

# 시맨틱을 이용한 연구 논문 검색 시스템

## Semantic based Research-Paper Searching System

김 영 민\*  
Young-Min Kim

이 상 준\*\*  
Sang-Joon Lee

### 요 약

웹의 발달로 정보의 수요가 증가하고 이를 위한 통합, 대용량 정보 저장 시스템들이 증가추세에 있다. 이런 정보 저장 시스템의 크기가 커져감에 따라 기존의 키워드 기반 논문 검색 방법은 검색결과의 양도 뿐만 아니라 검색자의 의도를 반영한 검색을 수행하기가 어렵다. 본 연구에서는 공학 논문들을 대상으로 논문 제목의 구성 형태를 분석하고, 제목 내의 키워드들의 역할 정보들을 RDF 시맨틱으로 구성하여 논문검색에 이용하는 방법을 제안한다. XML형태의 시맨틱을 이용하여 논문 검색에 이용한 결과 키워드 만을 이용하는 기존 방법보다는 훨씬 검색자의 의도를 잘 반영하면서도 필요한 결과만을 얻을 수 있었다.

### Abstract

Many information storage systems, such as database system, were needed to integrate much informations into one system and to provide more voluminous lump of informations. But as the size of information system becomes larger, the responded result size of existing keyword based searching system might be too large and couldn't do the exact search which the user intends to. In this study, we proposed a paper searching method which uses RDF semantic. For this, we analyzed the structural forms of the title of research papers and reconstructed them into RDF/XML. When we use these RDF descriptions of the titles to search papers, we could get more precise and accurate results than keyword based searching method.

· Key words : Semantic Search, Paper Search, RDF, Jena API

### 1. 서 론

데이터베이스와 웹의 발달과 함께 정보의 중요성이 부각되고 정보의 저장 활동이 활발하게 이루어지고 있다. 이와 더불어 저장된 정보들은 정보 이용자에 의해 검색되고 이용될 수 있어야 한다. 정보 검색에 있어서 가장 중요한 것은 정확한 결과와 정보 검색자의 의도를 반영하는 검색에 수행되어야 한다는 것이다. 하지만 기존 정보 저장시스템들은 여전히 검색 결과에 있어서 불필요한 정보까지도 제공하는 일이 허다하고, 정확한 검색자의 의도를 반영하기에는 아직 역부족이다.

현재 정보 검색의 방법으로는 문서 생성자에

의해 미리 정의된 키워드를 대상으로 검색하는 방법, 정보 데이터 내에서 주제어 및 주제문장을 찾아내게 하고 이를 통해 문서를 검색하는 주성분 분석방법을 이용하는 방법[1], 질의에 사용된 단어와 연관된 단어들까지 질의어에 포함시켜 검색하는 질의어 확장 검색방법, 질의에 사용되는 단어들의 시소리스를 구성하여 연관단어를 알아내는 방법[2,3,4,5], 키 팩트라는 문서의 주된 내용을 대표하는 (중심어, 종속어) 쌍을 추출하고 이 중 가장 빈도수가 많은 키팩트들에 가중치를 두어 이용자의 질의와 검색 대상의 유사도를 비교하는 방법 등이 연구되고 있다[6,7]. 또한 의미망 구축후 의미망을 통해 용어간의 연관성 측정, 개념적으로 관련된 용어를 추론, 검색에 반영하는 방법도 있으며[8], 검색어의 의미를 이용한 검색을 수행하기 위해 질의어에 대한 형태소 분석을

\* 준희원 : 제주대학교 대학원 컴퓨터공학과 박사과정  
mincando@cheju.ac.kr(제1저자)

\*\* 종신희원 : 제주대학교 컴퓨터공학과 부교수  
sjlee@cheju.ac.kr(공동저자)

통해 의미어를 추출하고, 의미어로 재구성된 템플릿을 사용하여 의미기반 검색을 하기도 한다[7,9]. 또한 최근에는 XML 및 XML 기반의 메타정보 표현 기술인 RDF를 이용한 시맨틱웹(Semantic Web)에 관한 연구도 이루어지고 있다[10,11,12,13,14].

이러한 수많은 연구에도 불구하고 연구 논문 검색과 같이 내용검색이 아닌 논문 요약 정보 및 문서 검색이 주된 경우 대부분 키워드와 제목, Abstract 내의 문자열과 질의어와의 유사성만을 이용한 정보 검색이 이루어지고 있어서 검색자가 원하는 정확한 범주의 결과만을 제시하지 못하고 있다. 이들의 경우 대부분 이용자가 원하던 정확한 정보를 검색하는 것이 아니라 이용자가 원하는 정보를 포함하고 있는 방대한 정보를 제시할 뿐이며, 이용자로 하여금 제시된 결과 내에서 또다시 검색을 수행하게 하는 이중 작업을 요구한다. 또한 키워드 생성에 있어서도 문서 생성자가 정의하고 있어서 정확하고, 획일적인 표현을 사용하지 않는 경우 검색의 정확성을 더욱 떨어뜨리게 된다.

본 논문에서는 이러한 검색 결과의 정확성을 개선하기 위하여 연구 논문들의 제목 내에서 나타나는 키워드들의 역할정보를 가지고 시맨틱을 구성하여 시맨틱 기반의 검색 방법을 제안하고 있다. 이를 위해 데이터 저장시 제목 내에서 키워드와 역할어를 추출하고 이를 XML기반의 메타정보인 RDF로 구성하고, 이를 이용하여 사용자 의도가 반영될 수 있는 검색을 수행하게 하였다.

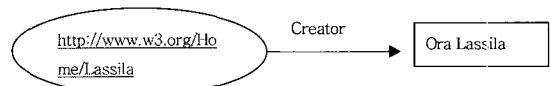
## 2. RDF Semantics

XML의 응용 애플리케이션중 시맨틱 웹(Semantic Web)의 구문적 기반을 제공하는 것이 RDF(Resource Description Language)라는 추상적 데이터 모델이다[13,14,15]. RDF 모델을 구체화하는 표현 방법으로는 Graph형식, N-triples(N3) 그리고 RDF/XML 등이 있다[14,15]. 이들 각각의 표현방법을 다루는 여러 RDF API들이 개발되고 있으며 본 논문에서는 Jena RDF API를 이용하여 RDF 쿼리를 수행

하고 있다[16,17,18].

RDF는 차세대 웹 모델인 시맨틱 웹의 구문적 기반을 제공한다. 즉 시맨틱 웹에서의 시맨틱을 구성하는 요소를 정의한다고 할 수 있다. XML은 HTML이 가질 수 없었던 구조정보를 갖는 게 장점이라면, RDF는 XML이 가질 수 없었던 Semantic 을 제공한다. 이러한 시맨틱간의 관계구조를 정의한 Ontology의 등장은 시맨틱의 교환 및 결합 개념까지도 지원하게 되었다[19].

아래 그림은 “Ora Lassila is the creator of the resource <http://www.w3.org/Home/Lassila>”라는 문장을 RDF Graph로 표현한 예이다.



(그림 1) RDF 그래프의 예

그림 1에서 동그라미는 URI가 대표하는 리소스를 나타내며, “Creator”는 리소스에 대한 프라퍼티(Property)의 이름을 나타내고, 사각형은 리소스에 대한 프라퍼티 값으로서 문자열 값을 의미한다. 이 그래프 표현은 RDF 데이터의 표현에는 적합하지만 실제 RDF 데이터 전송을 위해서는 직렬화 가능한 XML표현이 적합하다. 이 RDF 그래프를 RDF/XML으로 표현하면 다음과 같이 표현할 수 있다.

```

<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:s="http://description.org/schema/">
    <rdf:Description about="http://www.w3.org/Home/lassila">
        <s:Creator>Ora Lassila</s:Creator>
    </rdf:Description>
</rdf:RDF>

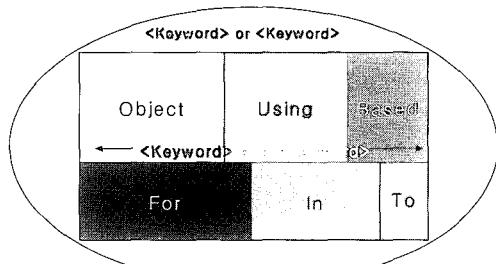
```

위의 RDF/XML에서 “about”은 리소스 그 자체를 의미하고 <s:Creator>라는 태그는 “s”라는 prefix가 나타내는 네임스페이스에 속하는 “Creator”라는 Property를 나타내고 있는 것이다. 우리는 이

Creator라는 용어의 의미를 이해한다면 “Ora Lassila”가 about이 나타내는 리소스의 creator라는 것을 알 수 있다. 즉, “Creator”라는 XML 태그는 처리 기에게 Semantic 정보가 되는 것이다. 최근의 Web Ontology라는 분야는 처리기에게 더욱 정교하면서도 구체화된 개념 표현방법을 제공하고 있다. 단지 용어의 정의만이 아닌 용어와 용어의 관계를 표현함으로써 처리기로 하여금 개념 관계에 기반을 둔 존재론적인 semantic을 이해하도록 도움을 주고 있다. 이러한 Web Ontology 언어로 RDFS, DAML+OIL, OWL등이 개발되었다[19].

본 논문에서는 이러한 개념관계정보는 추후 응용의 확장 시에 제공하기로 하고, 단순한 시맨틱을 제공함으로써 단순 키워드 검색시보다 검색정확도는 검색 범주를 줄이는데 그 목적을 두고 있다. 그림 2는 이러한 시멘틱을 이용하였을 경우 검색 결과를 효율적으로 줄일 수 있음을 보이고 있다.

그림 2에서 keyword 기반의 OR 검색은 전체 원에 해당하는 검색 결과를 보이고, And 검색은 큰 사각형 영역과 같은 검색결과의 크기를 갖는다. 하지만 본 논문에서 제안하는 RDF 메타정보를 이용한 검색은 그 안에 포함되어 있는 색상을 갖는 사각형처럼 부분집합을 형성한다는 것을 예측 할 수 있다. 즉 단순 키워드 기반의 검색보다는 Semantic을 이용한 검색 결과는 그 부분집합일 수밖에 없으며 최적의 경우 가장 검색 자가 의도하는 정확한 결과를 제시할 수 있게 된다. 그리고 최악의 경우도 기존 키워드 기반 검색과 동일한 경우일 뿐이라는 것을 알 수 있다.



(그림 2) 키워드기반 검색과 RDF 검색의 영역차이

### 3. 제안 검색 시스템의 설계

본 장에서는 논문 제목 내에서의 역할어 추출 및 검색 방법, 그리고 제안하는 검색시스템의 구조에 대해 설명한다.

#### 3.1 키워드 역할 정보 추출

본 논문이 검색 대상으로 하는 연구 논문들은 논문 제목에 있어서 일정한 패턴이 나타난다. 그 패턴을 유심히 살펴보면 키워드들은 일정한 역할을 맡고 있다는 것을 분석할 수 있다. 어떤 키워드들은 제목 내에서 목적으로 사용되고, 연구의 결과로 표현되기도 하며, 때로는 연구 결과를 이끌어 내기 위해 사용된 기술 및 배경 이론을 의미하기도 한다. 이를 키워드들의 역할 패턴을 정리하면 표 1과 같다.

이러한 가정 하에 연구 논문의 제목을 대상으로 각각의 키워드의 역할을 의미하는 어휘들을 추출하여 ‘역할어’라 명명한다면, 표 2와 같이 키워드별 역할어들을 나열할 수 있다.

(표 1) 연구 논문 제목 내에서의 키워드 역할 패턴

역할별 키워드 종류	역할 의미
In 키워드	연구의 대상, 범위를 나타냄
Based 키워드	기반 이론 및 기술을 나타냄
Using 키워드	사용한 도구 및 방법론을 나타냄
For 키워드	연구의 목적 등을 나타냄
To 키워드	연구 방향을 나타냄
Object 키워드	연구 결과물을 나타냄

(표 2) 키워드 역할과 역할어들

키워드 역할	역할어
In	에서 /
Based	기반 / 갖는 / 기초한 / 기반한 / 포함한
Using	이용한 / 통한 / 고려한
For	위한 / 지원하는 / 용 / 관한 / 지향하는
To	로 /
Object	역할어 정의 필요 없음

(표 3) 키워드 역할 기호와 변형된 제목

원 제목	키워드 역할 기호를 갖는 변형된 제목
XML문서에서의 엘리먼트 정보를 이용한 스키마 추출방법	XML 문서 -I 엘리먼트 정보 -U 스키마 추출방법 -O
XML 데이터의 RDBMS 로의 분할 저장시중복을 활용한 효율적 질의 처리	XML 데이터 RDBMS -T 분할 저장시중복 -U 효율적 질 처리 -O
XML기반 상품 카탈로그의 설계 및 적용	XML -B 상품 카탈로그 설계 및 적용 -O
객체모델을 이용한 XML DTD 의 RDB 스키마로의 변환	객체모델 -U XML DTD RDB 스키마 -T 변환 -O
EDMS를 위한 XML기반 목표문서 인식기법	EDMS -F XML -B 목표문서 인식기법 -O
객체 모델을 기반으로 한, XML 문서의 관계형 데이터베이스로의 저장을 위한 EJB 컴포넌트 개발	객체 모델 -B 한, XML 문서 관계형 데이터베이스 -T 저장 -F EJB 컴포넌트 개발 -O

```

<rdf:RDF xmlns:rdf
           = "http://www.w3.org/1999/22-rdf-syntax-ns#"
           xmlns:mincando='http://mincando.pe.kr/paper
                           /elements#'>
<rdf:Description>
  <mincando:object> object역할 키워드
  </mincando:object>
  <mincando:based> based 역할 키워드
  </mincando:based>
  <mincando:using>using역할 키워드
  </mincando:using>
  <mincando:for> for 역할 키워드 </mincando:for>
  <mincando:in> in 역할 키워드 </mincando:in>
  <mincando:to> to 역할 키워드</mincando:to>
</rdf:Description>
</rdf:RDF>

```

(그림 3) 연구 논문 제목에 대한 RDF 구조

### 3.2 키워드 역할에 대한 RDF 메타정보 생성

표 3과 같이 추출된 역할어를 이용하여 논문제목에 대한 메타정보인 키워드 역할정보를 나타내는 RDF문서를 생성한다. 표 2에서 추출된 역할어를 중심으로 논문제목을 분할하고, 각각의 분할된 키워드들을 역할어를 의미하는 그림 3의 6가지의 XML 엘리먼트를 이용하여 표현하면 표 4와 같이 각

(표 4) RDF로 표현된 연구 논문 제목들

원 제목	RDF형태로 표현된 제목
XML 문서에서의 엘리먼트 정보를 이용한 스키마 추출방법	<rdf:RDF xmlns:rdf='http://www.w3.org/1999/22-rdf-syntax-ns#' xmlns:mincando='http://mincando.pe.kr/paper/elements#'> <rdf:Description> <mincando:object>스키마 추출방법</mincando:object> <mincando:using>엘리먼트 정보</mincando:using> <mincando:in>XML 문서</mincando:in> </rdf:Description> </rdf:RDF>
XML기반 상품 카탈로그의 설계 및 적용	<rdf:RDF xmlns:rdf='http://www.w3.org/1999/22-rdf-syntax-ns#' xmlns:mincando='http://mincando.pe.kr/paper/elements#'> <rdf:Description> <mincando:object>상품 카탈로그 설계 및 적용</mincando:object> <mincando:based>XML</mincando:based> </rdf:Description> </rdf:RDF>
EDMS를 위한 XML기반 목표문서 인식기법	<rdf:RDF xmlns:rdf='http://www.w3.org/1999/22-rdf-syntax-ns#' xmlns:mincando='http://mincando.pe.kr/paper/elements#'> <rdf:Description> <mincando:object>목표문서 인식기법</mincando:object> <mincando:based>XML</mincando:based> <mincando:for>EDMS</mincando:for> </rdf:Description> </rdf:RDF>

제목들에 대한 RDF 메타정보를 생성할 수 있다.

### 3.3 RDF/XML 메타정보를 이용한 논문 검색

RDF/XML 데이터에 대한 접근 및 변환, 쿼리를

```

import com.hp.hpl.mesa.rdf.jena.model.*;
import com.hp.hpl.mesa.rdf.jena.mem.ModelMem;
import com.hp.hpl.mesa.rdf.jena.vocabulary.*;
import com.hp.hpl.mesa.rdf.jena.common.SelectorImpl;
public class QueryRDF
{
    .... 생략 ...
    public static void myQuery(String q, Model model){
        String result=null;
        try{
            Property
            DC=model.createProperty("http://mincando.pe.kr/paper/
elements#",q);
            StmtIterator iter=model.listStatements(new
            SelectorImpl(null,DC,(RDFNode) null));
            if(iter.hasNext()){
                while(iter.hasNext()){
                    result=
iter.next().getSubject().getProperty(DC).getString();
                    if(result.indexOf("객체") > -1)
System.out.println(result);
                }
            } else {
                result="No result";
            }
        }catch(RDFException e){ e.printStackTrace();}
    }
    ....
}

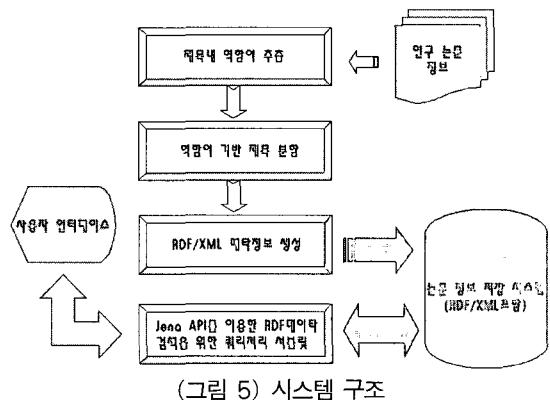
```

(그림 4) Jena API를 이용해 구현한 Java 서블릿 프로그램

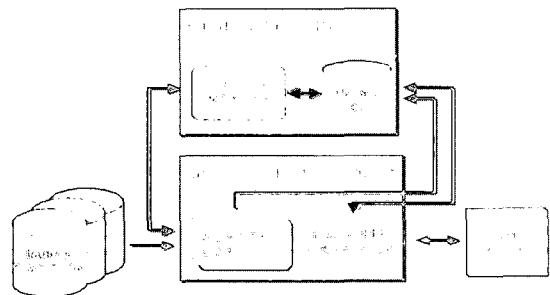
수행하는 여러 API들이 개발되고 있다. 이중 Jena RDF API는 RDF 그래프 모델을 생성, 조작하기 위한 Java 프로그래밍언어를 위한 API이다. Jena는 다른 구현들보다 쉬운 API를 제공하며 RDF 스펙에 적합한 구현 모델을 제시하고 있다. 그림 3은 Jena API를 이용한 RDF문서에 대한 쿼리를 수행하는 Java 소스의 일부분이다.

### 3.3 시스템의 구조

제안하는 시스템은 그림 5와 같이 구성된다. 이 시스템은 3가지 주요 구성요소로 RDF/XML



(그림 5) 시스템 구조



(그림 6) 구현 환경

기반 논문정보 저장 시스템, 논문정보 중 제목 내 키워드 역할 분석 및 RDF/XML메타정보 생성 모듈, 사용자 질의에 대한 Jena API기반의 쿼리 처리기를 갖는다.

## 4. 구현 및 결과

이 장에서는 앞에서 제시한 시스템을 구현하고 테스트한 결과를 기술한다.

### 4.1 구현 환경

본 논문은 Windows2000 운영체제상에서 MS-SQL DB를 논문정보 저장 시스템으로 사용하였고, 컴퓨터 연구정보센터(<http://cseric.cau.ac.kr>)에서 키워드 기반 검색을 이용해 각 xml관련 논문 546개, 에이전트 관련 논문 520개, 무선관련 논문 539개, 인식 관련 논문 805개, 보안 관련 논문 666개의

## 시맨틱을 이용한 연구 논문 검색 시스템

정보를 웹상에서 수집후, ASP시스템을 통해 역할 정보 추출후 RDF/XML 형식으로 DB에 저장도록 하였다. 검색을 위해서는 Jakarta-tomcat 기반의 Java Servlet을 이용해 사용자 쿼리를 수행도록 하였다.

### 4.2 역할정보 추출 결과

증복된 논문 정보를 제외한 총 2705개의 한글 논문들에 대해 패턴 분석을 수행한 후 각각에 대해서 RDF/XML 형태의 메타정보를 생성하였으며 아래 그림 7은 변환 결과 화면을 보여준다.

Mincando Paper Collection - 원문과 사전하고 비교하는 원문 검색	
제작자: 민간인 보기   보고서   증명서   저작권   도록방법   검색하기   대화방법   도록방법   목록보기   상세보기   상세검색   콘텐츠   인터넷   인터넷   주소록   http://mincando.pe.kr/tomino/openQuery.html	
제작자: 민간인	제작자: 민간인
제작자: 민간인	제작자: 민간인
제작자: 민간인	제작자: 민간인

(그림 7) 논문 제목에 대한 제목 분할 및 생성된 RDF/XML

2705개의 논문 제목을 대상으로 역할이 추출을 수행하였을 때 각 제목 내에서 역할어들이 검출된 횟수는 표 6과 같았다.

(표 6) 각 제목 내에서의 역할어 검출 횟수

object	based	for	using	in	to
2705	674	781	883	63	500

### 3.3 검색 결과

다음 그림 5는 사용자로부터 검색조건을 입력 받기 위한 사용자 인터페이스 화면이다.

다음 그림 9는 검색결과인 XML문서를 보여준다.

그림 10은 그림 5에 XSL을 적용하여 HTML로 표현한 결과이다.

The screenshot shows a search form titled "Query for Paper searching" with the instruction "within relation infomations between keywords in title". It contains several input fields for search terms and their relationships:

- object(what): [Title]
- from (기초하여, 포함하여): [based]
- to (위해(지원하는)): [for]
- using (이용하여): [using]
- in (에서(~상에서)): [in]
- to (으로의~): [to]

At the bottom are "Query Execute" and "reset" buttons.

(그림 8) 사용자 검색 인터페이스 화면

The screenshot displays XML search results with various titles and IDs:

- <title no=id="24">XML Schema의 이용한 병원정보교환시스템 HIES의 역할</title>
- <title no=id="48">XML과 SMIL을 이용한 전자 출판 시스템 설계</title>
- <title no=id="57">XML을 이용한 간접현장 EDI 시스템 구현</title>
- <title no=id="76">XML을 이용한 편집기와 학습지도인 활용 시스템 설계 및 구현</title>
- <title no=id="77">WBI를 이용한 XML소스간 이용한 교과 책인 시스템</title>
- <title no=id="91">WAP과 XML 기술을 이용한 전자상거래 시스템 구현에 관한 연구</title>
- <title no=id="98">WBI에서 XML을 이용한 교과 책인 시스템</title>
- <title no=id="100">XML을 이용한 건설현장 EDI 시스템 구현</title>
- <title no=id="105">MP3 태그의 XML 확장은 이용한 동화화된 재생 시스템</title>
- <title no=id="120">XML을 이용한 웹 기반 RMON 에이전트 시스템 설계 및 구현</title>
- <title no=id="146">XML을 이용한 문서운행 시스템 설계</title>
- <title no=id="216">인터넷 기반의 XML을 이용한 가상교育 시스템의 개발에 관한 연구</title>
- <title no=id="229">관계형 데이터베이스 시스템에서 XML-뷰를 통한 XML 데이터의 지원</title>
- <title no=id="258">워크플로우 시스템에서 XML을 이용한 실태 서리</title>
- <title no=id="297">XML을 이용한 실시간 입찰시스템의 설계 및 구현</title>
- <title no=id="301">XML 전자서명을 이용한 XML/EDI시스템의 설계 및 구현</title>
- <title no=id="329">XML을 이용한 내용기반 이미지 데이터베이스의 설계 및 검색 시스템 구현</title>
- <title no=id="344">XML을 이용한 중앙 의료정보시스템 구현</title>
- <title no=id="357">XML을 이용한 CNP&P방의 종합 검색시스템 설계</title>

(그림 9) 검색 결과화면(XML 문서)

The screenshot shows the same XML results as in Figure 9, but they are now displayed as an HTML table with columns for ID, Title, and Description.

ID	제목	내용
id : 24	[제목] XML Schema의 이용한 병원정보교환시스템 HIES의 역할	[내용]
id : 48	[제목] XML과 SMIL을 이용한 전자 출판 시스템 설계	[내용]
id : 57	[제목] XML을 이용한 간접현장 EDI 시스템 구현	[내용]
id : 76	[제목] XML을 이용한 편집기와 학습지도인 활용 시스템 설계 및 구현	[내용]
id : 77	[제목] WAP과 XML 기술을 이용한 교과 책인 시스템	[내용]
id : 91	[제목] WAP과 XML 기술을 이용한 전자상거래 시스템 구현에 관한 연구	[내용]
id : 98	[제목] WBI에서 XML을 이용한 교과 책인 시스템	[내용]
id : 100	[제목] XML을 이용한 건설현장 EDI 시스템 구현	[내용]

(그림 10) 검색 결과화면 XSL처리 후

(표 7) 키워드 기반 검색과 메타정보 검색 결과수의 비교

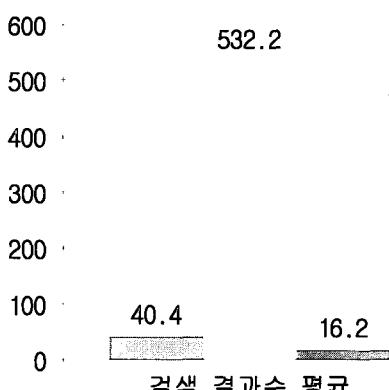
검색어	검색방법	검색 질의	검색수
시스템+xml	키워드 기반	시스템 <i>and</i> xml	127
		시스템 <i>or</i> xml	975
	메타정보 이용	시스템 <i>using</i> xml	26
		시스템 <i>for</i> xml	8
에이전트+보안	키워드 기반	에이전트 <i>and</i> 보안	19
		에이전트 <i>or</i> 보안	520
	메타정보 이용	에이전트 <i>for</i> 보안	2
		에이전트 <i>object</i> 보안	13
무선+변환	키워드 기반	무선 <i>and</i> 변환	11
		무선 <i>or</i> 변환	335
	메타정보 이용	무선 <i>object</i> 변환	10
		무선 <i>for</i> 변환	1
문서+인식	키워드 기반	문서 <i>and</i> 인식	6
		문서 <i>or</i> 인식	370
	메타정보 이용	문서 <i>for</i> 인식	1
		문서 <i>object</i> 인식	5
xml+검색	키워드 기반	xml <i>and</i> 검색	39
		xml <i>or</i> 검색	461
	메타정보 이용	xml <i>for</i> 검색	15
		xml <i>using</i> 검색	0

(표 8) 각 검색별 평균 검색결과의 평균 수

키워드 관계 기반 검색		RDF/XML 기반 검색의 결과수 평균
키워드간 <i>and</i> 조건	키워드간 <i>or</i> 조건	532.20 개

40.40 개	16.20 개
---------	---------



(그림 11) 키워드 기반 검색과 RDF/XML 메타정보 기반 검색법주의 비교

표 7과 표 8에서는 RDF/XML 메타정보를 이용한 검색과 키워드 기반 검색의 효율성 차이를 보이고 있다. 즉, RDF/XML 메타정보를 이용할 경우 검색 범주를 검색자가 의도하는 범위 내로 좁히기 때문에 검색의 정확도가 높다고 할 수 있다.

## 5. 결 론

결론적으로 키워드기반의 정보검색 시스템은 그 결과가 방대하여 사용자로 하여금 추가적인 검색을 수행하게 하는 단점을 안고 있으며 사용자 의도를 정확히 반영해 주지 못하는 설정이다. 여기에 논문 제목 속에 숨겨진 역할정보를 추출하여 이를 RDF/XML 메타정보로 구성하고 검색 대상으로 활용한다면 사용자가 의도하는 정확한 대상 논문을 검색할 수 있을 뿐 아니라 그 결과 집합의 수도 최적의 수로 줄어든다는 것을 알 수 있었다. 향후에는 공학 연구논문이외에 기타 논문들까지도 확장하기 위한 개념 및 역할이 연구가 필요하고 6가지의 패턴 정보뿐만이 아닌 다양한 패턴정보도 연구해볼 필요가 있으며, 또한 온톨로지를 이용한 외부 정보자원과의 연결 서비스까지 제공할 필요가 있다고 여겨지며, 이러한 제안은 디지털 도서관 및 웹 자원 검색에도 활용할 가치가 있을 것으로 여겨진다.

## 참 고 문 헌

- [1] 이창범, 김민수, 백장선, 이귀상, 박혁로. “주성분분석을 이용한 주제어 기반 문서 자동요약”, KDBC(SIGDB-KISS), 2001 1호, pp. 191~195, 2001. 06.
- [2] 김영천, 이재훈, 문유미, 박병권, 이성주. “정보검색에서 용어 가중치 재부여를 이용한 성능증진에 관한 연구”, 퍼지 및 지능시스템학회논문지, 11권 9호, pp. 811~816, 2001. 12.
- [3] 김주연, 김병만. “용어 발생 유사도와 퍼지추론을 이용한 질의 용어 확장 및 가중치

- [재산정”, 한국정보과학회 논문지 B, 27권 09호, pp. 961~972, 2000. 09.]
- [4] 우선미, 유춘식, 김용성. “용어 연관성 분석을 이용한 사용자 위주의 문서순위결정 기법”, 한국정보과학회 논문지 B, 28권 2호, pp. 149~156, 2001. 02.
- [5] 양승원, 노희영, “시소리스를 이용한 XML 태그 검색 시스템”, 정보과학회 2000년 추계학술대회, 27권 2호, pp. 145~147, 2000.
- [6] 김수희, 남효돈. “정보검색에서 정확률의 향상을 위해 키워드의 가중치 부여”, 한국정보과학회 논문지 D, 27권 4호, pp. 627~636, 2000. 12.
- [7] 장명길, 김현진, 장문수, 최재훈, 오효정, 이충희, 허정. “의미기반 정보검색”, 한국정보과학회지, 19권 10호, pp. 7~18, 2001. 10.
- [8] 노영희. “개념기반 검색 방법론”, 2001 국제컨퍼런스 IT21-정보처리학회, pp. 79~110, 2001. 6.
- [9] 박현규, 오종훈, 김명호, 최기선, 이광형. “퍼지 추론에 의한 자연언어 정보 검색”, 한국정보처리학회 논문지 B, 11권 9호, pp. 811~816, 2001. 12.
- [10] 김노환, 정충교, “XML DOM을 이용한 웹문서 검색 알고리즘”, 컴퓨터산업교육기술학회논문지, 2권 6호, pp. 775~782, 2001.
- [11] XML : eXtensible markup Language, <http://www.w3.org/XML/>.
- [12] Semantic Web, <http://www.w3.org/2001/sw/>.
- [13] <http://www.w3.org/TR/rdf-concepts/>.
- [14] <http://www.w3.org/TR/rdf-mt/>.
- [15] <http://www-uk.hpl.hp.com/people/bwm/papers/2001221-paper/>.
- [16] <http://www.hpl.hp.com/semweb/>.
- [17] [http://www.hpl.hp.com/semweb/doc/tutorial/RDF\\_API/index.html](http://www.hpl.hp.com/semweb/doc/tutorial/RDF_API/index.html).
- [18] <http://triple.semanticweb.org/>.
- [19] <http://infomesh.net/2001/swintro/#ontInferer.ce>.

## ● 저 자 소 개 ●

### 김 영 민

1993년 제주대학교 정보공학과 졸업(학사)  
1997년 제주대학교 대학원 정보공학과 졸업(석사)  
2000년~현재 : 제주대학교 대학원 컴퓨터공학과 박사과정  
관심분야 : XML, Agent, Semantic Web, Internet Computing etc.  
E-mail : mincando@cheju.ac.kr

### 이 상 준

1984년 중앙대학교 컴퓨터공학과 졸업(학사)  
1989년 중앙대학교 대학원 컴퓨터공학과 졸업(석사)  
1992년 중앙대학교 대학원 컴퓨터공학과 졸업(박사)  
1993년~현재 : 제주대학교 컴퓨터공학과 부교수  
관심분야 : Heuristic Algorithm, Web Database, Agent, XML 응용 etc.  
E-mail : sjlee@cheju.ac.kr