 다양한 관점과 지식을 사용자에게 제공하며 의미를 기술할 수 있는 온톨로지 언어인 OWL[3]을 사용하여 보다 정확하고 세부적인 검색과 추론을 가능하게 한다. 이같은 방식은 사용자 개인의 특성에 맞는 차별화된 서비스를 지원하며 온톨로지를 기반으로 사용자의 관점에 따라 블로그를 검색하거나 관심 커뮤니티도 동적으로 구성할 수 있도록 한다.

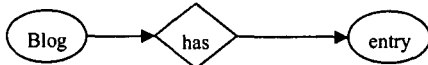
본 논문의 구성은 다음과 같다. 2장에서는 블로그 모델링과 구축 예를 설명하고, 3장과 4장에서는 시맨틱 블로그를 이용하여 활용할 수 있는 다양한 기능들에 대해 살펴본다. 마지막으로 결론 및 향후 연구를 제시한다.

2. 블로그 온톨로지 모델링 및 구축

블로그에 사용되는 개념들을 개체로 간주하고 개념 간의 관계 및 속성을 모델링 한 후 이들 간의 제약사항이나 관계 링크를 설정하여 블로그를 모델링하는 방법을 제안한다. 블로그의 개념과 관계 및 속성을 구조화하기 위해 [4]에서 제시한 온톨로지 그래픽 표기법을 사용한다. 우선 블로그에 사용되는 주요 용어들을 클래스(class)로 표현한다. 블로그가 이에 해당되며 ○로 표기한다. 클래스의 특정 객체가 되는 개념인 인스턴스(instance)는 실제 개설된 개개의 블로그를 지칭하며 각 블로그는 유일한 URI로 구분된다. 인스턴스는 그래픽 표기법으로 □로

표현한다. 블로그에서 사용되는 관계성(relationship)은 클래스 간의 관계를 단순한 연관성 정도를 표시하는 수준을 넘어 이웃, trackback, comment, reply 등과 같이 세부적 의미를 부여하기 위해 사용되며 ◇로 표기한다. 앞에서 정의한 모델링을 기반으로 다음과 같은 의미를 그래픽 표기법을 이용하여 표현할 수 있다.

- 블로그는 엔트리를 갖는다.



- 엔트리는 다른 엔트리에 trackback 할 수 있다.



- http://blog.chonbuk.ac.kr/kayang/490 은 엔트리 인스턴스이다.

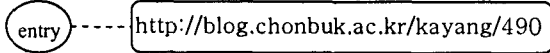


그림 1은 블로그의 개념과 개념 간 사용되는 세부적인 속성을 온톨로지로 모델링한 예이다. “GlobalCategory”는 블로그 도메인의 지식을 정의하는 온톨로지로서 블로그가 분류될 수 있는 카테고리를 클래스 계층으로 표현한 것이다.

블로그 클래스는 엔트리 클래스와 has 관계에 있으며 이 관계는 각각의 인스턴스에 상속된다. 예를 들어, 블로그 인스턴스인 “http://blog.chonbuk.ac.kr/kayang” -①는 엔트리 인스턴스인 “http://blog.chonbuk.ac.kr/kayang/490” -②를 가진다(has). 엔트리는 그 내용에 따라 “GlobalCategory”의 하위 클래스와 링크를 맺는다. 예를 들어 최근 데이터베이스 동향에 관한 내용을 담은 엔트리 ②를 “Database”라는 블로그 온톨로지 용어의 인스턴스로 정의함으로써 그 연관관계를 표현한다.

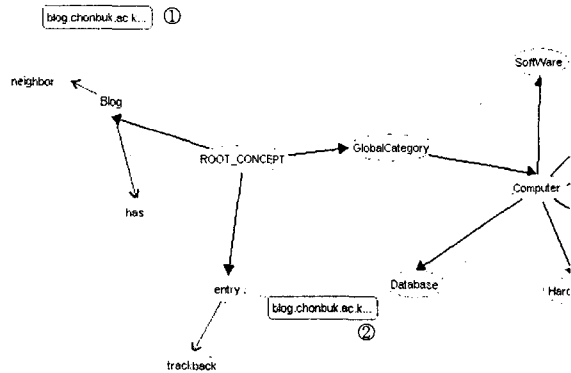


그림 1 블로그 모델링의 예

이처럼 블로그 개념과 이들의 세부적인 관계를 기계가 판독 가능한 온톨로지로 명백하게 명시함으로써 사용자에게 공유되어 지식으로 활용될 수 있게 된다. 다음은 블로그 온톨로지의 OWL 결과 중 일부를 나타낸 것이다.

```

...
<owl:Class
rdf:about="http://ktech.chonbuk.ac.kr/ontology#Blog"/>
<j.0:Blog
rdf:about="http://ktech.chonbuk.ac.kr/ontology#blog.chonbuk.ac.kr/kayang">
<j.0:has>http://ktech.chonbuk.ac.kr/ontology#blog.chonbuk.ac.kr/kayang/490</j.0:has>
</j.0:Blog>
<j.0:entry
rdf:about="http://ktech.chonbuk.ac.kr/ontology#blog.chonbuk.ac.kr/kayang/490">
rdf:type="http://ktech.chonbuk.ac.kr/ontology#Database"/>
...
    
```

그림 2 OWL로 표현한 블로그 온톨로지의 예

3. 온톨로지를 이용한 지능형 블로그 브라우징 및 검색

온톨로지를 활용하면 블로그 상에서 다음과 같은 지능적인 브라우징과 검색이 가능하다.

● 블로그 온톨로지 클래스 간의 관계를 이용한 검색
블로그 카테고리 클래스가 다른 카테고리 클래스와 맺고 있는 관계를 이용하여 검색을 할 수 있다. 그림 3은 블로그 카테고리 클래스 사이의 관계를 이용한 지능형 검색의 예이다. “디지털 카메라” 카테고리는 “영상처리”와 deviceOf 관계, “인터넷 쇼핑몰”과 sells 관계에 놓여 있으므로 사용자가 특정 카테고리에서 세부적 의미를 표현하는 관계들을 따라 개별적으로 설정된 관계 링크를 추적하여 원하는 정보를 검색할 수 있다.

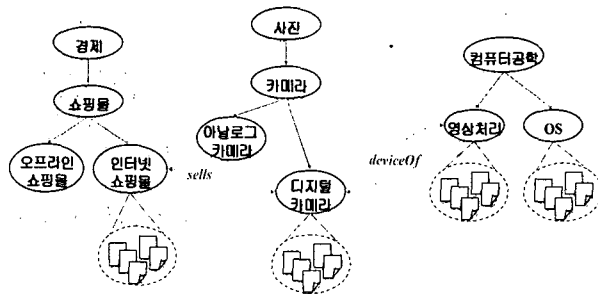


그림 3 블로그 클래스 간 관계를 이용한 지능형 검색

예를 들어 “영상처리” 블로그 카테고리에 접근한 사용자는 영상처리에 사용되는 장치들에 관련된 정보를 얻고자 할 때 “영상처리”가 가지고 있는 관계성 중 deviceOf 관계를 추적하여 “디지털카메라” 블로그 카테고리에 이를 수 있다. “디지털 카메라” 카테고리의 블로그를 둘러보던 사용자는 물품을 구매하고자 할 때 “디지털 카메라” 클래스와 sells 관계를 맺고 있는 “인터넷 쇼핑몰” 카테고리에 둘러 관심 물품을 구매할 수 있다.

● 온톨로지를 이용하여 일종의 패시화된 카테고리 작성 온톨로지를 통해 패시화된 기능을 하도록 함으로써 관점

에 따라 동적으로 블로그를 그룹핑하여 사용자가 원하는 관점을 반영할 수 있다. 일반적으로 하나의 블로그는 여러 개의 카테고리에 해당될 수 있으며 이는 각 블로그가 검색 관점에 따라 동적으로 분류되는데 사용될 수 있다. 그림 4에서 서양 미술 블로그들이 존재할 때 각 블로그는 시대별로 혹은 화풍별로 각 해당 카테고리 클래스에 링크될 수 있다. 블로그가 “화풍”에 대한 관점으로 블로그를 분류하고자 가정하자. “화풍” 카테고리의 하위 카테고리인 “추상파”, “인상파”, “입체파”를 기준으로 해당되는 블로그를 그룹핑할 수 있다. 이처럼 온톨로지의 도메인 지식을 이용하여 전문가가 사용하는 용어를 검색에 활용함으로써 사용자가 원하는 관점에서 적합한 콘텐츠를 접하게 된다.

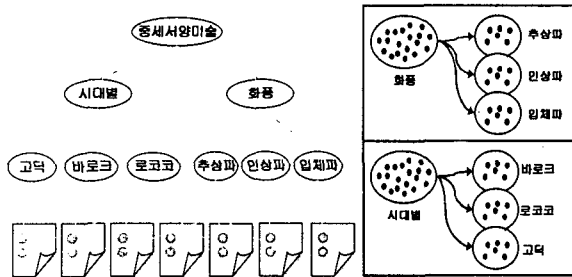


그림 4 온톨로지를 이용한 동적 블로그 그룹핑

4. 온톨로지를 이용한 지능형 블로그 추천 기능

온톨로지를 활용하여 다음과 같은 블로그 추천 기능을 수행할 수 있다.

- 엔트리 작성 시 카테고리 추천기능
 사용자가 엔트리가 해당하는 카테고리를 명시하지 않아도 엔트리에 사용된 단어의 시맨틱을 통해 자동으로 카테고리를 추천할 수 있다. 예를 들어, 카메라에 관련된 내용을 담고 있는 XML문서가 존재하고 OWL로 작성된 카메라 온톨로지가 시맨틱 웹 상에 존재할 때 XML 문서에 사용된 개념들을 구축된 온톨로지와 비교함으로써 이 문서가 카메라 카테고리에 해당하는 것을 알 수 있다. 이러한 추천기능을 제공하는 OWL 언어를 블로그에 적용한다면 블로그 시스템이 해당 엔트리가 포함될 수 있는 카테고리 리스트를 사용자에게 추천할 수 있으며 사용자는 엔트리에 대한 카테고리를 직접 작성할 필요 없이 해당리스트 중 선택하여 엔트리의 카테고리를 지정할 수 있다.
- 시맨틱 데이터 마이닝을 이용한 상위 관계성 생성
 서로 다른 두 개의 블로그 커뮤니티 간에 갖은 접근이 발생하고 카테고리에 속한 블로그 간에 이러한 현상이 일반화 될 때 이 두 블로그 카테고리 간 세부 관계성(specific relationship) 설정이 가능하다. 하위 카테고리 간의 관계가 일반화 되었다고 가정하면 상위 카테고리간의 보다 더 일반적인 관계성(general relationship)을 설정할 수 있다. 예를 들어 “요가”에 관심있는 사람은 “자연식품”에 관심을 가지며 “트레이닝”을 수행하는 사람은

“건강보조식품”에 관심을 갖는다고 가정하면 “요가” 카테고리와 “자연식품” 카테고리 간에 *assist* 라는 관계를, “트레이닝” 카테고리와 “건강보조식품” 카테고리는 *fillup* 라는 관계를 각각 부여할 수 있다. “운동”과 “식품” 카테고리의 하위 카테고리 간에 이러한 관계설정이 활발히 일어날 경우 이들 상위 카테고리간에 보다 온톨로지 정보를 기반으로 한 일반적인 의미의 관계설정이 가능하다.

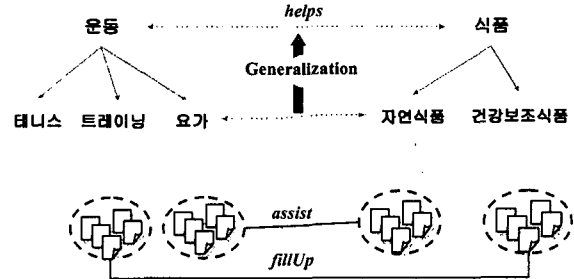


그림 5 하위 관계성을 이용한 상위 관계성 생성

5. 결론

본 논문에서는 시맨틱 웹 환경의 핵심기술인 온톨로지 기술을 블로그에 접목하기 위하여 블로그를 모델링하고 이를 이용하여 블로그 온톨로지를 작성하여 블로그 상에서 보다 세부적이고 정확한 검색 및 추천이 가능하게 하였다. 이러한 연구를 통해 온톨로지를 이용하여 사용자의 특성에 맞는 개별화된 블로그 서비스를 제공할 수 있다. 또한 사용자 관점에 따른 커뮤니티 동적 구성을 할 수 있으며 사용자가 중요하게 생각하는 관점에 따라 원하는 블로그 콘텐츠 검색이 가능하다. 추후 블로그 온톨로지와 블로그에서 사용가능한 S/W를 연계하여 효과적으로 작업할 수 있는 S/W를 사용자 특성에 맞게 제공할 수 있는 시스템 개발 연구를 수행하고자 한다.

참고문헌

1. Steve Cayzer, “ Semantic Blogging” HP Lab http://www.hpl.hp.com/personal/Steve_Cayzer/semblog.htm
2. K-Collector <http://www.evectors.com/itkcollector/>
3. OWL Web Ontology Language Overview <http://www.w3.org/TR/owl-features/>
4. 최재훈, 김기현, 양재동, 박천수, "OWL 기반 그래픽 바이오 온톨로지 관리 시스템의 설계 및 구현", 정보과학회 논문지(소프트웨어 및 응용), Vol. 32, No. 6, pp.232-234, 2005
5. Web Ontology Language Abstract Syntax and Semantics <http://www.w3.org/TR/2002/WD-owl-semantics-20021108/>
6. Edith Cohen and Balachander Krishnamurthy, “ A short walk in the Blogistan” , Computer Networks, In Press, Corrected Proof, Available online 21 July 2005