

# 웹 문서의 메타데이터 관리를 위한 XMP 및 온톨로지 기반의 시맨틱 어노테이션 지원도구 개발

양경모<sup>1\*</sup>, 황석형<sup>2</sup>, 최성희<sup>2</sup>  
<sup>1</sup>선문대학교 일반대학원 전자계산학과  
<sup>2</sup>선문대학교 컴퓨터공학부

## On Developing a Semantic Annotation Tool for Managing Metadata of Web Documents based on XMP and Ontology

Kyoung-Mo Yang<sup>1\*</sup>, Suk-Hyung Hwang<sup>2</sup> and Sung-Hee Choi<sup>2</sup>

<sup>1</sup>Department of Computer Science, graduate school, Sun Moon University

<sup>2</sup>Department of Computer Science & Engineering, Sun Moon University

**요약** 시맨틱 웹에서는 기계가 처리할 수 있는 의미 정보를 토대로 보다 효율적이고 효과적인 시맨틱 검색과 웹 서비스 제공을 목적으로 한다. 따라서, 다양한 웹 콘텐츠들에 대하여 컴퓨터가 이해 가능한 형식으로 메타데이터를 생성하고 추가하는 과정, 즉, 시맨틱 어노테이션이 시맨틱 웹의 중요한 기반기술 중의 하나이다. 최근에는 어노테이션 정보를 관리하기 위해서, 대상문서 내부에 메타데이터를 직접 내장시키는 기법이 주로 사용되고 있다. 그러나, 웹 문서의 시맨틱 어노테이션과 관련하여 기존의 지원도구들은, 주로 HTML 문서를 대상으로 하고 있고, 대부분의 도구들에서는 메타데이터를 활용한 시맨틱 검색기능을 제공하지 않고 있다.

본 연구에서는, 이와 같은 문제점들과 기존의 관련 연구결과들을 토대로, 웹 문서(HTML, PDF)들에 대한 시맨틱 어노테이션을 보다 효율적으로 지원하기 위한 온톨로지 기반의 시맨틱 어노테이션 지원도구(OSA)를 개발하였다. OSA에서는, RDFS(RDF Schema)로 시맨틱 어노테이션 모델을 정의하고, 이를 토대로 온톨로지 기반의 의미정보들을 표현하며, XMP(eXtensible Metadata Platform)표준에 맞추어서 해당 디지털 문서 내에 시맨틱 어노테이션 정보를 내장시킨다. 본 연구에서 개발한 어노테이션 도구를 활용함으로써, 웹 문서에 대한 효율적인 시맨틱 어노테이션이 가능하며, XMP를 기반으로 웹 문서 자체와 시맨틱 어노테이션 정보를 일체화시킴으로써 어노테이션 정보관리에 정합성을 유지할 수 있으므로, 향후 다양한 웹 콘텐츠에 대한 시맨틱 검색에 효과적으로 활용될 수 있다.

**Abstract** The goal of Semantic Web is to provide efficient and effective semantic search and web services based on the machine-processable semantic information of web resources. Therefore, the process of creating and adding computer-understandable metadata for a variety of web contents, namely, semantic annotation is one of the fundamental technologies for the semantic web. Recently, in order to manage annotation metadata, direct approach for embedding metadata into the document is mainly used in semantic annotation. However, many semantic annotation tools for web documents have been mainly worked with HTML documents, and most of these tools do not support semantic search functionalities using the metadata.

In this paper, based on these problems and previous works, we propose the Ontology-based Semantic Annotation tool(OSA) to efficiently support semantic annotation for web documents(such as HTML, PDF). We define a semantic annotation model that represents ontological-semantic information by using RDFS(RDF Schema). Based on XMP(eXtensible Metadata Platform) standard, the model is encoded directly into the document. By using OSA with XMP, user can perform semantic annotation on web documents which are able to keep compatibility for managing annotation metadata. Eventually, the integrated semantic annotation metadata can be used effectively in semantic search for a variety of web contents.

**Key Words** : Semantic Annotation, Ontology, EXtensible Metadata Platform

\*교신저자 : 양경모(yjj0309@gmail.com)

접수일 09년 04월 10일

수정일 (1차 09년 06월 25일, 2차 09년 07월 20일)

계재확정일 09년 07월 22일

## 1. 서론

웹 기술의 발전으로 인해 수많은 기관과 커뮤니티, 개인들에 의해서 생성되는 문서와 정보의 양 또한 급격히 증가하고 있다. 이로 인해 사용자들이 자신의 목적에 적합한 문서와 정보를 발견하는 것은 점점 어려워지고 있다. 이에 효과적으로 정보를 검색하기 위한 다양한 검색 기술들이 연구되고 있으나, 현재의 웹은 사람이 읽어보고 잘 이해할 수 있도록 하기 위한 디스플레이기술에 초점을 맞추고 있다. 따라서, 사람이 아닌 프로그램 또는 소프트웨어 에이전트가 문서로부터 의미를 자동 추출하여 정보 검색 및 정보의 통합과 같은 작업에 활용하는데 문제점이 발생하고 있다. 이러한 문제점을 해결하기 위하여, 팀 버너스리는 웹에 존재하는 자원(Resource)에 잘 정의된 의미를 부여하고 이를 사람과 컴퓨터가 이해하고 처리할 수 있도록 하는 시맨틱 웹(Semantic Web)을 제안하였다[1].

시맨틱 웹에서 웹 자원에 의미를 부여하고 컴퓨터의 자동화된 처리를 가능하게 하기 위해서는 메타데이터(Metadata)기술과 시맨틱 어노테이션(Semantic Annotation)기술이 필수적이다. 메타데이터는 데이터에 대한 데이터 또는 정보에 대한 정보를 의미하며, 시맨틱 웹에서 메타데이터를 표현하기 위한 기술들로는 자원을 식별하기 위한 URI(Uniform Resource Identifier), XML(Extensible Markup Language)의 표현 구문을 기반으로 하는 RDF(Resource Description Framework)[2], RDFS(RDF Schema)[3], 그리고 웹 온톨로지 언어인 OWL(Web Ontology Language)[4]이 있다.

일반적으로 어노테이션(Annotation)은 문서나 문서의 특정 부분에 개인의 의견, 해석, 강조, 비판 등의 추가정보를 기입하는 과정을 의미하며, 기입된 추가정보들은 사람이 이해할 수 있는 하이라이트, 밑줄, 도형, 노트 등과 같은 형태로 표현된다[5]. 한편, RDF, RDFS, OWL 등과 같은 메타데이터표현 언어를 사용하여 컴퓨터가 처리할 수 있는 형태로 메타데이터를 기술하고, 웹 콘텐츠에 메타데이터를 부여하는 과정을 시맨틱 어노테이션[6,7]이라고 하며, 시맨틱 웹의 활성화를 위하여 다양한 연구들이 진행되고 있다[8-14].

기존 연구들에서 사용된 시맨틱 어노테이션 정보의 부여 방식으로는, 어노테이션 정보를 문서에 직접 내장시키는 방식과 어노테이션 정보를 별도의 저장소에 저장, 관리하는 방법이 사용되고 있다. 특히, 메타데이터(또는, 어노테이션 정보)를 직접 파일에 내장시켜 메타데이터를 관리하기 위한 기술로는, 어도비(Adobe)에서 제안된 메타데이터 관리 플랫폼인 XMP(eXtensible Metadata

Platform)[15,16]가 주로 사용되고 있다.

시맨틱 어노테이션과 관련된 기존 대부분의 연구들에서는, 웹의 기본 문서형식인 HTML문서에 대한 시맨틱 어노테이션에 초점을 두고 있다. 그러나, 수많은 문서들이 다양한 문서형식(PDF, OpenOffice 등)으로 작성되어 웹에 배포되고 있다. 특히, 디지털 문서규격의 국제 표준인 PDF(Portable Document Format)[17]는 서적이나 신문, 잡지, 학술자료(논문) 등을 디지털 문서로 배포하기 위해 사용되는 가장 일반적인 형식이다. 따라서, HTML문서만을 처리대상으로 하는 기존 도구를 사용하여 시맨틱 어노테이션을 수행하는 것은 한계가 있다.

한편, 시맨틱 어노테이션 정보들은 정보관리 또는 정보의 검색측면에서 매우 중요한 역할을 수행한다. 그러나, 시맨틱 어노테이션 정보를 활용하지 않는 기존의 검색방법(키워드기반검색)을 사용하여, 시맨틱 어노테이션 정보가 포함된 문서들을 검색하는 것은 매우 비효율적이다[6,18]. 따라서, 시맨틱 어노테이션 정보를 활용한 시맨틱 검색을 제공함으로써, 사용자들은 자신들이 어노테이션한 정보를 기반으로 효과적인 문서검색을 할 수 있다.

본 연구에서는, 현재 인터넷에 급증하고 있는 디지털 문서들과 이에 포함된 유용한 정보를 효과적으로 관리하기 위한 디지털 문서(HTML, PDF)의 시맨틱 어노테이션을 지원하는 도구(OSA)를 개발하였다. 기존 대부분의 지원도구들[8-12]이 HTML문서만을 처리하는 반면, 본 연구에서는 HTML문서뿐만 아니라 디지털 문서의 국제표준 규격인 PDF문서도 처리대상으로 한다. 또한, RDFS로 기술된 어노테이션 모델과 메타데이터 관리 플랫폼인 XMP를 사용하여, 보다 풍부한 어노테이션 정보를 표현하고, 문서에 내장시킬 수 있게 되어 시맨틱 검색에 효과적으로 활용할 수 있다.

본 논문의 구성은 다음과 같다. 제2장에서는 관련연구로 기존 연구들의 제반 특징들에 대해서 정리하여 설명하고, 제3장에서는 본 연구에서 개발한 시맨틱 어노테이션 지원도구(OSA)를 소개한다. 제4장에서는 OSA에서 제공하는 XMP기반의 시맨틱 어노테이션 기법과 시맨틱 검색에 대해서 설명한다. 제5장에서는 기존 도구들과 본 연구결과(OSA)의 기능상의 특징들을 비교한 결과에 대해서 설명한다. 마지막으로 제6장에서는 결론과 향후 연구과제에 대해 설명한다.

## 2. 시맨틱 어노테이션 관련 연구

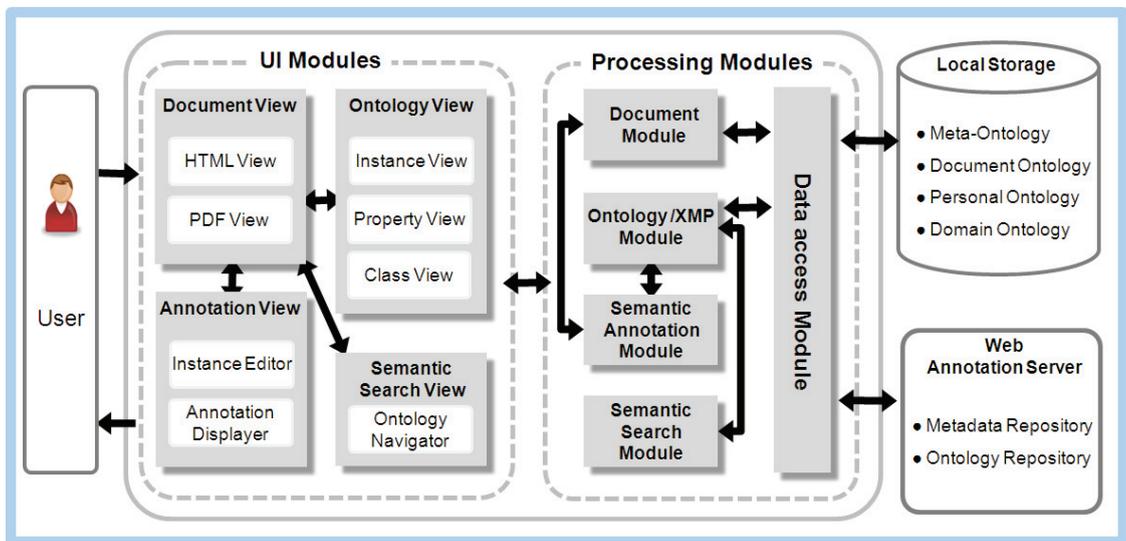
시맨틱 어노테이션을 지원하기 위한 도구들로서, Annozilla[8], MnM[9], OntoMat-Annotizer[10],

SHOE[11], KIM[12], SALT[13], PDFTab[14] 등이 개발되었다. 각 도구들의 제반 특징은 다음과 같다.

- Annozilla[8]은 W3C Annotate[20]프로젝트에서 정의한 어노테이션 프로토콜을 도입하여 개발된 웹기반의 시맨틱 어노테이션 지원도구이다. Annozilla에서 생성되는 어노테이션 정보들은 메타데이터 서버에 RDF형식으로 저장되고 공유된다. 또한, 어노테이션이 수행된 문서내부에서 어노테이션 정보의 위치를 파악하기 위하여 XPointer를 사용한다.
- MnM[9]은 HTML문서에 대한 시맨틱 어노테이션을 지원하기 위하여 자동과 반자동 어노테이션 기능을 모두 제공하고 있으며, 온톨로지 브라우저와 웹 브라우저가 통합된 사용자 인터페이스 환경을 제공한다.
- OntoMat-Annotizer[10]는 N3와 OWL파일을 지원하는 온톨로지 기반의 시맨틱 어노테이션 지원도구로 클래스와 속성, 그리고 인스턴스의 다양한 편집이 가능한 온톨로지 편집기와 웹 문서를 브라우징 할 수 있는 웹 브라우저를 포함하고 있다.
- SHOE(Simple HTML Ontology Extensions) [11]는 온톨로지를 기술하기 위한 지식표현 언어로서, HTML문서 내부에 시맨틱 어노테이션 정보를 기술할 수 있도록 HTML을 확장하여 정의하였다. SHOE의 시맨틱 어노테이션은 어노테이션 지원도구에 의해서 수행되고, 어노테이션이 수행된 문서

들은 시맨틱 검색 지원도구를 통해서 검색이 가능하다.

- KIM(Knowledge Information Management) [12]은 Ontotext Semantic Technology laboratory에서 개발한 웹기반의 자동 시맨틱 어노테이션지원도구로서, HTML문서의 시맨틱 어노테이션 정보와 인텍싱 정보를 저장, 관리하고 검색을 지원하기 위하여 별도의 KIM서버를 제공한다.
- SALT(Semantically Annotated LaTeX)[13]는 PDF 문서의 시맨틱 어노테이션을 지원하기 위한 도구이다. SALT는 LaTeX를 확장하여 PDF문서의 작성과 병행하여 어노테이션을 수행하고, 어노테이션 정보의 저장은 XMP를 사용하여 문서에 직접 부착하는 방식을 사용한다. 그러나, 기존에 작성되어 웹에 존재하는 PDF문서에 대해서는 어노테이션을 수행할 수 없으며, 시맨틱 검색기능 또한 제공하지 않는다.
- PDFTab[14]은 본 연구와 유사한 방식으로 PDF 문서의 시맨틱 어노테이션을 지원한다. PDFTab은 메타데이터의 저장관리를 위해서 XMP를 채택하였고, 온톨로지 기반의 시맨틱 어노테이션을 제공하기 위해서 온톨로지 저작도구인 Protégé의 플러그인으로 개발되었다. 그러나, 풍부한 어노테이션 정보를 표현하기 위한 어노테이션 모델과, 생성된 시맨틱 어노테이션 정보를 활용한 시맨틱 검색 등을 제공하지 않는다.



[그림 1] 시맨틱 어노테이션 지원도구(OSA)의 구조도

### 3. 시맨틱 어노테이션 지원도구

#### 3.1 사용자 인터페이스 모듈

사용자 인터페이스 모듈은 사용자에게 편의성을 제공하기 위하여, 다음과 같은 4개의 뷰들을 제공하고 있다. 그림1은 본 연구에서 개발한 시맨틱 어노테이션 지원도구(OSA)의 구조도이고, 그림2와 그림3은 각각 OSA의 실행화면과 시맨틱 검색화면이다.

- 문서 뷰(Document View) :

그림2(a)의 문서 뷰에서는 시맨틱 어노테이션의 처리 대상 문서들을 브라우징하기 위하여 HTML 뷰와 PDF 뷰를 제공하고 있다. 두 개의 뷰에서 공통으로 제공되는 기능들로는 문서내의 특정항목을 선택하는 기능과 어노테이션 기능, 그리고 어노테이션 된 항목을 표시하기 위한 하이라이트 기능들이 있다. 이외에도 HTML 뷰에서는 웹 브라우저의 기본 기능들을 추가로 제공한다. 문서 뷰에서 어노테이션한 결과는 온톨로지 뷰와 어노테이션 뷰를 통해 확인 가능하다.

- 온톨로지 뷰(Ontology View) :

그림2(b)의 온톨로지 뷰는 클래스, 인스턴스 그리고 속성 뷰로 구성된다. 클래스 뷰(Class View)는 온톨로지에 기술된 클래스의 계층 구조를 브라우징하는 기능과 클래스의 편집기능(이름, 계층 관계 등)을 제공한다. 인스턴스 뷰(Instance View)와 속성 뷰(Property View)는 클래스의 인스턴스 목록과 속성 목록을 각각 출력한다.

- 어노테이션 뷰(Annotation View) :

그림2(c)의 어노테이션 뷰는 인스턴스 에디터(Instance Editor)와 어노테이션 디스플레이어(Annotation Displayer)로 구성된다. 인스턴스 에디터에서는 온톨로지 뷰에서 선택된 인스턴스와 인스턴스의 속성값 편집이 가능하며, 이와 같은 편집과정에서 생성된 인스턴스 또는 인스턴스의 속성값들은 문서의 시맨틱 어노테이션 정보로 저장된다. 또한, 생성된 어노테이션 정보들은 어노테이션 디스플레이어에 포함된, PDF 개요 뷰(PDF Description View)와 PDF 어노테이션 정보 뷰(PDF Annotation Information View)에서 출력된다. PDF 개요 뷰에서는 PDF문서의 기본 메타데이터인 저자, 제목, 키워드 등을 편집할 수 있으며, PDF 어노테이션 정보 뷰는 문서에 포함된 시맨틱 어노테이션 정보를 출력한다.

- 시맨틱 검색 뷰(Semantic Search View) :

그림3의 시맨틱 검색 뷰에서는 어노테이션 정보가 기술된 온톨로지의 브라우징과 내비게이션 기능을 수행할 수 있는 온톨로지 내비게이터가 제공된다. 온톨로지 내비게이터(Ontology Navigator)는 다음과 같은 6개의 하위 뷰들로 구성된다. 타입 뷰(Type View), 패싯 뷰(Facet View), 그리고 인스턴스/리터럴 뷰(Instance/Value View)에서는 각각 온톨로지의 클래스, 속성, 그리고 속성값을 브라우징한다. 온톨로지 내비게이터의 검색은 사용자의 UI선택 이벤트에 의해서 단계적으로 수행되며, 수행 과정은 검색기록 뷰(History View)에 기록된다. 또한, 검색 결과인 문서목록은 결과 뷰(Result View)에서 출력되어지고, 정보 뷰(Information View)에서는 타입 뷰, 패싯 뷰, 인스턴스 뷰에서 선택된 온톨로지 자원의 추가 정보를 표시한다.

#### 3.2 프로세싱 모듈

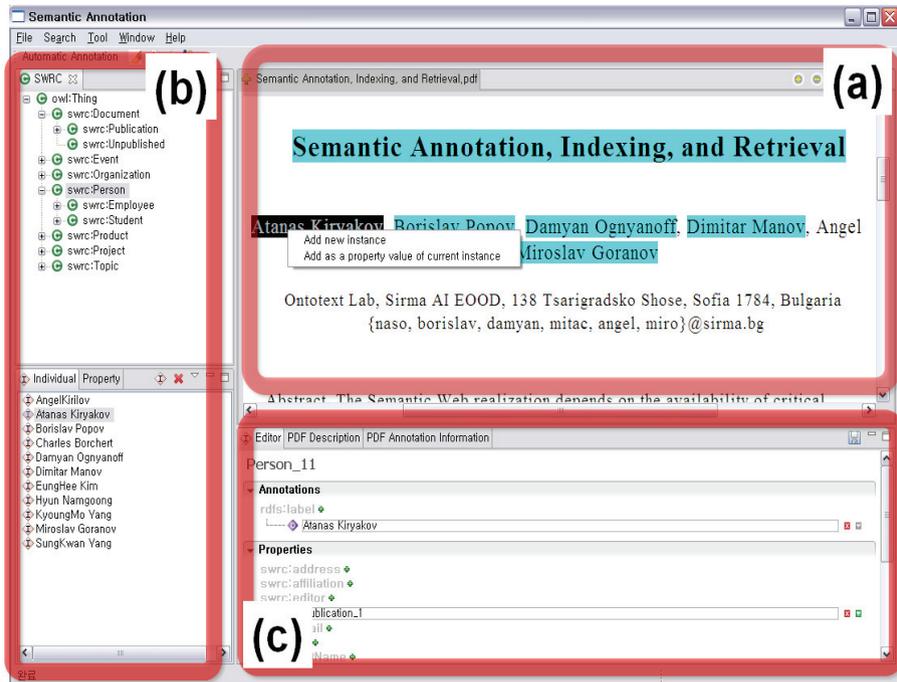
프로세싱 모듈은 사용자 인터페이스의 입력 데이터와 출력 데이터를 처리하기 위해서 다음과 같은 5개의 모듈들로 구성되어 있다.

- 문서 모듈(Document Module) :

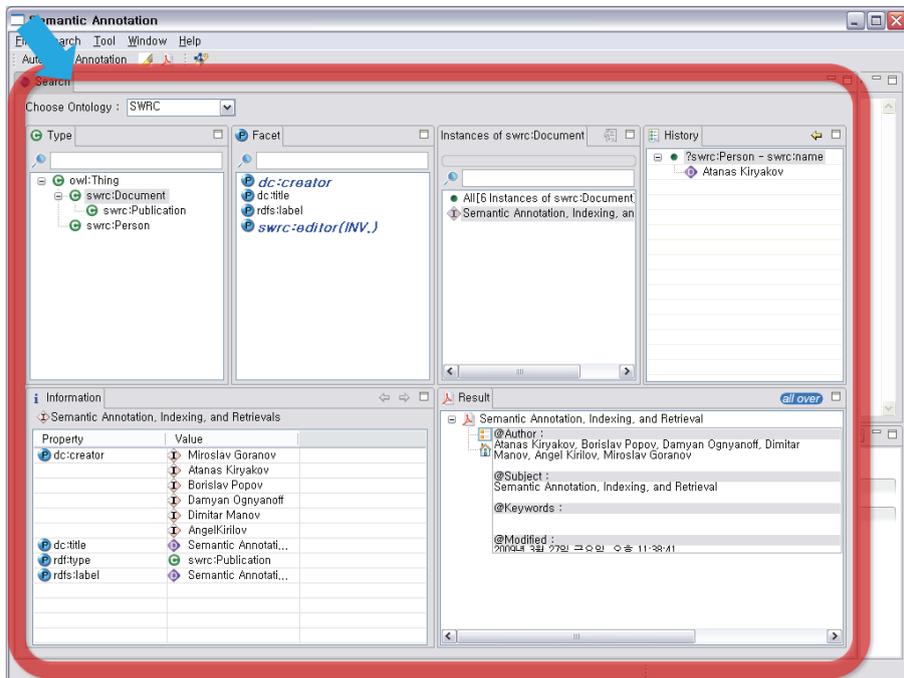
문서 모듈에서는 문서 뷰에 출력될 HTML문서와 PDF문서를 처리하기 위한 기능들을 제공한다. HTML문서의 처리는 Eclipse RCP(Rich Client Platform)에서 제공하는 SWT Browser를 확장하여 구현되었고, PDF문서를 처리하기 위한 처리모듈은 JAVA기반의 iText Library[21]를 기반으로 구현되었다.

- 온톨로지/XMP 모듈(Ontology/XMP Module) :

온톨로지 모듈은 HP Labs Semantic Web Research에서 개발한 온톨로지 관리 프레임워크인 Jena[22]를 기반으로 개발되었으며, 온톨로지 모듈에서 관리되는 온톨로지들은 시맨틱 어노테이션 모듈과 시맨틱 검색 모듈에서 사용된다. 또한, 온톨로지의 저장, 가져오기 그리고 내보내기는 데이터 접근 모듈을 통해서 처리된다. XMP모듈은 PDF문서에 시맨틱 어노테이션 정보를 부착시키거나 또는 부착된 시맨틱 어노테이션 정보의 추출 및 관리기능을 수행하며, 이와 같은 기능을 지원하기 위해서 Adobe에서 배포한 XMP SDK[23]를 확장하여 구현하였다.



[그림 2] OSA의 문서 뷰(a), 온톨로지 뷰(b) 그리고 어노테이션 뷰(c) 화면



[그림 3] OSA의 시맨틱 검색 뷰 화면

● 시맨틱 어노테이션 모듈(Semantic Annotation Module) :

시맨틱 어노테이션 모듈에서는 RDFS로 기술된 시맨틱 어노테이션 모델의 관리를 담당한다. 시스템에서 생성되는 모든 어노테이션 정보는 시맨틱 어노테이션 모델로 표현되어 관리되고, 시맨틱 어노테이션 모델은 문서의 저장시 XMP모듈을 통해서 PDF문서에 부착된다.

● 시맨틱 검색 모듈(Semantic Search Module) :

시맨틱 검색 모듈은 온톨로지 내비게이터에서 발생하는 이벤트를 감지하여 온톨로지의 질의를 생성하고 검색을 수행하는 이벤트기반의 온톨로지 검색 모듈이다. 이벤트를 기반으로 생성되는 질의의 표현과 온톨로지의 검색은 온톨로지 질의언어인 SPARQL[24]과 Jena프레임워크에서 구동되는 ARQ[25]질의엔진을 사용한다.

● 데이터 접근 모듈(Data Access Module) :

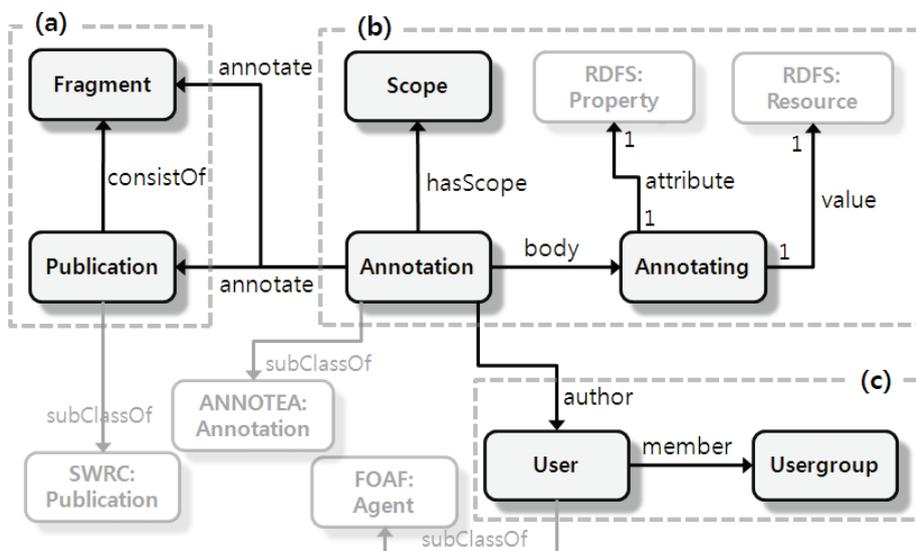
데이터 접근 모듈에서는 시스템에서 사용되는 온톨로지들을 로컬 저장소에 저장하고 관리하는 기능과 웹 어노테이션 서버와 연계하여, 웹 어노테이션 수행시 생성되는 어노테이션 정보들을 저장관리 한다. 또한, 어노테이션의 대상문서(PDF문서)를 파일 시스템으로부터 불러오고 저장하는 기능을 수행한다.

3.3 로컬 저장소와 웹 어노테이션 서버

로컬 저장소(Local Storage)는 OSA에서 사용되는 온

톨로지들의 저장 공간으로서, 사용자 온톨로지와 시스템 온톨로지를 저장하고 관리한다. 사용자 온톨로지란, 사용자 개인의 선택에 의해서 등록된 것으로서, 개인 사용자의 관심 분야를 정의한 개인 온톨로지(Personal Ontology)와 특정 도메인 정보를 정의한 도메인 온톨로지(Domain Ontology)를 의미한다. 한편, 시스템 온톨로지는 시스템에 필요한 데이터들을 표현하고 관리하기 위해서 사용되는 것으로서, OSA에는 문서 온톨로지(Document Ontology)와 메타-온톨로지(Meta Ontology)가 등록되어 있다. 문서 온톨로지에는 시맨틱 검색에 활용하기 위한 문서의 정보(문서의 경로 및 요약정보)와 해당 문서에 첨부한 시맨틱 어노테이션 정보가 기술되고, 메타-온톨로지에는 시스템에서 사용되는 온톨로지들의 관리를 위해서 온톨로지들의 메타정보(version, comment, label, baseURI 등)들이 기술된다.

웹 어노테이션 수행시, 즉, HTML문서의 시맨틱 어노테이션 수행시 어노테이션 정보를 HTML문서에 직접 부착시키는 것은 불가능하다. 따라서, 어노테이션 정보들을 저장관리하기 위한 별도의 저장소를 필요로 하며, 웹 어노테이션에서 생성되는 어노테이션 정보들을 저장관리는 저장소의 기능을 웹 어노테이션 서버(Web Annotation Server)가 담당한다. 또한, 웹 어노테이션 서버에서는 도메인 온톨로지들을 로컬 저장소에 등록할 수 있는 서비스를 제공하여 다양한 도메인 온톨로지의 사용이 가능하다.



[그림 4] OSA의 시맨틱 어노테이션 모델의 개요

## 4. OSA의 시맨틱 어노테이션 모델과 사용 시나리오

### 4.1 시맨틱 어노테이션 모델

본 장에서는, OSA에서 시맨틱 어노테이션 정보를 표현하기 위해 RDF Schema로 기술된 시맨틱 어노테이션 모델에 대해서 설명한다. 본 논문에서 제안하는 시맨틱 어노테이션 모델은 크게 3개의 부분으로 구성되어 있고, 그림4는 OSA의 시맨틱 어노테이션 모델의 개요를 나타내고 있다.

- Publication부분

그림4(a)의 Publication부분에서는 시맨틱 어노테이션 대상문서들을 표현하기 위해서 SWRC[26] 온톨로지의 어휘를 도입하여 “Publication”클래스를 정의하였다. Publication클래스는 SWRC:Publication클래스에 정의된 “dc:title”, “dc:creator”, “swrc:citedBy” 등의 다양한 속성들을 재사용할 수 있다. 또한, 문서의 특정 부분들을 표현하기 위한 “Fragment”클래스를 정의하였고, Publication클래스와는 “consistOf”관계로 연결되어 있다.

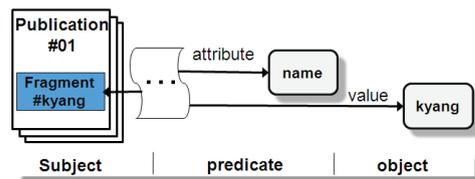
- Annotation부분

그림4(b)의 Annotation부분에서는 Annotea Project에서 제공하는 Annotea Annotation Schema[27]에 정의된 “ANNOTEA:Annotation”클래스의 서브클래스로 “Annotation”클래스를 정의 하였다. ANNOTEA:Annotation클래스의 속성들로는, 어노테이션의 작성자를 표현하기 위한 “author”속성, 온톨로지 기반의 의미정보와 어노테이션을 연결시키기 위한 “body”속성 그리고 생성일과 변경일을 표현하기 위한 “created”, “modified”속성들이 있다. 또한, “annotate”속성은 어노테이션 대상 자원과 어노테이션을 연결하기 위해 사용된다. “Annotating”클래스는 온톨로지 기반의 의미정보를 표현하기 위한 클래스로서, “attribute”속성과 “value”속성을 사용하여 온톨로지의 트리플 정보(Subject, Predicate, Object)를 표현한다. 그림5는 attribute속성과 value속성을 사용하여 온톨로지의 트리플 정보를 표현한 예이다. 추가적으로, Predicate와 Object의 쌍을 기술하기 위해서 각 속성이 취할 수 있는 값의 개수(Cardinality)를 1로 제약하였다. “Scope”클래스는 어노테이션 정보의 접근 권한에 관한 정보를 기술한다. 본 어노테이션 모델에서는, 작성자만이 접근 가능한 “Private”, 작성자가 포함된 커뮤니티 내에서 공유가 가능한 의미의 “Shared”, 그리고 공개의 의미인 “Public”인스턴스들을 정의 하였다.

- User부분

그림4(c)의 User부분에서는, 사용자들의 정보를 기술하기 위해서 FOAF[28]의 어휘를 도입하였고, 커뮤니티를 표현하기 위해서 “Usergroup”클래스를 추가로 정의하였다. 또한, 어노테이션의 작성자를 표현하기 위해서 “author”속성을 사용한다.

본 장에서 설명한, OSA의 시맨틱 어노테이션 모델에서는 Dublin Core와 같은 기본 어휘뿐만 아니라, 기존에 잘 정의된 SWRC, ANNOTEA, FOAF 등의 어휘를 도입하여, 사용자정보, 문서의 정보, 온톨로지 기반의 의미정보 등을 보다 풍부하게 기술할 수 있다.



[그림 5] 온톨로지 트리플의 표현 예

### 4.2 XMP기반의 시맨틱 어노테이션

지속적으로 증가하는 디지털 문서의 효과적인 검색과 관리 및 처리를 위해서 시맨틱 웹이 제안되었고, 시맨틱 웹에서는 웹 자원들을 의미적으로 연결하는 메타데이터의 생성과 효과적인 관리기술이 요구되고 있다. 이와 같은 요구들을 충족시키기 위해서, W3C에서는 메타데이터를 기술하기 위한 RDF를 제안하였으며, Adobe사에서는 디지털자원을 분류하거나 관리를 가능하게 하는 메타데이터 관리 플랫폼인 XMP[15,16]를 제안하였다. XMP는 메타데이터를 기술하기 위해서 RDF를 사용하고 있으며, 사용자가 XMP에 사용자지정 스키마를 기술할 수 있도록 확장성을 제공한다. 본 연구에서는 4.1절에서 소개한 RDF Schema로 기술된 시맨틱 어노테이션 모델을 사용하여 XMP스키마를 확장하는 방식으로, 풍부한 시맨틱 어노테이션 정보를 저장하고 관리할 수 있도록 하였다.

시맨틱 어노테이션의 사용 시나리오를 설명하기 위해서, OSA의 HTML문서와 PDF문서에 대한 시맨틱 어노테이션 중 PDF문서에 대한 시맨틱 어노테이션의 사례에 대해서 설명한다. 본 사례에서 사용되는 온톨로지는 SWRC(Semantic Web for Research Communities)에서 배포한 SWRC온톨로지[26]를 사용한다. SWRC온톨로지는 사람, 조직, 간행물(학술자료, 도서)과 같은 연구 커뮤니티 정보들의 관계를 기술하기 위한 온톨로지이다.

다음에 설명할 어노테이션 시나리오는 참고문헌[6]의

저자 7명 중에서 “Atanas Kiryakov”에 대한 시맨틱 어노테이션을 수행하는 과정으로 기술하였다. 그림6은 시맨틱 어노테이션 수행 화면을 나타낸다.

- ① 온톨로지의 클래스 뷰에서 Person클래스를 선택하면, 인스턴스 뷰에 Person클래스의 인스턴스목록이 나열된다. 나열된 인스턴스목록 중에서 “Atanas Kiryakov”를 지칭하는 인스턴스가 존재하면 아래의 ②를 수행하고 존재하지 않으면 ③을 수행하게 된다.
- ② 인스턴스 뷰에 “Atanas Kiryakov”를 지칭하는 인스턴스가 존재하면, 해당 인스턴스를 선택한 후에 인스턴스의 name속성값으로 “Atanas Kiryakov”를 할당한다.
- ③ 인스턴스 뷰에 “Atanas Kiryakov”의 인스턴스가 존재하지 않으면, 문서 뷰에서 “Atanas Kiryakov”항목을 선택하고 어노테이션 팝업메뉴에서 “Add new instance”를 선택하여 저자 “Atanas Kiryakov”를 지칭하는 인스턴스를 생성한다. 생성된 인스턴스는 rdf:label속성의 기본값으로 “Atanas Kiryakov”가 할당된다. 추가적인 어노테이션이 필요한 경우, 과정 ②를 반복 수행한다.

특히, 과정②의 인스턴스에 속성값을 할당하는 방식은 다음과 같은 3가지 방식이 가능하다. 첫째, 문서의 텍스트를 선택하여 끌어놓기(Drag&Drop)기능으로 할당하는 방식, 두 번째는 직접 인스턴스 에디터 뷰에서 인스턴스에 속성값을 추가하는 방식, 그리고 세 번째는 문서 뷰에서 제공하는 팝업메뉴에서 “Add as a property value of current instance”를 선택한 후 속성선택 다이얼로그에서 해당 속성을 선택하는 방식이다.

시맨틱 어노테이션 과정에 의해서 인스턴스와 인스턴스의 속성값으로 기술된 시맨틱 어노테이션 정보가 생성되고, 이때 생성된 어노테이션 정보는 그림7의 어노테이션 정보 뷰에서 확인이 가능하다. 정보 탭(Information Tab)에는 항목 탭(Entity Tab)에서 선택된 Atanas Kiryakov항목의 시맨틱 어노테이션 정보를 표시한다. 문서내의 “Atanas Kiryakov”는 rdf:type이 Person클래스이고 이름이 Atanas Kiryakov이다. 추가적으로, “Semantic Annotation, Indexing, and Retrieval”인스턴스, 즉, 참고문헌[6]의 저자라는 정보를 확인할 수 있다. 또한, 그림8의 PDF 개요 뷰는 PDF문서에 대한 기본적인 타이틀, 저자, 주제, 키워드 그리고 문서의 생성일 및 변경일과 같은 메타데이터를 보여준다. 여기서, 문서의 생성일 및 변경일을 제외한 모든 정보들은 사용자에게 의해서 수정이 가능하다.

본 논문에서는 OSA의 시맨틱 어노테이션수행 과정을

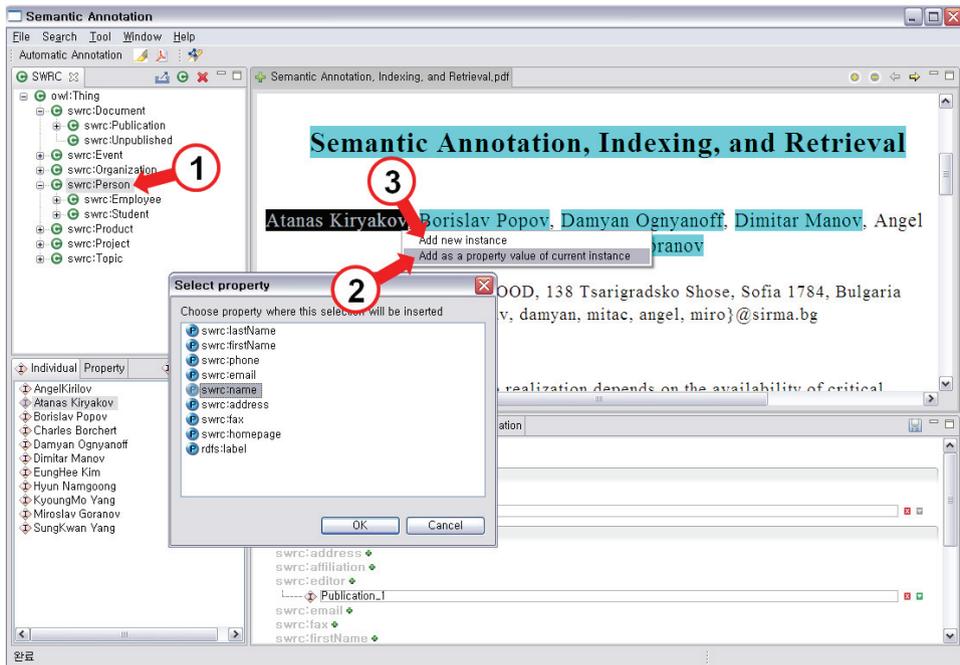
설명하기 위하여 어노테이션 수행 시에 참조되는 온톨로지를 단 1개(SWRC온톨로지)만으로 한정하여 사용하였다. 그러나, 개별 문서에는 다양한 영역의 지식들이 포함될 수 있고, 이러한 다양한 지식에 하나의 온톨로지만을 참조하여 어노테이션을 수행하는 것은 분명히 한계가 있다. 이와 같은 한계점을 극복하기 위해서 OSA에서는 2개 이상의 온톨로지들을 참조하여 어노테이션을 수행하는 것이 가능하다. 예를 들어, 참조문헌[6]의 저자인 “Atanas Kiryakov”와 “Dimitar Manov”에 어노테이션을 수행하기 위하여 FOAF 온톨로지를 추가로 로딩하고 FOAF에 정의된 많은 어휘들을 사용하여 보다 풍부한 의미정보를 생성할 수 있다.

### 4.3 시맨틱 검색

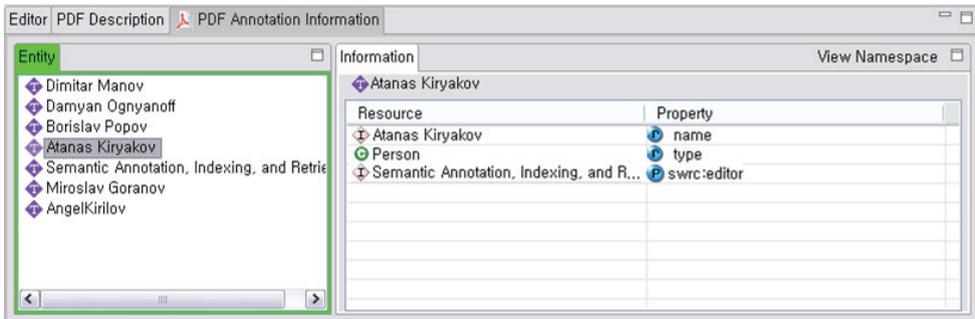
문서에 시맨틱 어노테이션 정보를 부여하는 목적은 저장되어있는 다량의 정보들 중에서 사용자가 요구하는 정보를 효율적으로 검색하여 사용자에게 전달하고자 하는데 있다. 그러나, 기존의 검색방법(키워드검색)을 사용하여 시맨틱 어노테이션 정보가 부여된 문서를 검색하는 것은 매우 비효율적이다[6,18].

한편, 패시 브라우징은 패시 분류(Faceted Classification)과정을 통해서 정보영역을 직교개념의 차원(Orthogonal Conceptual dimensions)들로 분할하여 사용자 인터페이스 기반으로 브라우징과 내비게이션을 수행하는 검색기술이다[29]. 패시 분류과정에서 분류의 기준이 되는 것을 패시(Facet)이라고 하며, 예를 들어, 예술 작품정보의 분류를 위해 패시들을 선정한다면, 작품의 주제, 아티스트 이름, 제작 년도 등이 될 수 있다. 또한, 아티스트의 이름인 “레오나르도 다빈치”와 같은 값을 분할된 정보영역을 제약하여 축소시키기 위한 제약값(Restriction-value)이라고 한다. 이와 같은 패시 분류를 통해서, 아티스트의 이름이 “레오나르도 다빈치”인 예술품의 검색이 가능하다. 이와 유사하게, 주어(Subject), 술어(Predicate), 목적어(Object)로 구성된 RDF의 구조는, 주어가 검색의 대상이 되는 정보를 표현하고 술어는 패시으로 고려될 수 있다. 또한, 목적어는 패시로 분류된 정보의 제약값(Restriction-value)이 된다[30].

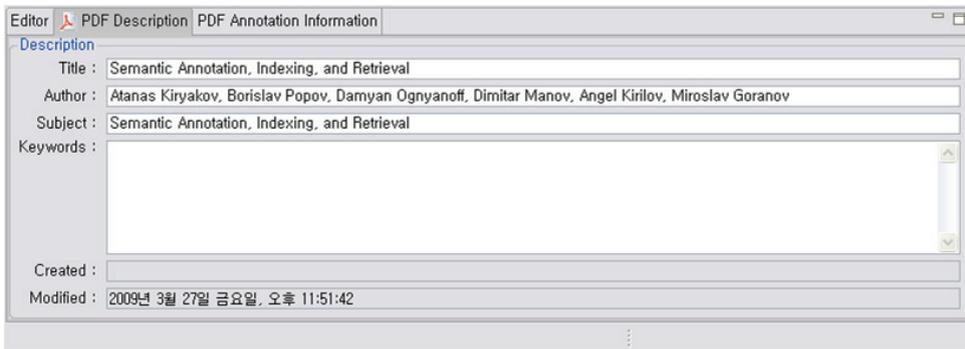
본 논문에서는, 시맨틱 어노테이션 정보가 기술된 온톨로지(어노테이션시에 참조된 온톨로지)를 패시 브라우징기술과 온톨로지 질의 언어인 SPARQL을 사용하여 사용자 인터페이스 기반의 시맨틱 검색을 제공한다. 패시 브라우징에서는, 사용자가 단계적으로 인터페이스를 사용하여 제약사항을 추가하는 과정에 의해서 질의가 생성된다. 또한, 사용자들은 중간결과와 함께 추가 질의를 생성하기 위한 내비게이션 기능을 제공받는다.



[그림 6] OSA의 시맨틱 어노테이션 수행 화면

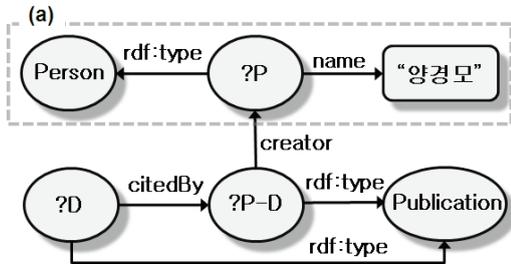


[그림 7] OSA의 PDF 어노테이션 정보 뷰 화면



[그림 8] OSA의 PDF 개요 뷰 화면

시맨틱 검색을 수행하기 위해서 제공되는 시맨틱 검색 뷰는 6개의 서브 뷰로 구성된다. 그 중에서, 타입 뷰(Type View), 패킷 뷰(Facet View), 그리고 인스턴스/리터럴 뷰(Instance/Value View)는 질의생성 뷰로서, 각각 온톨로지의 클래스, 속성, 그리고 인스턴스와 리터럴값을 브라우징할 수 있다. 각 뷰에서 질의를 생성할 수 있는 이벤트는 다음과 같으며, 그림9는 해당 이벤트들에 의해서 생성될 수 있는 질의의 예를 나타내고 있다.



[그림 9] 질의 예제

● 타입 선택 :

타입 선택은 검색대상의 타입을 결정하고 검색대상의 분류기준인 패킷 목록을 추천받는다. 타입 선택의 예로, 사람을 검색하기 위해서 Person클래스를 선택함으로써, Person클래스와 관련된 속성목록이 패킷 뷰에 브라우징된다.

● 패킷 선택 :

패킷 선택은 검색대상을 제약시키기 위한 패킷을 결정하고 제약값 목록을 추천받는다. 예를들어, name속성으로 선택하면, 제약값의 목록으로는 Person클래스 인스턴스들의 name속성값이 브라우징된다. 또한, 패킷의 종류를 주어가 속성의 도메인(Domain)이 되는 기본 패킷과 기본 패킷의 역을 표현한 역(Inverse) 패킷으로 분류하여, 역관계의 검색 또한 가능하다.

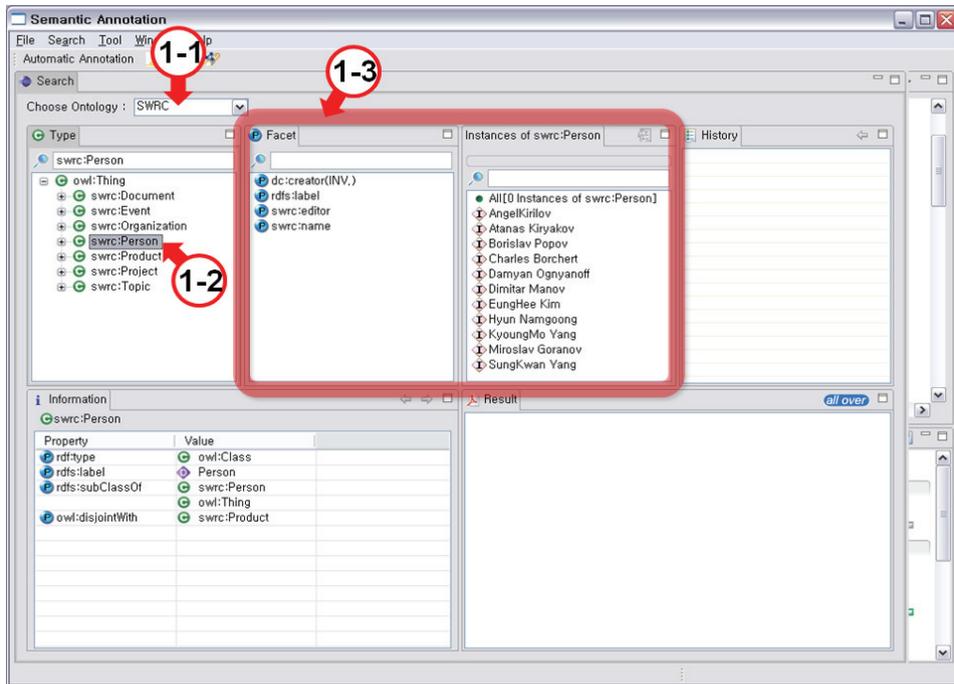
● 제약값 선택 :

제약값 선택은 하나의 SPARQL 질의를 결정한다. 그림 9(a)에서는 제약값으로 “양경모”를 선택한 경우이고, 추가적으로, “양경모”가 작성한 문서(?P-D)를 참고문헌으로 사용하는 문서(?D)를 검색하기 위한 복잡한 질의 또한 단계적으로 생성할 수 있다. 또한, 제약값을 효율적으로 결정할 수 있도록, 정규표현식을 사용한 문자열의 매칭과 날짜형식과 숫자형식의 특정 구간을 제약값으로 결정할 수 있는 기능들을 정의하였다.

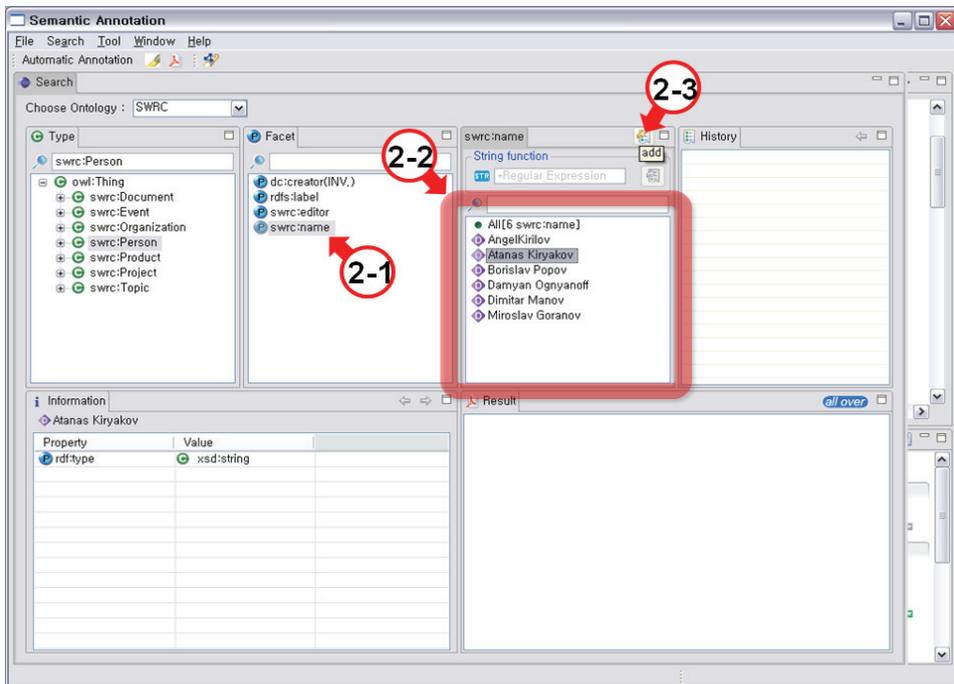
검색기록 뷰(History View)에서는 사용자의 검색기록을 목록으로 나열하여 표시하고 취소기능을 사용하여 이전 검색결과로 되돌아갈 수 있다. 정보 뷰(Information View)에서는 서브 뷰에서 선택된 온톨로지 자원(클래스, 인스턴스, 속성)의 정보를 테이블 형태로 제공한다. 마지막으로 결과 뷰(Result View)에서는 검색의 최종 결과인 문서를 나열하여 표시한다.

OSA에서 제공하는 시맨틱 검색기능을 간단한 “Atanas Kiryakov가 저자인 문서는?”이라는 질의어를 토대로 설명하기로 한다. 그림10-13은 OSA의 시맨틱 검색을 순서대로 실행한 화면이다.

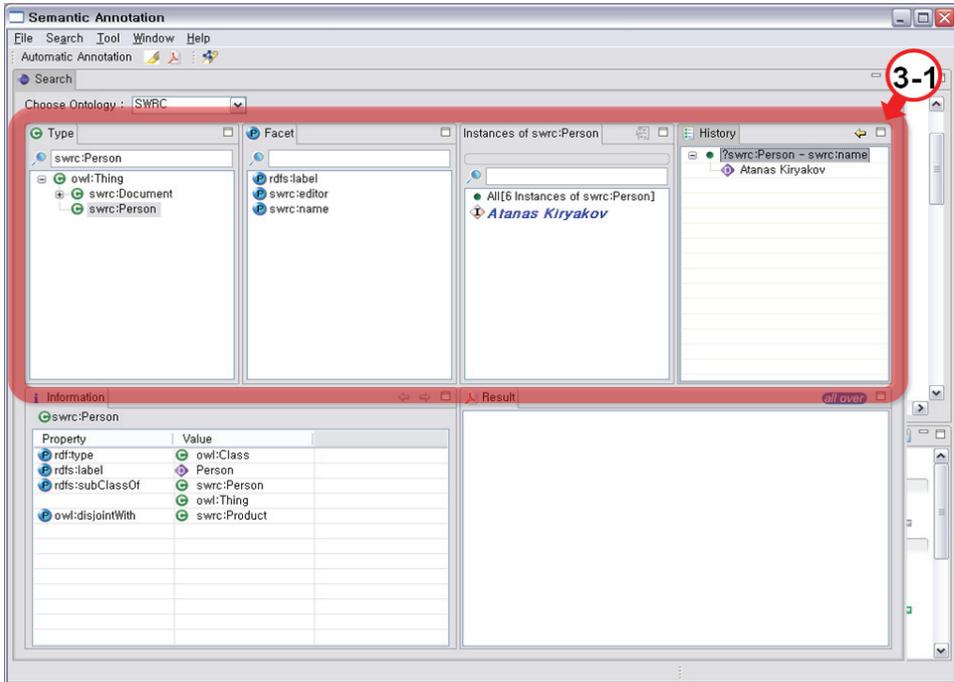
- ① 그림10의 “Choose Ontology” 콤보상자에서 시맨틱 어노테이션에 사용된 SWRC온톨로지를 선택하고, “Atanas Kiryakov”를 검색하기 위해서 타입 선택으로 Person클래스를 선택한다. Person클래스가 선택되면, 인스턴스/리터럴 뷰에 Person클래스의 모든 인스턴스 목록이 나열되고 패킷 뷰에는 인스턴스/리터럴 뷰에 나열된 인스턴스들이 가지고 있는 모든 속성들이 나열된다.
- ② 그림11과 같이, 사람의 이름에 대한 제약사항을 추가하기 위해서 name속성을 선택한다. 패킷 뷰에서 name속성이 선택되면 인스턴스/리터럴 뷰에는 Person클래스의 모든 인스턴스들이 가지고 있는 name속성값 목록이 나열된다. 나열된 목록 중에서 name속성의 제약값으로 “Atanas Kiryakov”를 선택하고 우측 상단의 Add버튼을 선택한다.
- ③ 제약값이 추가되면 시맨틱 검색 컴포넌트는 SPARQL로 기술된 질의를 수행하여 검색된 인스턴스와 연관성이 없는 인스턴스들을 필터링한다. 이 같은 필터링 과정을 통해서 사용자에게 선택이 가능한 항목을 추천해주는 내비게이션 기능을 제공할 수 있다. 그림12에서는, Document클래스의 인스턴스와 관련된 사항들로 추가적인 검색이 가능하다.
- ④ 마지막으로, 그림13은 문서의 검색에 관한 질의를 수행하는 과정이다. 타입 뷰에서 Document클래스를 선택하면, 인스턴스/리터럴 뷰에는 이전 단계에서 검색된 “Atanas Kiryakov”의 인스턴스와 관련된 Document클래스의 인스턴스가 나열된다. 또한, 패킷 뷰와 정보 뷰에서 나열된 인스턴스의 작성자가 “Atanas Kiryakov”라는 정보를 확인할 수 있다. 검색된 인스턴스를 선택하면, 결과 뷰에 해당 인스턴스를 참조하여 시맨틱 어노테이션을 수행한 문서의 정보가 표시된다.



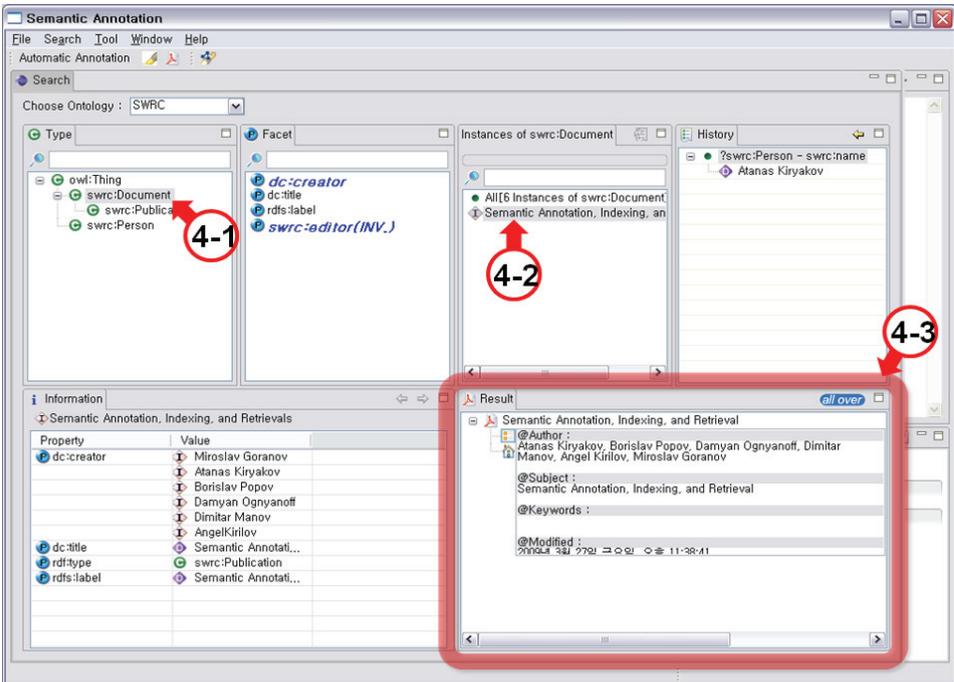
[그림 10] 타입 선택 화면



[그림 11] 패킷 선택 화면



[그림 12] “Atanas Kiryakov”의 인스턴스 검색 화면



[그림 13] “Semantic Annotation, Indexing, and Retrieval”문서 검색 화면

【표 1】 시맨틱 어노테이션 지원도구 특징 비교

	처리대상 문서형식	메타데이터 내장	메타데이터 내장 기술	시맨틱 검색 기능 지원	시맨틱 어노테이션/검색에 활용 가능한 온톨로지
Annozilla[8]	HTML	×	-	×	-
MnM[9]	HTML	○	XML	×	-
OntoMat-Annotizer[10]	HTML	×	-	×	-
SHOE[11]	HTML	○	HTML 확장	○	SHOE 온톨로지
KIM[12]	HTML	×	-	○	KIM 온톨로지
SALT[13]	PDF	○	XMP	×	-
PDFTab[14]	PDF	○	XMP	×	-
OSA	HTML, PDF	○	XMP	○	RDF/OWL로 기술된 온톨로지

### 5. 기존 도구들과의 기능 비교

본 연구의 주요 목적은, XMP 및 온톨로지 기반의 시맨틱 어노테이션 지원도구(OSA)를 개발하여, 웹 문서에 대한 풍부한 시맨틱 어노테이션을 수행하고 시맨틱 어노테이션 과정에서 참조한 온톨로지와 생성된 메타데이터를 기반으로 시맨틱 검색을 지원하는 것이다. 본 연구결과로서 개발된 OSA의 기능상의 특징과 장점에 대해서 기존 도구들[8-12]과 비교해 보면 표1과 같다.

- ① 기존의 도구들[8-14]은 시맨틱 어노테이션 대상 문서 형식으로 HTML문서 또는 PDF문서 중에서 한 가지만을 처리하는 반면에, 본 연구결과(OSA)에서는 HTML문서와 PDF문서를 모두 시맨틱 어노테이션 처리할 수 있다.
- ② 기존의 도구들[8-14]은 생성된 메타데이터를 관리하기 위해서, 메타데이터를 문서 내부에 직접 부착하는 방식 또는, 별도의 서버에 저장하여 관리하는 방식을 사용하고 있다. 한편, OSA에서는, 메타데이터의 관리에 대한 호환성을 높이기 위해서 메타데이터 관리 플랫폼으로서 국제표준으로 제정된 XMP[15,16] 표준기술을 도입하여 채용하고 있다.
- ③ 기존의 도구들[8-14] 중에서 시맨틱 검색 기능을 제공하고 있는 것으로는 SHOE와 KIM뿐이며, 이들 도구들은 각각의 시스템에서 정의한 온톨로지만을 기반으로 시맨틱 어노테이션 기능과 시맨틱 검색 기능을 제공한다. 따라서, SHOE와 KIM은 특정 도메인의 정보 또는 개인 사용자가 요구하는 정보를 시맨틱 어노테이션 처리하여 검색하는데 많은 제약이 있다. 한편, OSA에서는, W3C표준으로 제정된 온톨로지 기술언어인 RDF 및 OWL을 사용하여 표현된 개인/도메인 온톨로지(Personal/Domain Ontology)를 처리하기 때문에 SHOE와 KIM에 비

해서 시맨틱 어노테이션 기능과 시맨틱 검색 기능의 범용성이 더 높다.

한편, 시맨틱 검색 기능을 제공하고 있는 SHOE와 KIM, 그리고 OSA에 대해서, 각각의 시맨틱 검색결과를 비교하는 작업을 수행함으로써, 본 연구에서 개발된 도구의 장점을 살펴볼 수 있다. 그러나, SHOE의 경우, 시맨틱 검색 기능을 웹 사이트

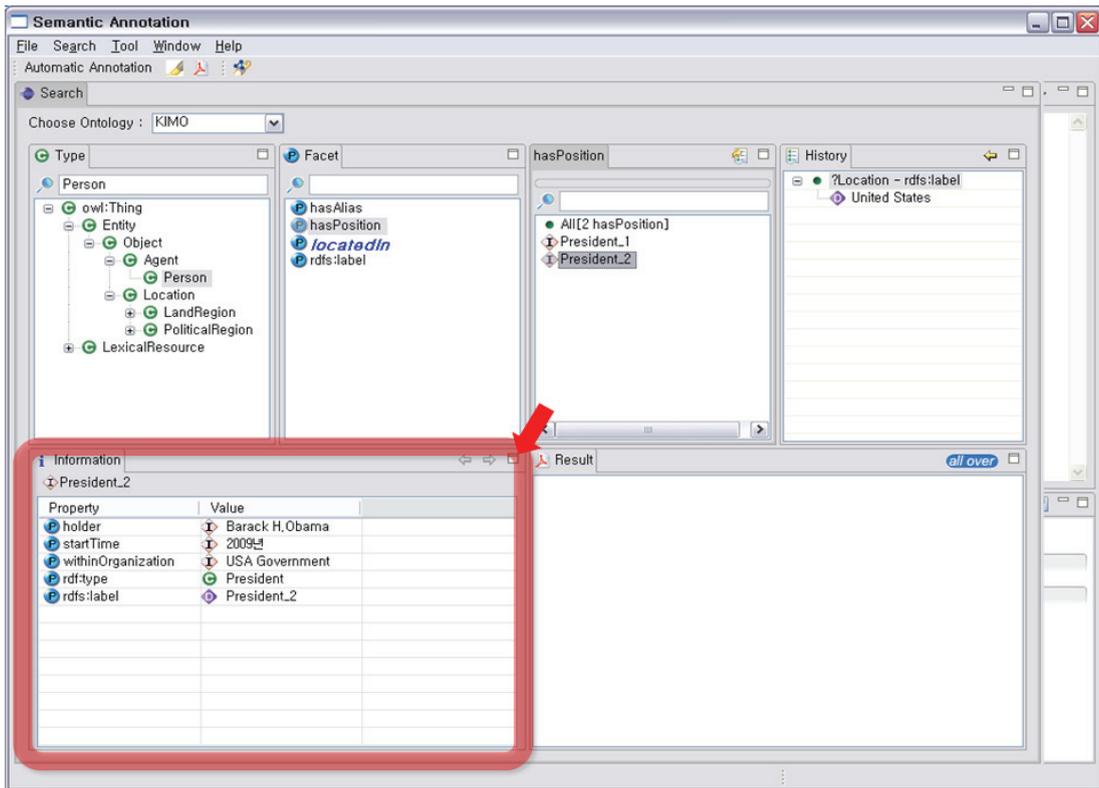
(<http://www.cs.umd.edu/projects/plus/SHOE/search/>)에서 자바 애플릿으로 제공하고 있으나, 2009년 06월 19일 현재, 시스템의 오류로 인하여 서비스가 중단된 상태이다. 따라서, KIM과 본 연구의 결과(OSA)의 시맨틱 검색 결과를 비교해보기로 한다. 먼저, 시맨틱 검색기능을 비교하기 위하여, 임의의 질의로서 “현재의 미국대통령과 관련된 정보는?”을 선택하였다. KIM과 OSA 각 시스템에 대해서, 위에서 선택된 질의 “현재의 미국대통령과 관련된 정보는?”을 검색하는 과정을 수행하였다.

KIM에서는, 그림14의 “Location”에서 “United States” 항목을 선택을 한 후 “Job Positions”에서 현재 미국 대통령을 지칭하는 “President”를 선택한다. 그러나, KIM에서는 검색에 활용할 수 있는 항목을 클래스와 인스턴스만으로 한정하여 제공하고 있다. 따라서, 사용자가 현재의 미국 대통령에 대한 정보(이름 등)를 알지 못하는 상태에서 정확한 “President”항목을 선택할 수 없다. 또한, 미국의 대통령과 관련된 “President”항목을 선택하였더라도, “Person”의 항목들 중에서 현재의 대통령을 검색하는 것은 불가능하다.

한편, OSA에서는, 그림15와 같이 “Location”클래스를 선택하고 “United States”를 검색한다. “United States”와 관련된 항목들로 검색의 범위가 축소되고, 타입으로 “Person”클래스를 선택한 후, “hasPosition”속성을 패킷으로 선택을 하여 미국 대통령에 대한 정보를 검색할 수 있다. 또한, 정보 뷰를 통해서 대통령의 당선 연도



[그림 14] KIM의 검색 결과 화면



[그림 15] OSA의 검색 결과 화면

(startTime)를 선택하여 현재의 대통령에 대한 정보에 대해서 확인할 수도 있다.

위와 같은 시맨틱 검색기능에 대한 KIM과 OSA의 검색결과를 비교해 보면, 기존의 도구들 중에서 시맨틱 검색 기능을 지원하는 KIM에서 검색 불가능했던 것을 본 연구결과(OSA)에서는 보다 정확하게 검색해 낼 수 있다. 따라서, 본 연구결과(OSA)는 기존의 도구들과 비교했을

때, 보다 정확하고 효율적인 시맨틱 검색 기능을 지원하고 있음을 알 수 있다.

## 6. 결론 및 향후과제

본 연구에서는, 급증하는 디지털 문서들과 그에 포함

된 유용한 정보를 효과적으로 관리하기 위해서, 디지털 문서의 시맨틱 어노테이션 모델을 제안하고, 제안된 모델을 기반으로 생성된 시맨틱 어노테이션 정보들을 XMP를 사용하여 저장·관리하는 시맨틱 어노테이션 지원도구(OSA)를 개발하였다.

OSA의 특징 및 장점을 요약하면 다음과 같다.

- ① 디지털 문서(HTML, PDF)에 포함된 단어, 문장과 같은 요소들을 대상으로 시맨틱 어노테이션을 수행할 수 있으며 생성된 시맨틱 어노테이션 정보들은 메타데이터 관리 플랫폼인 XMP를 사용하여 저장하고 관리된다.
- ② OSA의 시맨틱 메타데이터 모델을 RDF Schema로 정의하고 XMP를 확장하여 정의함으로써, Dublin Core와 같은 기본 메타데이터 스키마를 사용하는 것에 비해서 매우 풍부한 시맨틱 어노테이션 정보를 표현할 수 있다.
- ③ 시맨틱 어노테이션 정보가 기술된 온톨로지를 페킷 브라우저 기술과 온톨로지 검색기술을 사용한 사용자 인터페이스 기반의 시맨틱 검색을 제공함으로써, 검색의 편의성이 증가하여 사용자가 요구하는 정보를 보다 정확하고 효율적으로 검색 할 수 있다.

향후 연구과제로는, OSA의 다중 온톨로지 참조기능을 보완하는 것과, 디지털 문서들의 다양한 문서형식을 지원하도록 OSA가 처리할 수 있는 문서형식을 다양화(OpenOffice, Image 등)하는 부분이 있다. 특히, 다중 온톨로지 참조기능의 보완은, 시맨틱 어노테이션을 수행하기 위하여 참조되는 도메인 온톨로지 또는 개인 온톨로지들을 하나의 통합된 온톨로지로 생성하기 위한 온톨로지 통합기술을 토대로 하고 있다. OSA에 온톨로지 통합기술을 도입하여 다양한 도메인 온톨로지에 공통으로 정의된 용어들을 하나의 통합된 개인 온톨로지로 관리함으로써, 시맨틱 어노테이션 정보가 중복생성되는 것을 방지할 수 있으며, 통합된 개인 온톨로지는 시맨틱 검색에서 유용하게 재사용될 수 있다. 또한, 통합된 개인 온톨로지들을 사용하여 다양한 사용자들이 참여할 수 있는 협력적인 시맨틱 어노테이션 환경을 구성할 수 있다.

## 참고문헌

- [1] Berners-Lee, T., J. Hendler and O. Lassila “The Semantic Web”, In Scientific American, May, Vol. 284, No.5, pp. 34-43, 2001.
- [2] Klyne G., J. J. Carroll and B. McBride, “Resource Description Framework (RDF): Concepts and Abstract Syntax”, W3C Recommendation , <http://www.w3.org/TR/rdf-concepts/>, February, 2004.
- [3] Brickley D., R.V. Guha and B. McBride, “RDF Vocabulary Description Language 1.0: RDF Schema”, W3C Recommendation, <http://www.w3.org/TR/rdf-schema/>, February, 2004.
- [4] Deborah L. McGuinness and F. Harmelen, “OWL Web Ontology Language Overview”, W3C Recommendation, <http://www.w3.org/TR/owl-features/>, February, 2004.
- [5] 광승진, “디지털 콘텐츠의 어노테이션에 관한 연구”, 한국문헌정보학회지, 제40권, 제4호, pp.267-286, 12월, 2006.
- [6] Kiryakov A., B. Popov, D. Ognyanoff, D. Manov, A. Kirilov and M. Goranov, “Semantic Annotation, Indexing, and Retrieval”, 2nd International Semantic Web Conference (ISWC2003), Florida, USA. LNAI Vol.2870, pp.484-499, 2003.
- [7] Euzenat J., “Eight questions about Semantic Web annotations”, IEEE Intelligent Systems, Vol.17, No.2, pp.55-62, 2002.
- [8] mozdev, “Annozilla”, <http://annozilla.mozdev.org/>, 2007.
- [9] Knowledge Media Institute, “MnM”, <http://kmi.open.ac.uk/projects/akt/MnM/>, 2004.
- [10] Braun M., K. Kuehn and L. Meyer, “OntoMat-annotizer”, <http://annotation.semanticweb.org/ontomat/index.html>, 2002.
- [11] PLUS Group, “Simple HTML Ontology Extensions”, <http://www.cs.umd.edu/projects/plus/SHOE>, 2002.
- [12] Ontotext Semantic Technology laboratory, “KIM Semantic Annotation”, <http://www.ontotext.com/kim/semanticannotation.html>, 2007.
- [13] Tudor Groza K. M., S. Handschuh and S. Decker, “SALT - Semantically Annotated LaTeX for scientific publications”. In 4th European Semantic Web Conference, pp.518-532, 2007.
- [14] Henrik Eriksson, “An Annotation Tool for Semantic Documents”. Lecture Notes In ComputerScience, Springer-Verlag, Berlin, Heidelberg, Vol.4519, pp.759-768, 2007.
- [15] Alex Ball, “Briefing Paper: The Adobe eXtensible Metadata Plaform(XMP)”, UKOLN research organization, February, 2007.
- [16] Abobe, “XMP Specification”, sSeptember, [http://www.adobe.com/devnet/xmp/pdfs/xmp\\_specificati on.pdf](http://www.adobe.com/devnet/xmp/pdfs/xmp_specificati on.pdf), 2005.
- [17] Adobe, “PDF Reference Sixth Edition version 1.7”, [http://www.adobe.com/devnet/acrobat/pdfs/pdf\\_reference](http://www.adobe.com/devnet/acrobat/pdfs/pdf_reference)

\_1-7.pdf, November, 2006.

[18] Li Ding, Tim Finin, Anupam Joshi, Rong Pan, Scott R Cost, Yun Peng, Pavan Reddivari, Vishal Doshi and Joel Sachs, "Swoogle: a search and metadata engine for the semantic web", Proceedings of the thirteenth ACM conference on Informaiton and Knowledge management, pp.652-659, 2004

[20] Ahan K., J. Ojvunen K., Prud Hommeaux E. and Swickr R., "Annotea: An Open RDF Infrastructure for Shared Web Annotations". In The Tenth International World Wide Web Conference, Hong Kong, pp.623-632, May, 2001.

[21] Lowagie B., "iText, a Free Java-PDF Library", <http://www.lowagie.com/>, 2008.

[22] HP Labs Semantic Web Research, "Jena 2 - A Semantic Web Framework", <http://www.hpl.hp.com/semweb/jena2.htm>, 2008.

[23] Adobe, "Extensible Metadata Platform SDK", <http://www.adobe.com/devnet/xmp/sdk/eula.html>, 2008.

[24] Prud Hommeaux E. and Seaborne A., "SPARQL - Query Language for RDF", W3C Recommendation, <http://www.w3.org/TR/rdf-sparql-query/>, January, 2008.

[25] HP Labs Semantic Web Research, "<http://jena.sourceforge.net/ARQ/>", 2008.

[26] Sure Y., S. Bloehdorn, P. Haase, J. Hartmann and D. Oberle, "The SWRC Ontology - Semantic Web for Research Communities", In Carlos Bento, Amilcar Cardoso, Gael Dias, Proceedings of the 12th Portuguese Conference on Artificial Intelligence - Progress in Artificial Intelligence (EPIA 2005), Vol.3803, pp. 218-231, 2005.

[27] Annotea project, "Annotea Annotation Schema", <http://www.w3.org/2000/10/annotation-ns>, October, 2000.

[28] Dan Brickley, Libby Miller, "FOAF Vocabulary Specification 0.91", November, 2007.

[29] K. P. Yee, K. Swearingen, K. Li, and M. Hearst, "Faceted metadata for image search and browsing", CHI., pp.401-408, 2003.

[30] Oren E., Delbru R. and Decker S., "Extending Faceted Navigation for RDF Data", Lecture Notes in Computer Science, Springer Berlin, Heidelberg, pp.559-579, 2006.

**양 경 모(Kyoung-Mo Yang)**

[준회원]



- 2006년 2월 : 선문대학교 컴퓨터 정보학과 (이학사)
- 2008년 3월 ~ 현재 : 선문대학교 전자계산학과 (석사과정)

<관심분야>

데이터 마이닝, 시맨틱 웹, 온톨로지 공학, Formal Concept Analysis

**황 석 형(Suk-Hyung Hwang)**

[정회원]



- 1991년 8월 : 강원대학교 전자계산학과 조기졸업 (이학사)
- 1993년 4월 : 일본 오사카대학교 대학원 정보공학과 (공학석사)
- 1997년 4월 : 일본 오사카대학교 대학원 정보공학과 (공학박사)
- 1997년 3월 ~ 현재 : 선문대학교 컴퓨터공학부 교수

- 2007년 1월 ~ 2008년 1월 : 아일랜드국립대학교 DERI 객원연구원

<관심분야>

소프트웨어공학, 객체지향, 온톨로지공학, 시맨틱 웹, Formal Concept Analysis, Semantic Web Mining

**최 성 희(Sung-Hee Choi)**

[정회원]



- 1977년 2월 : 서강대학교 수학과 졸업 (이학사)
- 1988년 8월 : 미국 펜실베이니아주립대학교 대학원 전자계산학과 (이학석사)
- 1994년 5월 : 미국 켄터키대학교 대학원 전자계산학과 (이학박사)
- 1995년 3월 ~ 현재 : 선문대학교 컴퓨터공학부 교수

<관심분야>

Computational Complexity, 수치해석학, 계산이론, 소프트웨어공학, Formal Concept Analysis