

## 의미 기반 정보 검색을 제공하는 멀티미디어 게시판 시스템

정의현\*

# A Multimedia Bulletin Board System Providing Semantic-based Searching

Eui-Hyun Jung\*

### 요 약

게시판 시스템은 텍스트 외에도 다양한 멀티미디어 데이터를 포함하는 구조로 발전해왔다. 그러나 기존 시스템은 사용자가 게시물의 내용을 파악하기 위해서는 많은 시간과 노력을 들여야 한다는 단점을 갖고 있다. 이런 문제를 해결하기 위해 대부분의 게시판 시스템에서 검색 기능을 제공하지만, 렉시컬 수준의 정보 접근으로 사용자의 의도에 적합한 게시물을 찾는 것은 어려운 일이다. 더욱이 게시물이 멀티미디어를 포함하는 경우에는 검색이 거의 불가능하게 된다. 본 논문에서는 이를 해결하기 위하여 시맨틱 웹을 적용한 새로운 게시판 시스템을 제안한다. 제안된 게시판 시스템은 게시물의 분야와 멀티미디어 속성을 표현할 수 있는 온톨로지를 제공한다. 또한, 사용자들이 제안된 온톨로지로 분야별 게시판 온톨로지를 기술할 수 있도록 설계되었다. 시스템의 수정 없이 다양한 분야별 의미기반 검색을 제공하기 위하여, 관리자가 기술한 게시판 온톨로지를 게시판 시스템에서 도입할 경우 자동으로 입력/검색 인터페이스와 RDF 데이터의 접근 구조가 형성되도록 구현하였다. 제안된 게시판 시스템은 사용자의 의도에 적합한 게시물 검색을 하는 의미기반 검색의 가능성과 유효성을 보여주었다.

### Abstract

Bulletin board systems have evolved to support diverse multimedia data as well as text. However, current board systems have a weakness: it takes much time and efforts for users to figure out contents of articles. Most board systems provide a searching function with lexical level data access for solving that problem, however it fails to serve users' intended searching results. Moreover, it is nearly impossible to search proper articles if they contain multimedia data. This paper proposed a bulletin board system adopting the Semantic Web to solve this issue. The proposed system provides users with new ontology which is used for describing articles' domain knowledge and multimedia features. Users can describe their own board ontology using the proposed ontology. To support semantic-based searching for diverse domain knowledge without modification of the system, the system dynamically generated input/query interface and RDF data access module according to the board ontology written by administrators. The proposed board system shows that semantic-based searching is feasible and effective for users to find their intended articles.

▶ Keyword : 시맨틱 웹(Semantic Web), 온톨로지(Ontology)

---

• 제1저자 : 정의현  
• 접수일 : 2005.10.14, 심사완료일 : 2005.10.29  
\* 안양대학교 디지털미디어공학과 교수

## I. 서론

웹(Web)이 대두되기 이전부터 게시판 시스템(Bulletin Board System)은 동일한 관심을 공유하는 사용자들 사이에 매우 유용한 정보창구였으며, 단방향적인 정보 게시형태인 웹과 달리 여러 사람의 참여로 정보 공간을 만든다는 장점을 갖고 있었다. 이러한 이유 때문에 고퍼(gopher)나 아키(archie) 등의 정보 시스템이 사용자들의 외면을 받은 것에 비해, 게시판 시스템은 자연스럽게 웹의 한 부분으로 통합되었고, 문자 중심의 게시판 시스템의 한계를 극복하기 위해 멀티미디어의 수용과 다양한 필터링(filtering) 방식을 통한 효율적인 정보 접근에 대한 연구가 많이 이루어졌다 [1][2].

그러나 기존의 게시판 시스템은 게시물의 내용을 파악하기 위해서 사용자가 일일이 내용을 확인해야 하는 구조적인 단점을 갖고 있다. 물론 기존 게시판 시스템도 검색 기능이 제공되어 특정 키워드가 포함된 검색 결과를 사용자에게 제공하지만, 이것은 단순히 렉시컬(lexical) 수준에서의 텍스트 패턴 매칭(matching)에 의한 결과라는 한계점을 갖는다 [3]. 웹과 달리 게시판은 여러 사용자들이 참여하기 때문에 사용자들은 동일한 정보를 여러 형태의 렉시컬 형태로 표시할 수 있고, 혹은 그 반대로 동일한 렉시컬이 다른 시맨틱(semantics)을 의미할 수 있다. 따라서 단순한 키워드 검색을 통해서 사용자가 원하는 의미를 포함한 정확한 검색 결과를 얻는 것은 무리가 있다. 예를 들어, "고양이(cat)"라는 정보를 어떤 사용자는 "고양이"로 표기하지만, 다른 사용자는 "나비"로 표기할 수 있는데, 이러한 경우에 "고양이(cat)"이라는 시맨틱을 "고양이"라는 렉시컬 수준으로 검색한 결과는 "나비"나 기타 다른 렉시컬로 표기된 결과를 포함할 수 없게 된다. 이와 반대로 "나비"라고 렉시컬 수준 검색을 하게 되면, "고양이(cat)"와 "나비(butterfly)"의 두 가지 시맨틱을 가진 결과가 나오게 되어, 사용자가 의도한 정보라는 것을 보장할 수 없다.

따라서 기존 게시판 시스템에서는 사용자가 일일이 문서의 내용을 확인해야만 정보 교환이 가능했는데, 참여도가 높은 게시판의 경우에는 게시되는 게시물의 양이 많기 때문에 이러한 수작업을 통한 정보의 접근 방식은 사용자의 시간과

노력을 많이 요구하게 된다. 더욱이 텍스트 패턴 매칭으로 검색하기 어려운 이미지나 비디오와 같은 멀티미디어 데이터가 게시물에 포함되어 있는 경우에는 기존 검색 방식으로 데이터를 찾아내는 것은 매우 어려운 문제이다[3].

본 논문에서는 이러한 문제점을 해결하기 위한 방식으로 시맨틱 웹(Semantic Web)[4][5] 기술을 적용한 게시판 시스템의 구조에 관해 제안한다. 시맨틱 웹을 적용하는 경우 게시물에 의미기반 태그(tag)를 추가하는 것이 가능하기 때문에 사용자가 의도한 정보 검색을 할 수 있으며 멀티미디어 데이터에 대한 정확한 검색 결과를 얻을 수 있게 된다. 시맨틱 웹 기술을 게시판 시스템에 적용하기 위하여 새로운 온톨로지(Ontology)인 SBoard 온톨로지를 설계하였으며, SBoard 온톨로지 기술(description)된 게시판 온톨로지(Board Ontology)에 근거하여 게시물에 대한 다양한 의미기반 입출력을 지원하는 게시판 시스템을 구현하였다.

SBoard 온톨로지는 다양한 분야의 온톨로지를 조합하여 게시판 온톨로지를 기술할 수 있도록 설계되었으며, 구현된 게시판 시스템은 게시판 온톨로지에 대응되는 시맨틱 입력 및 검색을 제공하는 HTML 폼의 자동생성을 지원한다. 또한, 입력된 시맨틱 데이터를 게시판 온톨로지에 대응되는 RDF 문서 데이터 구조와 검색에 필요한 RDQL[6] 쿼리(query)를 생성하는 기능을 갖도록 구현하였다.

본 논문은 다음과 같은 구성을 갖는다. 먼저 2장에서는 기존 연구들과 시맨틱 웹의 역할, 그리고 시맨틱 웹의 게시판 시스템 적용시의 기술적 고려사항에 대해서 살펴본다. 3장에서는 설계된 SBoard 온톨로지의 구조와 그 역할에 대해서 논하며, 4장에서는 제안된 게시판 시스템의 구조에 대해서 살피고, 그 효용에 대해서 평가하며, 5장에서 결론을 맺는다.

## II. 기존 연구 및 기술적 고려사항

### 2.1 기존 연구 및 시맨틱 웹

사용자간의 비동기적 정보 교환을 지원하는 시스템은 크게 E-mail, 게시판 시스템, 검색 아카이브(searchable storage archive)로 구분할 수 있다[7]. 이 시스템의 분류는 프로토콜과 응용의 형태와는 별도로 커뮤니티(community)의 폐쇄

성에도 영향을 받는다. 즉, E-mail은 친밀도와 상관없이 임의의 사용자간의 메시지 전달에 사용되는데 반해, 검색 아카이브의 경우에는 회사나 조직 같이 사용자간의 결합도가 더 높은 경우에 주로 사용된다. 이에 비해 게시판은 같은 관심사를 가진 사용자들에게 커뮤니티 결합력을 제공하면서, 동시에 유연한 사용자의 접근이 가능하여 여러 정보 시스템의 기본적인 형태로 제공되고 있다. 또한, 게시판 시스템은 웹과 결합되면서 단순한 텍스트 정보 외에 멀티미디어 데이터를 포함하는 복합적인 형태로 발전되고 있다. 이런 속성 때문에 게시판에 대한 연구는 교육(8), 행정(9) 등의 다양한 분야에서 사용되고 있으며, 멀티미디어 기반의 게시판 시스템(1)에 대한 연구도 이루어지고 있다. 그러나 지금까지의 게시판 시스템은 사용자가 관심 있는 게시물을 모두 일일이 확인해야 하는 문제점을 갖고 있다. 이런 문제점은 게시물의 정보가 기계가 이해할 수 없는 정보라는 것에 기인한다. 게시물의 정보는 사람만이 이해할 수 있는 의미이기 때문에, 사용자가 원하는 정보를 찾기 위해서는 많은 시간과 노력을 기울여야 한다.

그러나 이러한 문제점은 단순히 게시판 시스템만의 문제는 아니며, 현재의 정보 시스템은 대부분 같은 문제점을 안고 있다. 대표적으로 웹의 경우에도 검색엔진을 통해 검색된 결과는 대부분 사용자의 요구사항을 포함하는 의미 중심의 검색보다는 렉시컬 수준의 키워드 검색에 근거한 결과가 반환된다(3)(4). 이것은 지금까지의 정보 시스템이 시맨틱 중심의 데이터 처리가 아니고 렉시컬 중심의 데이터 처리를 이용하기 때문이다. 이것은 정보 시스템에 저장되는 정보들이 기계 입장에서는 단순한 바이트(byte) 정보로 저장되기 때문이다. 따라서 사용자가 정보를 검색하는 경우, 기계는 단순히 주어진 키워드의 대한 패턴 매칭에 근거한 접근만을 제공하며, 사용자가 의도한 의미(semantic)를 정확히 파악하여 처리하는 것은 어려운 문제이다. 이를 해결하기 위해 웹의 검색 시스템은 개인화 기법(Personalization)(10)과 페이지 랭크(Pagerank)(11)와 같은 다양한 방법론이 나왔지만, 근본적인 해결 방법은 되지 못하고 있다.

이러한 문제를 해결하기 위해서 여러 연구들이 있었는데, 현재 가장 주목을 받고 있는 것은 시맨틱 웹(Semantic Web)이다(5)(12). 시맨틱 웹은 Tim Berners-Lee에 의해 제안된 개념으로 웹을 구성하는 리소스들에 특정한 방식으로 시맨틱 정보를 추가하는 새로운 개념의 웹이다. 기존 웹이 사람에게 정보를 전달하기 위한 표현(presentation) 중심의 웹이라면, 시맨틱 웹은 기계가 의미(meaning)를 파악하고 자동화된 처리를 할 수 있는 웹이라고 할 수 있다

[13]. 따라서 시맨틱 웹을 정보 시스템에 적용할 경우 기계에 의한 추론, 기계간의 시맨틱 전달, 그리고 시맨틱에 근거한 정확한 정보 접근이 가능한 특징을 갖고 있다. 시맨틱 웹에 대한 연구는 초기 단계이지만, W3C를 중심으로 매우 활발하게 표준화 활동이 일어나고 있으며, 기존의 응용들과는 접근방법이 다른 MusicBrainz(14)나 RDF-PIC(15)과 같은 응용들이 발표되고 있다.

## 2.2 기술적 고려 사항

시맨틱 웹이 정보 시스템에서 시맨틱을 처리할 수 있는 기능을 제공하지만 게시판 시스템에 적용하기 위해서는 서로 대치되는 기술적 난점을 해결해야 한다. 기본적으로 시맨틱 웹을 정보 시스템에 적용하기 위해서는 각 리소스에 웹 저술자(author)가 태깅(tagging) 작업을 해주어야 하며, 이를 위해 저술자가 전문적인 지식을 보유해야 한다. 이에 비해 게시판이나 지식베이스(Knowledge Base)와 같이 다수의 참여자가 구축하는 정보 시스템의 경우에는 사용자가 수작업으로 시맨틱 태깅하는 것을 피하고 있다. 이는 수작업으로 시맨틱 태깅을 하게 되면, 시스템 참여자에게 전문적인 지식을 요구하게 되며, 또한 입력 시의 오류 발생 가능성 때문에 시스템의 일관성(consistency)을 해칠 수 있기 때문이다. 따라서 이러한 다수 참여자 정보 시스템에서는 사전에 정의된 온톨로지에 맞는 입력 인터페이스를 통해 시맨틱 데이터를 추가할 수 있게 하는 것이 일반적인데, 게시판 시스템은 웹 정보 시스템의 특성상 HTML 폼(form)으로 시맨틱 데이터를 입력받아야 한다.

그러나 이렇게 HTML 폼 형태로 게시판 시스템이 시맨틱 데이터의 입력을 지원하는 경우에는 적용될 분야의 시맨틱 정보 입력이 게시판 시스템의 코드에 고착되어야만 한다. 이러한 고착이 문제가 되는 이유는 게시물이 시맨틱 데이터를 지원하기 위해서는 온톨로지별로 다른 술어(predicate)들을 필요로 하기 때문이다.

예를 들어, (그림 1)에서 볼 수 있는 것처럼 회의 관련 게시판의 온톨로지에는 회의 개최자, 회의 날짜 등이 중요한 술어가 되지만, 사진 관련 게시판의 온톨로지에는 사진의 종류, 카메라 종류 등이 중요한 술어가 될 수 있다. 그러나 기존 게시판 시스템의 구조에서는 온톨로지가 시스템에 고정되는 문제를 가지게 되므로, 여러 온톨로지를 지원하기 위해서는 매년 새로운 입력 인터페이스와 저장 데이터베이스 구조, 그리고 게시판 코드를 개발해야 한다. 즉, 다중 사용자 환경에서 온톨로지의 오용(誤用)을 피하기 위해 불가피한 "형식화된 입력의 제공"이 단일한 게시판 시스템이 "다

양한 술어 집합"을 지원하기 위한 재사용성 측면에서 걸림돌이 되는 것이다.

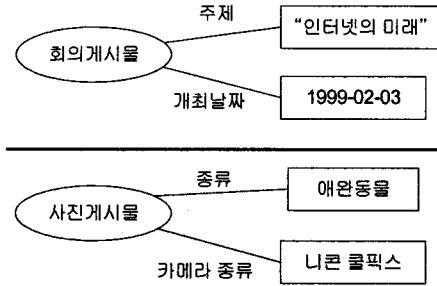


그림 1. 게시 분야별 술어  
Fig 1. Predicates dependent on Domains

이를 해결하기 위해서는 개별 온톨로지의 특성에 따라서 자동으로 입력 인터페이스, 저장 데이터베이스 구조가 생성되는 게시판 시스템의 구조와 게시판에 적용될 개별 온톨로지를 기술(description)할 수 있는 기술 방식이 필요하다. 본 논문에서는 3장에서 기술 방식에 대해서 설명하고, 4장에서 게시판 시스템의 구조에 관해서 논한다.

### III. SBoard 온톨로지

#### 3.1 개요

게시판 시스템에서 여러 술어 집합이 동적으로 지원되기 위해서는 술어 집합을 게시판 시스템에서 사용할 수 있도록 기술하는 방법이 필요하다. 본 논문에서는 이 목적을 위하여 SBoard라는 새로운 온톨로지를 설계하였다. SBoard 온톨로지의 주요한 목적은 게시물을 리소스로 하는 술어 집합들의 개별 술어에 대해 입력/검색폼을 생성하고, RDF 문서로 저장할 때 필요한 정보를 기술하는 것이다. 먼저, 웹 인터페이스의 생성과 관련된 정보는 폼 입력 컴포넌트의 종류, 라벨, 기본값, 입력 컴포넌트의 변수 이름, 검색 대상 여부가 중요하다. 이 정보들을 이용하여 게시판 시스템은 적절한 입력/검색 폼을 만들 수 있다. 둘째로 RDF 문서로 저장되는 부분에서 중요한 정보는 URL과 Prefix, 데이터 형이다.

예를 들어, Dublin Core 온톨로지의 creator 술어를 게시물에 사용하기 위해 게시물 관리자는 (그림 2)의 술어 기술 부분에서처럼 필요한 정보들을 명시해야 한다. 이렇게 명시된 술어는 그에 대응되는 HTML 문서(웹 인터페이스 파트)와 저장되는 구성요소(RDF 문서 파트)를 생성하는데 이용된다. 그리고 서버 코드에서 처리하기 위한 기본 변수명은 술어와 Prefix, 데이터형의 결합된 형태로 나타나게 된다. 실제 HTML의 입력에 John이라는 값을 넣으면, (그림 2)에서 볼 수 있는 것처럼 대응되는 RDF XML 문서로 변환되어 저장되게 되어, 검색 시 의미기반 처리가 가능하게 된다.

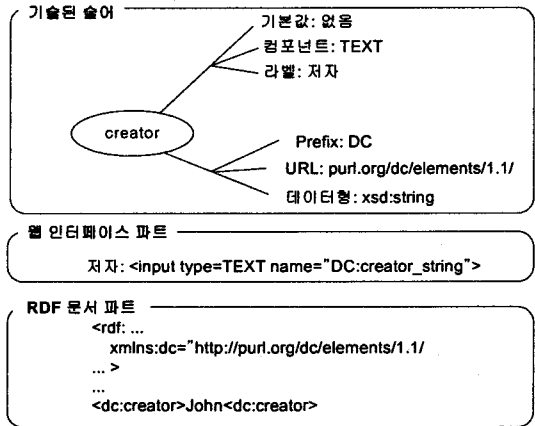


그림 2. 술어 기술에 필요한 정보들  
Fig 2. Information needed for describing a Predicate

이렇게 술어들에 대한 정보를 명시하기 위해 정의된 SBoard 온톨로지는 RDF 문서 파트와 웹 인터페이스 파트의 2개의 부분으로 구성된다. RDF 문서 파트는 개별 술어에 대해 사용자가 입력한 값을 저장할 때 필요한 정보를 가지고 있다. 그리고 웹 인터페이스 파트는 개별 술어가 HTML 폼 태그로 만들어질 때 필요한 정보를 제공하게 된다. (그림 3)은 게시판 온톨로지를 SBoard 온톨로지로 명시하였을 때, 각 파트가 어떻게 연동되는지를 나타낸 그림이다. 게시물 온톨로지에서는 게시물에 필요한 정보를 추가하기 위한 술어별로 두 파트들에 대한 정보를 기술하도록 구성되어 있다.

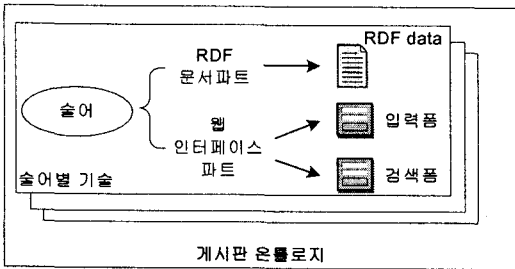


그림 3. 게시판 온톨로지 내의 술어 기술  
Fig 3. Predicate Description in Board Ontology

설계된 SBoard 온톨로지는 (그림 3)과 같은 구조를 가짐으로써, 다수의 외부 온톨로지의 술어들을 혼재해서 게시물 온톨로지를 정의하는데 사용할 수 있으며, 각 술어에 대해 형식화된 입력/검색폼을 만드는 것이 가능하다. <표 1>은 SBoard 온톨로지의 술어에 대해 설명한 표이다.

표 1. SBoard 온톨로지 술어  
Table 1. Predicates in the SBoard Ontology

분류	로컬 이름	기능
RDF	datatype	해당 술어의 데이터형을 결정한다. XSD의 날짜, 문자, 수치형의 데이터형을 가진다.
	prefix	해당 술어의 prefix를 지정한다
웹 인터페이스	formtype	해당 술어를 입력에 사용할지, 검색에 사용할지를 결정한다. inputform과 queryform, both의 값을 가질 수 있다. both의 경우에는 입력과 검색 모두에 사용이 가능하다.
	label	HTML 입력 컴포넌트 앞에 붙는 문자열을 지정한다.
	formtag	입력 컴포넌트를 어떤 것으로 선정할지를 지정한다. text, selection, textarea, hidden의 태그가 지원된다.
	range	formtag가 selection 컴포넌트인 경우에만 사용되며, selection 컴포넌트에 들어갈 데이터를 명시한다. 이를 사용하기 위해서 RDF:Bag을 이용한다.
	default	입력 컴포넌트에 사전에 들어가야 할 값을 지시하는 태그이다.

### 3.2 온톨로지 사용 예: DC:Creator

SBoard 온톨로지서 정의된 술어를 설명하기 위해, DC:creator 술어에 대해 기술한 예를 보기로 한다. 여기서는 DC:creator 술어를 웹을 통해 입력받기 위해, 입력 컴포넌트는 TEXT 컴포넌트로 지정하며, 저장 데이터는 xsd:string으로 저장하는 것으로 가정한다.

```

<!-- rdf:about에는 DC:creator에 대한 URI가 들어간다. -->
<rdf:Description rdf:about="http://purl.org/dc/elements/1.1/creator">
  <sboard:formtype>inputform</sboard:formtype>
  <sboard:datatype>xsd:string</sboard:datatype>
  <sboard:prefix>DC</sboard:prefix>
  <sboard:label>저자</sboard:label>
  <sboard:formtag>text</sboard:formtag>
  
```

기술된 내용을 살펴보면, rdf:about은 DC:creator 술어의 URI에 해당하며, 술어의 이름인 creator와 URL을 얻는데 사용된다. 그리고 formtype이 "inputform"으로 지정되어 있으므로, 해당 술어가 입력폼에 사용되는 것을 알 수 있다. 그리고 prefix는 RDF 문서 작성과 HTML 입력 컴포넌트의 name 속성을 구성하는데 필요한 정보들이다. datatype은 "xsd:string", label이 "저자"로 formtag가 "text"로 지정되었으므로, (그림 4)와 같은 입력 컴포넌트에 대응되는 HTML 문서가 생성된다.

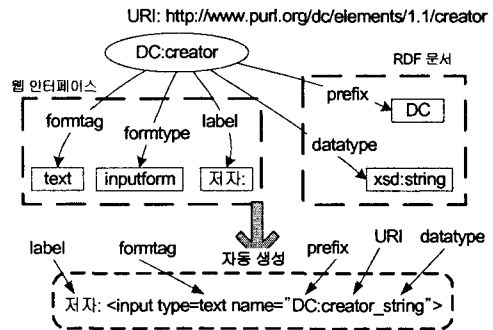


그림 4. 술어 기술에 대한 HTML 문서 생성  
Fig 4. Generation of HTML Document from Predicate Description

## IV. 시스템 구조

### 4.1 전체 구조 및 클래스 관계

일반적으로 웹 기반 게시판 시스템은 파일 시스템이나 데이터베이스를 CGI(Common Gateway Interface) 코드로 연결함으로써 구축할 수 있다. 그러나 시맨틱 기반의 웹 서비스는 사용자가 입력하는 텍스트 정보 외에 해당 게시물에 대한 시맨틱 정보를 따로 저장할 수 있어야 한다. 또한, SBoard 온톨로지로 기술한 게시판 온톨로지를 동적으로 적용할 수 있는 구조가 필요하다. 이를 위해 본 논문에서 제시한 게시판 시스템은 일반적인 구조와 달리 (그림 5)와 같은 형태의 구조를 갖고 있다.

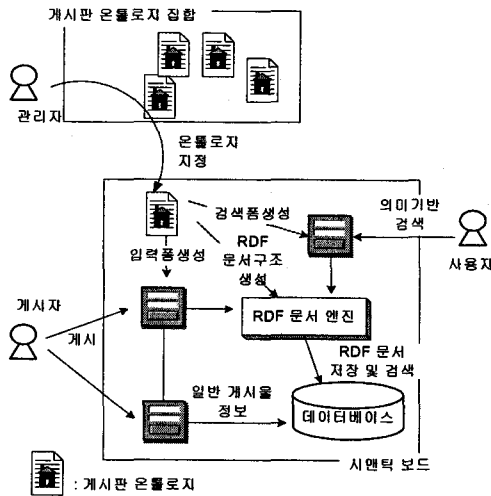


그림 5. 시스템 내부 구조  
Fig 5. Internal Structure of the System

기본적으로 관리자는 게시판 시스템에서 사용될 게시판 온톨로지를 시스템에 지정한다. 이때 게시판 온톨로지는 인터넷에서 구할 수도 있고, 관리자가 직접 정의할 수도 있다. 시스템은 시작 시점에 지정된 게시판 온톨로지를 이용해서 입력/검색 화면의 생성과 RDF 문서 저장 모듈을 동적으로

생성한다. 입력폼과 검색폼의 경우에는 게시물 작성/검색 화면에 자동으로 결합되도록 구성되었다. 이러한 방식을 이용하게 되면, 관리자가 지정하는 어떤 분야의 술어집합이라도 게시판 시스템에서 지원이 가능하게 된다. 또한, 여러 분야의 온톨로지를 결합한 새로운 술어집합의 사용도 가능하다.

구현된 게시판 시스템은 JSP(Java Server Page)로 만들어져 있으며, 관리자가 지정한 온톨로지를 기술한 문서를 읽어 전체 게시판의 입력/검색/RDF 저장 및 검색을 담당하는 부분은 자바빈즈(Java Beans)로 구성되어 있다. 시스템의 간략화된 클래스 다이어그램은 (그림 6)과 같다.

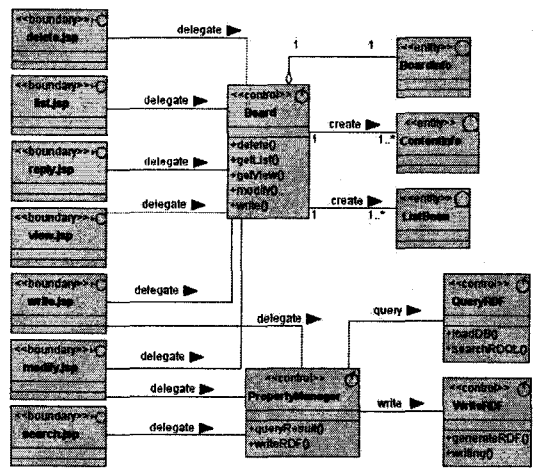


그림 6. 시스템의 클래스 다이어그램  
Fig 6. Class Diagram of the System

컨트롤 기능을 담당하는 Board 클래스는 일반 게시물에 대한 정보를 처리하는 모듈이다. 이 클래스는 외부의 JSP 파일의 요청을 받아서, 화면에 보일 데이터를 만들거나, 게시물의 목록을 만들어주는 역할을 한다. 이에 비해서 시맨틱 정보를 관리하는 주요 역할은 PropertyManager 클래스가 담당한다. 문서를 저장하고 입력 폼을 생성하는 부분은 WriteRDF 클래스가 하며, 검색 처리와 검색폼을 생성하는 부분은 QueryRDF 클래스가 담당하게 된다. 그리고 RDF 문서의 저장과 검색은 HP의 Jena[16] 클래스 라이브러리와 MySQL[17] 데이터베이스를 연동하여 처리하였다.

### 4.2 검색식의 구성

SBoard 온톨로지에 의해 기술된 입력폼이 RDF 구성에 해당하는 술어들을 생성하여 저장하는 다소 단순한 구조인

데 반해, RDF 문서의 검색을 처리하는 것은 그리 용이한 것이 아니다. 이것은 RDF 문서를 검색하기 위해서는 RDQL(6)과 같은 검색 언어(query language)를 사용해야 하기 때문인데, 일반 사용자들이 이러한 검색 언어를 사용하는 것은 그리 쉬운 일이 아니다. 따라서, 본 연구에서는 <표 2>에서 예시한바와 같이 데이터형 별로 검색 메서드를 만들었다. 이 검색 메서드는 게시판 온톨로지에서 지정된 데이터형에 따라서 검색폼에서 자동으로 검색 메서드가 selection HTML 컴포넌트로 지정되게 된다.

표 2. 데이터형에 따른 검색식 메서드  
Table 2. Query Method corresponding to Data Type

데이터형	메서드
수치형	GT(>), GE(>=), EQ(=), LE(<=), LT(<)
문자형	SAME
날짜형	FROM, UNTIL

이 검색 메서드를 실제 RDF 문서 검색에서 지원하기 위해서는 적절한 메커니즘이 필요하다. 본 논문에서는 HP에서 개발된 RDQL(6)과 제안된 검색 메서드를 매핑하여 이를 해결하였다. RDQL은 SQL과 비슷한 형태로 RDF 문서를 트리플(triple) 형태로 검색이 가능하도록 구성된 식이다. 그러나 RDQL은 RDF에 대한 검색 기능이 제공되지는 않지만, 아직까지는 표준 SQL처럼 여러 데이터형에 대해서 검색 기능이 제공되지는 않고 있다. 현재 지원되는 검색 데이터형은 수치형, 문자형이다. 따라서 날짜형은 수치형을 변형하여 사용하였다.

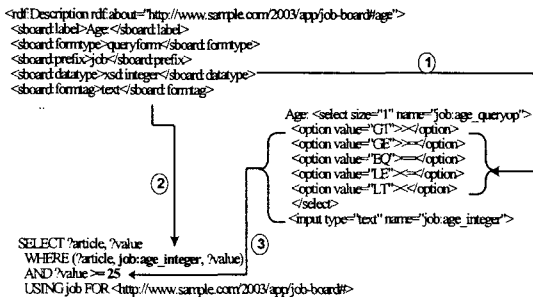


그림 7. 게시판 온톨로지, 검색 HTML 폼, RDQL 검색식의 관계  
Fig 7. Relation among Board Ontology, Query HTML Form, and RDQL Query Statement

각 검색 메서드는 해당 술어의 데이터형에 의해서 자동으로 HTML에 지정된다. 예를 들어, xsd:integer인 "age"라는 술어를 검색 술어로 지정하면, (그림 7)에서 ①로 표시된 것과 같이 자동으로 수치형에 대한 검색 메서드가 검색 HTML 폼 안에 생성된다. 그리고 생성된 검색 폼에서 사용자가 검색 메서드를 "GE"로 주고, 나이를 "25"라는 값을 주었을 때, 게시판 온톨로지인 ②에서 검색의 대상에 해당되는 술어를 얻고, 실제 입력된 ③의 값으로 검색의 조건을 이용하여 RDQL 문장이 생성된다. 이렇게 게시판 온톨로지서 자동으로 검색 HTML 폼이 생성됨으로써 사용자가 복잡한 RDQL을 직접 사용할 필요 없이, 검색 HTML 폼에서 의미 기반 검색을 할 수 있게 된다.

### 4.3 시스템 평가

본 논문에서 제안한 SBoard 온톨로지와 게시판 시스템이 설계 목적에 적합하게 동작하는 지를 확인하기 위해, 이미지 파일을 포함할 수 있는 멀티미디어 게시판에 두 개의 게시판 온톨로지를 정의하여 실제 적용이 가능한지 평가를 하였다. 첫 번째 게시판 온톨로지는 애완동물 게시판 온톨로지이다. 이 온톨로지에서는 애완동물의 종류를 술어 집합으로 선택하였으며, 기술된 내용은 다음과 같다.

```

<rdf:Description
  rdf:about="http://www.pet.com/2005/pet-board#type">
  <skboard:formtype>both</skboard:formtype>
  <skboard:prefix>pet</skboard:prefix>
  <skboard:datatype>xsd:string</skboard:datatype>
  <skboard:label>애완동물종류</skboard:label>
  <skboard:formtag>selection</skboard:formtag>
  <skboard:range>
    <rdf:Bag>
      <rdf:li rdf:resource="http://www.pet.com/pettype/고양이"/>
      <rdf:li rdf:resource="http://www.pet.com/pettype/강아지"/>
      <rdf:li rdf:resource="http://www.pet.com/pettype/새"/>
    </rdf:Bag>
  </skboard:range>
</rdf:Description>
  
```

애완동물 게시판 온톨로지를 지정하였을 때 나온 입력 폼과 검색 폼은 (그림 8)과 같다. 게시판 온톨로지에는 애완동물의 종류를 나타내는 술어가 입력/검색 화면에 모두 사용되도록 지정되어 있다. (그림 8)을 보면 알 수 있듯이 지정된 술어에 대해서 자동으로 의미 기반 태그를 추가할 수 있도록 HTML 폼 들이 생성된 것을 확인할 수 있다.

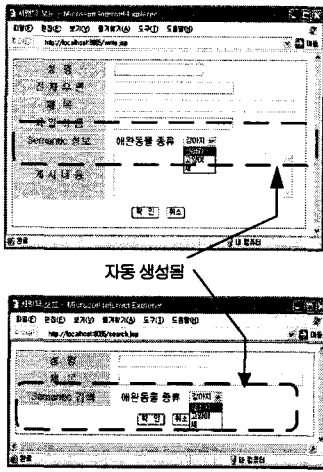


그림 8. 게시판 온톨로지에 의한 폼 생성  
Fig 8. Form Generation from Board Ontology

실제로 RDF 문서에 대한 의미 기반 검색이 가능한지 확인하기 위해, 검색 폼에서 제목에 "귀여운"을 주고, 애완동물 종류 술어에 "강아지"로 검색하였으며, 그 결과는 (그림 9)와 같다. 게시판 내용 중에는 귀여운 고양이나 새와 같은 애완동물의 게시물도 있었지만, 의미기반 검색에 의해서 강아지의 의미(semantic)를 가진 "귀여운" 애완동물(강아지)에 대한 검색 결과를 얻을 수 있었다. 만일 "귀여운 강아지"라는 텍스트 패턴만으로 검색하였다면, 귀여운 개나 귀여운 멍멍이 등의 제목을 가진 문서들은 검색 결과로 얻을 수 없을 것으로 평가된다.



그림 9. 의미기반 검색 결과  
Fig 9. Results from Semantic-based Searching

게시판 시스템의 코드의 수정 없이 게시판 온톨로지의 교체만으로 새로운 의미기반 검색이 가능함을 보여주기 위해, 본 논문에서는 새로운 게시판 온톨로지를 기존 게시판 시스템에 설정하였다. 설정된 게시판 온톨로지는 컴퓨터 부품 판매를 위한 온톨로지로서 컴퓨터 부품의 종류, 가격을 술어로 정의하였다.

```

<rdf:Description
  rdf:about="http://www.comsell.com/2005/com-board#price">
  <sbboard:formtype>inputform</sbboard:formtype>
  <sbboard:prefix>comsell</sbboard:prefix>
  <sbboard:datatype>xsd:integer</sbboard:datatype>
  <sbboard:label>가격</sbboard:label>
  <sbboard:formtag>text</sbboard:formtag>
</rdf:Description>
<rdf:Description
  rdf:about="http://www.comsell.com/2005/com-board#device">
  <sbboard:formtype>both</sbboard:formtype>
  <sbboard:prefix>comsell</sbboard:prefix>
  <sbboard:datatype>xsd:string</sbboard:datatype>
  <sbboard:label>부품종류</sbboard:label>
  <sbboard:formtag>selection</sbboard:formtag>
  <sbboard:range>
    <rdf:Bag>
      <rdf:li rdf:resource="http://www.comsell.com/device/본체"/>
      <rdf:li rdf:resource="http://www.comsell.com/device/마우스"/>
      <rdf:li rdf:resource="http://www.comsell.com/device/키보드"/>
      <rdf:li rdf:resource="http://www.comsell.com/device/모니터"/>
    </rdf:Bag>
  </sbboard:range>
</rdf:Description>
  
```

이 온톨로지를 적용하고 시스템을 재시동하였을 때, (그림 10)과 같은 입력/검색 화면이 자동으로 생성되었으며, 설계 시에 가정했던 게시물의 시맨틱 태깅과 게시물에 추가된 시맨틱 정보에 근거한 검색이 시스템 코드의 수정이 전혀 없이도 가능함을 확인하였다.



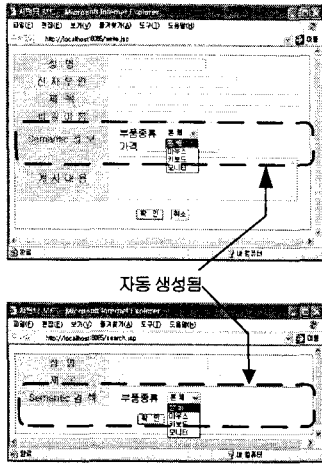


그림 10. 폼 생성  
Fig 10. Form Generation

이러한 실험 결과는 제안된 시스템이 “형식화된 입력의 제공”을 유지하면서 동시에 “다양한 슬어집합”의 지원이 가능함을 의미한다. 특히, 사진과 같은 멀티미디어 데이터가 게시물에 포함된 경우, 기존 시스템은 게시물에 첨부된 텍스트 패턴 매칭만으로 데이터를 찾는데 비해, 본 시스템은 사진에 포함된 속성(애완동물의 종류나 컴퓨터 부품의 종류)에 대한 시맨틱 태깅과 해당 속성에 맞춘 검색이 가능함을 확인하였다. 물론 게시판 온톨로지 설계 시에 다양한 속성이 포함되지 않으면, 세밀한 의미 기반 검색이 어려운 단점을 갖고 있으나, 이것은 게시판의 관리자가 다양한 온톨로지를 포함할 수 있도록 구성하는 것에 따라 보완될 수 있을 것이라 판단된다.

## V. 결론 및 향후 과제

기존의 게시판 시스템에서 원하는 게시물을 검색하기 위해서는 렉시컬 기반의 키워드 매칭만을 이용할 수밖에 없었다. 그러나 의미 중심적인 검색과 접근이 점차적으로 중요도를 더해가는 시점에서 시맨틱 웹 기술을 이용한 정보 저장 및 접근은 연구자들의 많은 관심을 불러일으키고 있다. 이러한 이슈를 해결하기 위해 본 논문에서는 시맨틱 웹 기술을 적용하여 게시물의 의미 기반 검색을 지원하는 게시판

시스템의 구조를 제안하였다. 다양한 분야의 의미 기반 검색이 가능하도록 하기 위해 새로운 온톨로지인 SBoard 온톨로지를 제안하였으며, SBoard 온톨로지로 작성된 게시판 온톨로지를 적용하여 시스템의 변화 없이 바로 다양한 분야의 의미 기반 처리가 가능하도록 시스템을 구현하였다. 제안된 시스템은 사용자가 작성한 게시판 온톨로지를 분야별로 적용할 수 있으며, 적용된 온톨로지를 기반으로 게시물 작성시에 시맨틱 정보를 태깅할 수 있는 입력폼과 검색폼을 자동으로 생성하는 기능을 제공한다. 따라서 지식 수준이 다른 일반 사용자들의 편리한 시맨틱 태깅을 위한 형식화된 입력을 지원하며, 동시에 임의적인 시맨틱 태깅으로 인한 오류를 방지할 수 있어, 자연스러운 시맨틱 기반의 정보 접근을 제공한다. 또한, 게시판에 시맨틱 태그가 고정되지 않으므로, 다양한 분야의 게시판에서 시스템의 수정 없이 독자적인 온톨로지 적용이 가능하다. 본 논문에서 제시한 기법은 다른 종류의 정보 시스템에도 동적인 시맨틱 정보를 추가할 때 유용한 접근 방안이 될 수 있을 것이라 기대된다.

## 참고문헌

- [1] Hyunmo, K. and Shneiderman, B., and Wolff, G.J., “Dynamic layout management in a multimedia bulletin board,” Proc. IEEE 2002 Symposia on Human Centric Computing Languages and Environments, pp.51 -53, 2002
- [2] David, G., David, .A.N, Brian, .M.O, and Douglas, B.T, “Using collaborative filtering to weave an information tapestry,” CACM, Vol.35, No.12, pp.61-70, 1999.
- [3] 윤보현, 서창호, “개념 속성기반 정보검색,” 컴퓨터정보학회 논문지 제 10권 3호, 2005
- [4] Lawrence,S. and Giles,C.L., “Searching the Web: general and scientific information access,” IEEE Communications Magazine, Vol.37, No.1, pp.116 -122, 1999
- [5] Berners-Lee,T., Hendler, J., and Lassila,O, “The Semantic Web,” Scientific American, Vol.284, No.5, pp.34-43, 2001

- [6] Andy, S., "A Programmer's Introduction to RDQL," <http://www.hpl.hp.com/semweb/doc/tutorial/RDQL/>, 2002
- [7] Rose, D.E., Borenstein, J.J., and Tiene, K., "MessageWorld: A new Approach to Facilitating Asynchronous Group Communication," Proc. CIKM'95, pp.266-273, 1995
- [8] Kato, S., Ota, G., and Akahori, K., "Differences between communications using e-mail and BBS in an e-counseling environment," Proc. Computers in Education-2002, pp. 423-427, 2002
- [9] Hong, H.C. and Chen, Y.C. "Design and implementation of a Web-based bulletin system for official documents," Proc. COMPSAC-2000, pp.60-65, 2000
- [10] 김은수, 송강수, 이원돈, 송정길, "웹 마이닝을 이용한 개인 광고기법에 관한 연구," 컴퓨터정보학회 논문지 제 8권 4호, 2003
- [11] Swartz, A., "MusicBrainz: A Semantic Web Service," IEEE Intelligent Systems, Vol.17, No.1, pp.76-77, 2002
- [12] Heflin, J. and Hendler, J., "A Portrait of the Semantic Web in Action," IEEE Intelligent Systems, Vol.16, No.2, pp.54-59, 2001
- [13] Hendler, J., "Agents and the Semantic Web," IEEE Intelligent Systems, Vol.16, No.2, pp.30-37, 2001
- [14] Lafon, Y. and Bos, B., "Describing and retrieving photos using RDF and HTTP," <http://www.w3.org/TR/2002/NOTE-photo-rdf-20020419>, 2002
- [15] Brin, S. and Page, L., "The Anatomy of a Large Scale Hypertextual Web Search Engine," Proc. The Seventh International WWW Conference, pp.107-117, 1998
- [16] McBride, B., "Jena: a semantic Web toolkit," IEEE Internet Computing, Vol.6, No.6, pp.55-59, 2002
- [17] MySQL Home Page, "www.mysql.com," 2003[1]  
Ian F.Akyldiz, Weilian Su, Yogesh Z Survey on Sensor Networks," IEEE Communications Magazine, vol. 40, no. 8, pp. 102-114, 2002.

## 저자 소개



### 정의현

1999년 2월 한양대학교 전자공학박사  
 2004년~현재 한양대학교 디지털미디어  
 공학과 전임강사  
 <관심분야> 시맨틱 웹, 디지털  
 컨버전스