

# 의미 기반 정보 검색을 제공하는 이미지 게시판 시스템

<sup>1</sup>정의현, <sup>2</sup>조동찬<sup>0</sup>

<sup>1</sup>안양대학교 이공대학 디지털미디어공학과

<sup>1</sup>jung@anyang.ac.kr

<sup>2</sup>한양대학교 공과대학 전자전기컴퓨터공학부

<sup>2</sup>wantseeu@nownuri.net

## An Image Bulletin Board System providing Semantic-based Searching

<sup>1</sup>Eui-Hyun Jung and <sup>2</sup>Dong-Chan Cho<sup>0</sup>

<sup>1</sup>Dept. of Digital Media, Anyang University

<sup>2</sup>Division of Electrical and Computer Engineering, Hanyang University

### 요 약

게시판 시스템은 양방향으로 정보를 교환하는 정보 시스템으로서의 높은 효용을 지니고 있으며, 웹과 결합하여 다양한 정보 시스템의 핵심 요소로 자리잡고 있다. 또한 이미지 등의 멀티미디어 정보를 게시물에 포함하여 효율적인 정보 공유에 사용되고 있다. 그러나 지금까지의 게시판 시스템은 게시물의 내용에 접근하기 위해, 단순한 텍스트 패턴 매칭에 의존하고 있다. 이러한 접근 방식은 텍스트 중심의 게시판에서는 어느 정도 효용을 갖지만, 멀티미디어를 포함하는 게시판의 경우에는 적용되기 어려운 단점을 갖고 있다. 본 논문에서는 이의 해결을 위해 이미지 데이터를 포함하는 게시물에 대해 시맨틱 태깅을 할 수 있는 게시판 시스템에 관하여 논한다. 제안된 시스템은 사전에 정해진 태깅 정보가 코드에 고착되지 않고, 외부에서 지정한 시맨틱 태깅을 동적으로 수용하는 구조를 갖고 있다. 이러한 구조를 통하여 이미지의 종류나 성격에 가장 적합한 태깅을 동적으로 지정할 수 있게 되며, 의미 기반의 검색을 지원하게 된다.

### 1. 서 론

게시판 시스템은 여러 사용자들이 동시에 정보 공간(Information Space)을 만드는 구조로 인하여, 사용자들의 정보 공유와 의견 교환에 매우 적합한 정보 시스템으로 간주되어 왔다[1]. 이러한 이유 때문에 게시판 시스템은 교육, 행정 시스템 등의 다른 정보 시스템의 중요한 요소로 연구되었고, 멀티미디어 데이터를 접근하는 다양한 형태의 확장에 대해서 연구가 진행되었다[2]. 일반적으로 게시물에 포함된 멀티미디어 데이터의 포함은 등록된 게시물을 선택한 후, 게시물을 통해서 브라우저할 수 있는 구조가 된다. 이런 구조로 인하여 멀티미디어 데이터의 검색은 게시물을 통한 1차적인 검색을 통해서 접근이 가능하게 된다. 그러나, 게시물의 내용을 기준으로 한 검색은 단순한 텍스트 패턴 매칭을 이용하게 됨으로 포함하고 있는 멀티미디어 데이터에 대한 정확한 검색은 어려운 문제로 남게 된다.

이것은 게시판 시스템에 등록(posting)되는 정보의 속성이 렉시컬(lexical) 형태로 저장된다는 특징 때문이다. 즉, 등록되는 게시물은 사용자별로 같은 의미를 가진 정보를 다른 형태로 저장하게 되는 경우가 일반적이다. 예를 들어, 고양이 사진을 올릴 때, 어떤 사람은 “고양이”

라고 명기하지만, 어떤 사람은 “나비”라고 표기할 수 있다. 이런 구조에서는 “고양이”라고 검색을 한다 하더라도 “나비”라고 표기된 고양이 사진을 검색할 수 없게 된다.

이렇게 같은 의미를 다른 렉시컬을 통해 표현할 수 있는 정보 표현의 상이함은 게시판을 통해서 정보에 접근하려는 사용자가 올바른 정보에 접근하는데 어려움을 야기시킨다. 이러한 렉시컬 기반의 정보 처리에 대한 문제는 단순히 게시판 시스템만의 문제는 아니며, 대부분의 정보 시스템에서도 같은 문제점을 안고 있다. 이를 해결하기 위해, Tim-Berners Lee 등은 시맨틱 웹(Semantic Web)을 제안하였다[3]. 시맨틱 웹은 기존 정보에 시맨틱 정보를 추가하여, 기계가 의미를 이해하고 사용자에게 의도한 정보를 검색할 수 있도록 해준다.

본 논문에서는 시맨틱 웹 개념을 이미지 게시판 시스템에 적용하여 게시물을 등록할 때, 형식화된 의미기반 태깅을 제공하고, 의미기반 검색을 가능하게 하는 구조에 관하여 논한다. 제안된 게시판 시스템은 외부에서 정의한 프로퍼티 기술(Property Description)을 읽어들이며, 동적으로 각 게시판의 영역(domain)에 맞는 등록/검색 HTML 태그와 적합한 RDF 문서를 생성한다. 또한, RDQL을 통한 의미기반 검색을 지원한다.

본 논문은 2장에서 시맨틱 태깅을 게시판에 적용하기 위한 방안과 동적으로 온토로지를 수용(importing)하기 위한 방법에 관하여 설명한다. 3장에서는 설계된 이미지 게시판의 구조와 의미기반 검색에 관해서 설명하고, 4장에서 결론을 맺는다.

2. 온토로지 적용 방안

2.1. 형식화된 시맨틱 태깅

게시물에 대한 시맨틱 태깅을 수작업으로 하는 것은 사용자에게 시맨틱 웹 태깅 언어에 대한 상당한 수준의 사전 지식을 요구하며, 입력 오류를 발생시킬 가능성이 매우 높다. 이를 피하기 위해서는 형식화된 시맨틱 태깅을 제공해야 한다. 웹 브라우저에서 형식화된 태깅을 제공하기 위해서는 각 프로퍼티에 대응되는 HTML 폼 태그(Form Tag)가 생성되어야 한다. 그림 1에서 볼 수 있는 것처럼, 시맨틱 웹은 특정 리소스에 대해서 프로퍼티와 해당 값을 얻어야 하므로, 본 논문에서는 프로퍼티와 태그 컴포넌트의 ID를 매핑하고, 프로퍼티의 값과 컴포넌트에 저장된 값을 매핑하는 구조를 갖도록 하였다. 본 논문에서 사용되는 시맨틱 태깅 언어는 RDF(Resource Description Framework)이다.

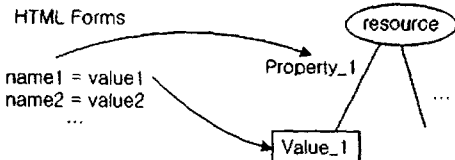


그림 1. RDF 프로퍼티에 대한 지정

2.2. 동적인 온토로지의 적용

앞 절에서 명시한 방식으로 형식화된 시맨틱 태깅을 구성하는 것은 기존의 CGI(Common Gateway Interface) 기법을 이용해서 어렵지 않게 구성할 수 있다. 그러나, 이런 단순한 적용은 해당 게시판의 코드에 형식화된 태깅이 종속되는 결과를 낳는다. CGI 코드는 사전에 처리해야 할 HTML에 대해서 종속적으로 작성되고, 처리된 데이터의 저장도 고정되기 때문이다. 그러나, 일반적으로 게시판의 종류에 따라서 요구되는 시맨틱이 달라지게 되므로, 이런 접근 방안은 확장성에 큰 문제점을 야기하게 된다.

본 논문에서는 이를 해결하기 위해서 사용하려는 온토로지의 프로퍼티에 대해서 기술(description)할 수 있는 새로운 온토로지를 설계하였다.

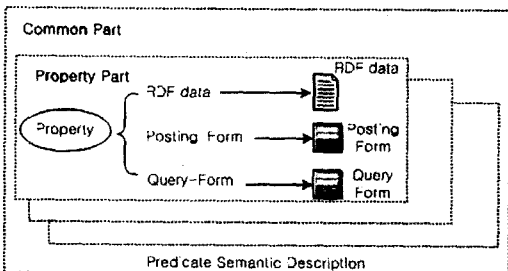


그림 2. 설계된 온토로지의 역할

설계된 온토로지는 그림 2에서 볼 수 있는 것처럼, 특정 프로퍼티에 대해서 등록/검색 HTML 생성과 RDF 문서 생성에 대한 프로퍼티 기술을 할 수 있도록 되어 있다. 이런 구조를 통해서 다른 온토로지의 프로퍼티를 사용할 수 있고, 자신만의 온토로지를 구성하는 것이 가능해진다.

2.3. 예: 이미지 온토로지

제안된 온토로지를 이용해서 여러 종류의 프로퍼티 기술을 할 수 있지만, 본 논문에서는 다음과 같은 애완동물과 관련된 게시판의 프로퍼티 기술을 갖고 예를 들기로 한다. 애완 동물에 대한 게시물은 “동물 종류”, “구입 가격”, “원산지” 등의 프로퍼티를 가질 수 있다. 이 중에서 “동물 종류” 프로퍼티를 나열형(enumeration)의 데이터로 지정하면, 다음과 같은 프로퍼티 기술을 갖게 된다. 동물 종류는 나열형(enumeration)으로 RDF:Bag으로 지정하였다.

```
<rdf:Description
rdf:about="http://www.pet.com/2004/app/imageboard#pettype">
<rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
<sboard:prefix>Pet</sboard:prefix>
<sboard:datatype
rdf:datatype="http://www.sboard.org/2003/sboard-rdf#enumeration"/>
<sboard:label>동물 종류</sboard:label>
<sboard:formtag
rdf:resource="http://www.sboard.org/2003/sboard-rdf/1.0#selection"/>
<sboard:defaultvalue>고양이</sboard:defaultvalue>
<sboard:range rdf:resource="#Prop_001_Bag"/>
</rdf:Description>
```

이렇게 지정된 프로퍼티 기술은 그림 3에서 볼 수 있는 것처럼, 대응되는 HTML 태그가 생성된다. 동물 종류의 데이터형은 나열형이기 때문에, HTML 태그로는 SELECT 태그가 사용되게 된다.

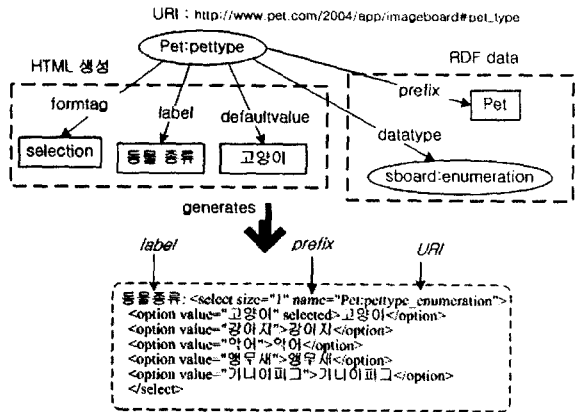


그림 3. 생성된 HTML 태그

### 3. 시스템 설계

#### 3.1 시스템 구조

제안된 시스템은 그림 4와 같은 구조를 갖고 있다. 일반적인 게시판 시스템과 달리 시맨틱 정보를 포함하는 부분이 추가되어 있다. 이미지의 업로딩은 HTML의 FILE 태그를 이용하여, 일반적인 파일 업로딩과 같은 방식을 사용한다. 단, 이미지 업로딩시에 해당 이미지에 대한 시맨틱 태깅을 할 수 있는 구조를 갖고 있다.

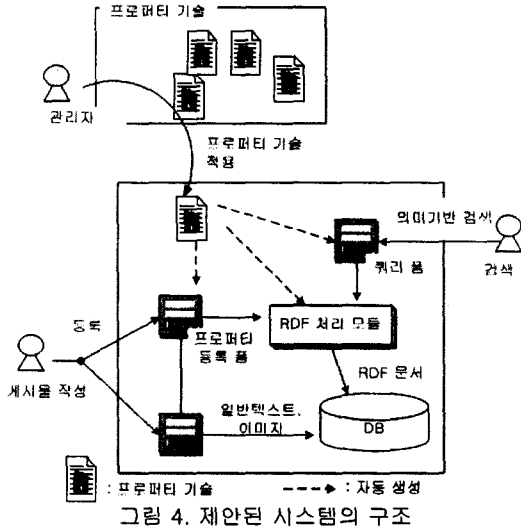


그림 4. 제안된 시스템의 구조

게시판의 관리자는 해당 보드의 이미지 종류에 따라서, 적합한 프로퍼티의 기술 문서를 게시판에 적용할 수 있다. 적용되는 프로퍼티 기술 문서는 직접 작성하거나, 이미 작성된 기술 문서를 이용하는 것이 가능하다.

#### 3.2 의미기반 이미지 검색

이미지 검색을 하는데 있어서, 제안된 시스템은 적용된 프로퍼티들을 검색에 이용하며, 이를 통해 이미지에 단순히 추가된 텍스트 정보를 이용하는 방식을 피할 수 있게 된다. 이러한 프로퍼티를 이용한 의미기반 검색은 프로퍼티의 데이터 형태에 따라서, 자동으로 질의문(query statement)이 형성되도록 구성하였다. 본 논문에서는 등록된 게시물의 시맨틱 정보가 RDF 문서로 저장되므로, 이의 검색을 위해 RDQL(RDF Data Query Language)[4]을 사용하였다. RDQL은 SQL과 비슷한 형태의 질의 언어로 RDF 문서의 검색에 적합하게 구성되어 있다.

그러나, RDQL의 경우에 일반 사용자가 다루기 어려운 질의문(query statement)을 작성해야 하는 어려움이 있다. 따라서, 질의문을 형식화된 입력을 통해서 제공하는 것이 바람직하다. 본 논문에서는 질의의 대상이 되는 프로퍼티를 *프로퍼티 기술*에서 지정하면, 자동으로 형식화된 HTML 품이 생성되는 구조를 갖도록 하였다. 이때 중요한 요소는 프로퍼티의 데이터형이다. 이것은 데이터형의 종류에 따라서, 생성되는 RDQL의 필터에 적용되는 질의연산(query operation)이 달라지기 때문이다. 표 1에는

프로퍼티의 데이터형에 따라서 HTML 품으로 생성되는 태그를 보여주고 있다.

표 1. 데이터형에 의해 생성되는 HTML 품

데이터형	품의 질의연산
수치형	GT(>), GE(>=), EQ(==), LE(<=), LT(<)
나열형	SAME
날짜형	FROM, UNTIL

본 논문에서는 프로퍼티를 나열형, 수치형, 날짜형의 세가지 데이터형을 가질 수 있다고 설계하였다. 이것은 하부의 RDF 문서 검색에 사용되는 RDQL(RDF Data Query Language)[5]이 현재 문자형과 수치형 밖에는 지원하지 않기 때문이다.

실제 의미기반 검색을 통해서 검색된 화면은 그림 5와 같다. 그림 5에서는 “고양이”라는 검색어를 주었고, 해당 의미정보를 갖고 있는 데이터를 자동으로 추출하여 화면에 보여주고 있다. 다른 형태의 프로퍼티 기술을 수용(importing)하면, 인물이나 풍경 이미지에 적합한 의미기반 검색도 가능하다.

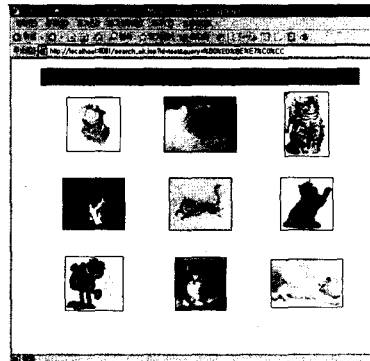


그림 5. 제안된 시스템의 모습

### 5. 결론

현재의 게시판 시스템의 의존하고 있는 렉시컬 기반의 정보 처리는 사람이 의도한 시맨틱을 효과적으로 처리할 수 없는 단점을 갖고 있다. 특히 멀티미디어가 포함된 경우에는 이런 문제점이 더욱 심화된다. 본 논문에서는 이를 해결하기 위한 방안으로 게시판에 시맨틱 태깅을 제공하는 방안과 그 시스템에 관해 설명하였다. 동적인 시맨틱 태깅의 적용을 통해 시맨틱 정보를 코드와 분리하여, 여러 다양한 온토로지를 지원하는 게시판 시스템으로 확장이 가능하도록 하였다. 본 논문에서 제시한 방법은 다른 정보시스템에서도 응용이 가능할 것으로 기대한다.

### 참고문헌

- [1] Brisaboa, N.R., et al., "A collaborative framework for a digital library," Proc.CRIWG 2000, 18-20, 2000
- [2] Hyunmo, K., et al., "Dynamic layout management in a multimedia bulletin board," Proc. Human Centric Computing Languages and Environments, 51-53, 2002
- [3] Berners-Lee, T., et al., "The Semantic Web," Scientific American, Vol.284, No.5, 34-43, 2001
- [4] Andy, S., "A Programmer's Introduction to RDQL," <http://www.hpl.hp.com/semweb/doc/tutorial/RDQL/> 2002