

웹 기반의 RDF 저작도구

이민희*, 이무훈*, 안병규*, 김동혁*, 백주현**,
고병오***, 조현규****, 송병렬****, 최의인*

RDF Authoring Tool based on the Web

Min-Hee Lee, Moo-Hoon Lee, Byung-Gyu Ahn,
Dong-Hyuk Kim, Joo-Hyun Baek, Byung-Oh Goh,
Hyun-Kyu Cho, Byoung-Youl Song, Eui-In Choi

요 약

웹의 보편화로 인하여 수많은 데이터가 산출되고 있으나 데이터 관리를 위한 메타데이터 표준은 특정 도메인에 제한적이고 너무 광범위하게 정의되어 있어 데이터가 비효율적으로 관리되고 있는 실정이다. 이를 해결하기 위해 RDF를 활용하여 자원간의 관계를 정의하고 있으나, RDF의 특성에 적합한 편집도구가 전무한 상태이다. 따라서 본 논문에서는 RDF 저작도구를 설계 및 구현함으로써 RDF를 기반으로 메타데이터의 유효성 검증을 수행하고, 기술된 자원간의 관계를 가시적으로 분석할 수 있는 노드 앤 아크 다이어그램을 지원하며, 웹 자원을 N-triple로 표현하여 데이터의 효율적인 관리 및 분석을 가능하게 하였다.

Key Word : Metadata, Semantic Web, RDF

본 연구는 한국전자통신연구원의 "ebXML 메타정보 저작도구 개발"의 연구 결과임

* 한남대학교 컴퓨터공학과

** 국방과학연구원

*** 공주교육대학교 컴퓨터교육과

**** 한국전자통신연구원

1. 서론

웹 사용이 보편화되면서 다양한 유형의 서비스가 웹을 통해 지원되고 있으며 대량의 데이터가 지속적으로 생성되고 있다. 이에 따라 웹 상의 데이터를 효율적으로 관리할 수 있는 기술의 필요성이 크게 요구되고 있다. 그러나 데이터 관리를 위해 기존에 사용되고 있는 데이터베이스의 경우 특정 도메인에 의존적으로 스키마를 미리 정의하여 사용하여 웹 상의 다양한 데이터를 수용하는데 필요한 유연성을 제공하지 못하고 있다. 또한 웹과 데이터 관리를 위해 일반적으로 사용되고 있는 관계형 데이터베이스의 경우 속성들을 정하고 관계를 생성하고 있지만 데이터 자체에 대한 정보를 가지지 못하기 때문에 데이터에 대한 의미해석은 불가능하다. 따라서 이와 같은 문제를 해결하기 위해 웹 상의 자원을 효율적으로 관리할 수 있는 메타데이터의 표준으로 RDF를 제안하였다. RDF는 웹 분야에서 대두되고 있는 시맨틱 웹(Semantic Web)의 기반으로, 기존의 Dublin Core와 W3C Framework가 가졌던 한계를 해결하고, 기계가 자원의 의미를 이해하고 관리할 수 있는 메타데이터의 표준이다[1,2,3,4,7].

RDF는 특정 도메인에 종속적이지 않기 때문에 활용 범위에 제한이 없다. 그러나 RDF를 이용하여 데이터를 관리한다고 해도 RDF를 기반으로 작성된 메타데이터 문서가 유효하지 않다면, 효율적인 데이터 관리가 불가능하다. 이는 데이터 문서가 유효하지 않을 경우 웹 상에서 데이터를 효율적으로 검색할 수 없게 된다. 따라서 RDF 문서가 유효한지에 대한 유효성 검사를 수행하여 유효성을 보장할 수 있는 저작도구의 개발이 절실히 필요하다.

본 논문에서는 RDF와 RDFS를 기반으로 RDF 문서를 텍스트, 트리, 그래프 등의 형태로 저장할 수 있는 편집기, 유효성 검증이 가능한 검증기, 자원의 관계를 Triple 형태로 생성하는 N-Triple 생성기, N-Triple을 기반으로 자원간의 관계를 그래프 형태로 출력할 수 있는 Graph Viewer, 그래프를 SVG(Scalable Vector Graphics)로 변환한 후 저장하는 SVG 변환기를 설계 및 구현하여 다양한 편집 환경을 지원할 수 있게 하였다.

2. 관련 연구

현재 여러 기관들에 의해 RDF 저작도구 개발을 위한 프로젝트가 진행되어 왔다. RDF 저작도구로 대표적인 것으로는 IsaViz와 RDFedit가 있다. 각각의 저작도구는 기본적으로 RDF 문서의 저작을 목적으로 하지만 구현 환경, 적용기술, 인터페이스에 있어서 차이점을 가지고 있다.

IsaViz는 W3C에서 Jena와 Xerces 파서를 이용하여 개발된 것으로, 그래프 형태로 RDF 문서를 브라우징하고 RDF 문서를 저작할 수 있기는 하나 노드 앤 아크(Node & Arc) 다이어그램 상에서의 컴포넌트 제어가 어렵고 텍스트 편집을 지원하지 못함으로 인하여 다양한 양식의 저작에 적합하지 않다는 단점을 가지고 있다.

RDFedit는 RDF/RSS 문서를 생성할 수 있는 저작도구로서 RDF 모델과 구문, RDF 스키마, Dublin Core, RSS를 지원하고 트리 구조에 기반하고 있으며 저작 중인 자원에 대한 링크를 지원한다. 그러나 RDF/XML 문서의 유효성 검사를 수행하지 못하고 자원들간의 관계를 보여줄 수 있는 인터페이스를 제공하지 못한다는 단점을 가지고 있다.

3. RDF 저작도구의 구성 및 기능

3.1 RDF 저작도구의 구현 환경

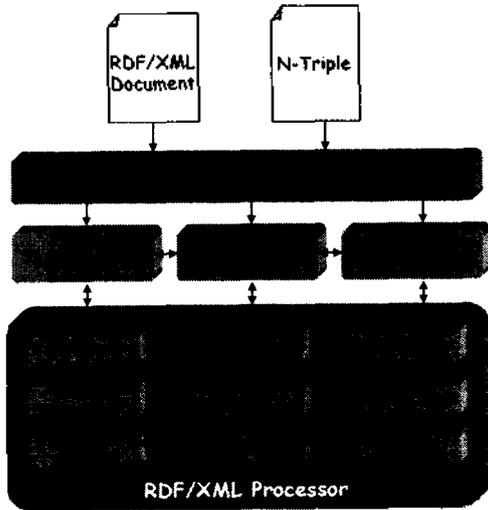
본 논문에서는 RDF 저작 도구의 구현환경은 Java 기반의 J2SE v1.4이며 파서(parser) 구현을 위한 API로 Xerces 2 Java Parser 2.0.2 Release와 JAXP(Java API for XML Processing) v1.2를 사용하였다[5]. RDF 모델을 처리하기 위해 Jena v1.4를 사용하였다[6].

3.2 RDF 저작도구의 구성

RDF 저작 도구는 파일 형태의 입력을 처리하기 위한 입력 인터페이스와 텍스트, 트리, 그래프 방식의 문서 저작 지원을 위한 3개의 인터페이스, RDF/XML 문서의 처리를 위한 RDF/XML 프로세서로 구성되어 있다.

<그림 1>과 같이 각각의 구성 요소들은 상호 관계를 가지고 있다. 입력 인터페이스는 텍스트 인터페이스, 트리 인터페이스, 그래프 인터페이스에 RDF/XML 문서를 전송하여 텍스트 인터페이스에서는 텍스트 노드로, 트리 인터페이스에서는 트리 노드로, 그래프 인터페이스에는 노드와 아크를 포함한 그래프 노드로 각각 변환하여 자원을 표시

할 수 있도록 한다.



<그림 1> RDF 저작도구의 구성도

텍스트 인터페이스는 보편적으로 지원하는 인터페이스로서 본 논문에서도 텍스트 편집 환경을 지원하기 위해 사용된다. 이 인터페이스는 트리 인터페이스 및 그래프 인터페이스로 상호연동이 가능하다.

트리 인터페이스는 RDF/XML 구문을 ELEMENT, ATTRIBUTE, CDATA, COMMENT, DOCTYPE, ENTITY, ENTITYREF, TEXT로 구분하고 구분된 각각의 노드를 관계에 따라 트리 구조로 표현하는 인터페이스이다. 이 인터페이스 역시 텍스트 인터페이스 및 그래프 인터페이스로 상호연동이 가능하다.

그래프 인터페이스는 N-Triple 형식의 입력을 받아 자원과 속성값은 노드로, 속성 타입은 아크로 표현할 수 있는 인터페이스이다. 이 인터페이스는 SVG Exporter, N-Triple Generator와 내부적으로 연동되어 저작 중인 그래프를 각 포맷에 맞게 출력할 수 있다. 이 인터페이스는 텍스트 인터페이스 및 트리 인터페이스로 상호연동 가능하다.

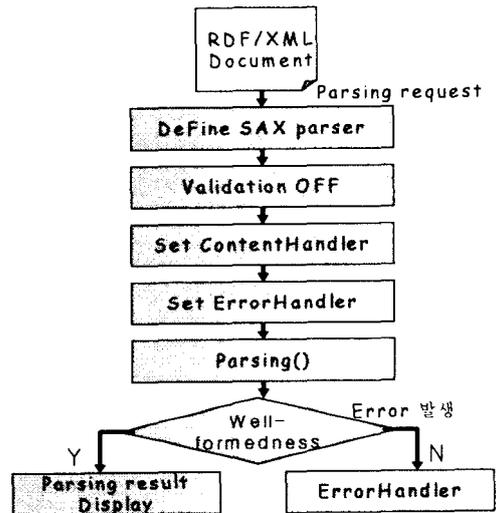
RDF/XML Processor는 RDF/XML 문서를 처리하기 위한 처리 모듈이다. 이 모듈은 RDF 문서의 파싱과 N-Triple 생성을 위한 ARP Parser, RDF/XML 문서의 파싱을 수행하는 XML Parser, RDF/XML 문서의 유효성 검사를 수행하는 XML Validator, RDF 문서에 기술된 메타데이터를 N-Triple 형태로 구성해 주는 N-Triple 생성기, SVG 형식의 그림 파일로의 저장을

지원하는 SVG Exporter, N-Triple 구조의 텍스트 파일로 RDF N-Triple을 저장할 수 있는 N-Triple Exporter를 포함한다.

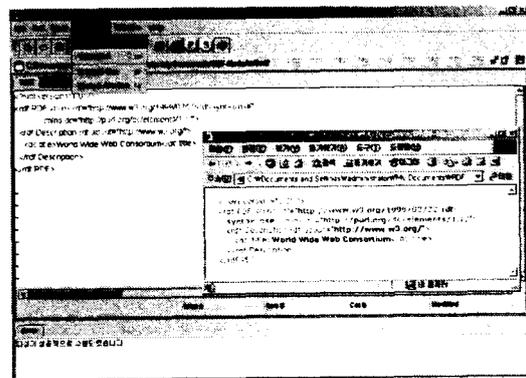
4. RDF 저작도구의 주요 처리모듈

4.1 파싱 모듈

파싱은 RDF 문서가 문법에 맞게 구성되어 있는지를 검사하는 처리 과정을 의미한다. 즉, RDF의 모든 표현 형식을 제공하기 위해서는 RDF 전용 파서를 통해 파싱 과정을 거쳐야 한다. 본 논문에서는 다양한 RDF 문서의 편집과 웹 상에서 직접 사용 가능한 RDF 문서의 생성을 위해 RDF 파서의 구현이 선행된다. RDF 파서는 RDF 문서의 Well-formedness를 검사하도록 하고, DTD 및 Schema에 따른 유효성 검증기를 통해 수행한다. RDF 파서의 파싱 루틴은 <그림 2>와 같다.



<그림 2> 파싱 루틴의 흐름도

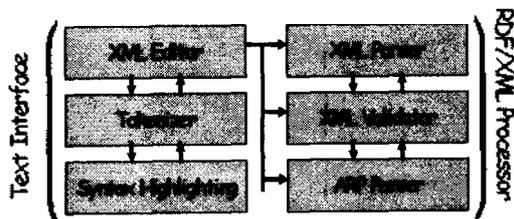


<그림 3> RDF 파서의 실행 화면

파싱 루틴은 텍스트와 트리 인터페이스를 통해 편집되거나 편집 중인 RDF 문서에 대해 파싱을 지원한다. SAX를 기반으로 하여 여러 문서에 대해 빠르게 처리할 수 있도록 구성되어 있으며 텍스트 인터페이스에 있는 RDF 문서에 의해 파싱 요청이 왔을 때 문자열을 읽어 들여 Well-formedness를 검사한다.

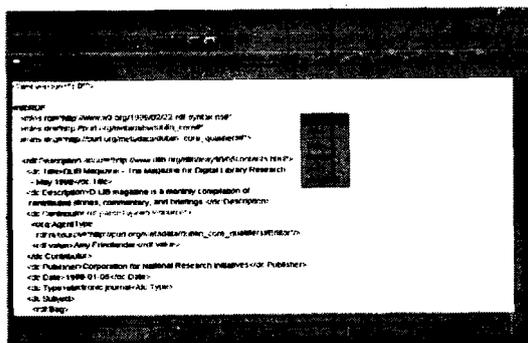
4.2 텍스트 기반 편집 모듈

기존의 메타데이터를 생성하는데 있어서 가장 보편적인 방법이 텍스트 기반의 메타데이터 편집 방식으로 본 연구는 기존 메타데이터 생성 방식을 수용하여 텍스트 기반의 메타데이터 편집기를 구현했다. 대부분의 저작도구들은 오직 한 가지 방식의 인터페이스만을 채택하여 사용하기 때문에 다른 유형의 편집 방식과 연동할 수 없어 전체적인 RDF 자원의 분석이 비효율적으로 처리되는 단점을 해결하기 위하여 본 논문에서는 그래프 기반 방식과 트리 기반 방식으로 쉽게 전환하여 메타데이터를 처리하기 위한 방법을 설계 및 구현하였다.



<그림 4> 텍스트 기반 편집 모듈의 구성

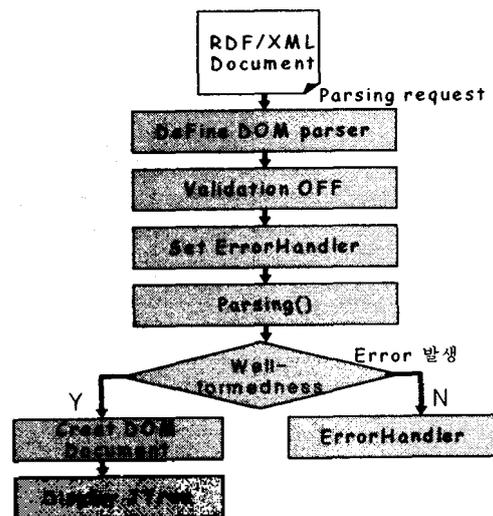
텍스트 인터페이스는 RDF/XML Processor의 XML Parser, XML Validator, ARP Parser와 연동되어 저장하는 RDF/XML 문서에 대한 파싱, 유효성 검사, N-Triple 생성을 처리한다.



<그림 5> 텍스트 기반 편집 모듈의 실행 화면

4.3 트리 기반 편집 모듈

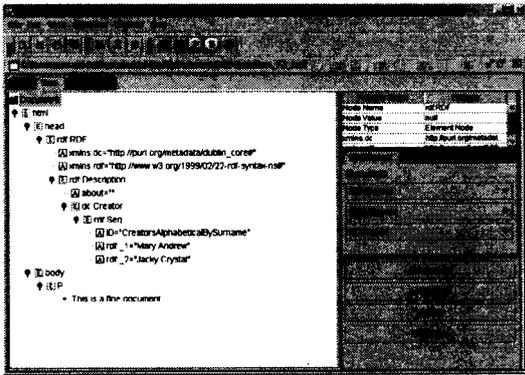
계층 구조 안에 정보를 그룹화해 놓을 수 있는 트리 구조는 항목 간 서로 관련되어 있는 객체간의 부모/자식(parent/child) 관계나 형제/형제(sibling/sibling) 관계를 표현할 수 있다. 본 논문에서는 RDF에서 제안한 데이터 모형과 동일한 의미를 가지면서 RDF 메타데이터를 편집할 수 있는 또 다른 유형의 편집 방법을 적용한다.



<그림 6> DOM을 이용한 트리 기반 편집 모듈의 흐름도

트리 기반 메타데이터 편집기에서는 SAX 파서에 비해 처리 속도는 좀 느리지만 RDF/XML 문서를 수정 및 변경할 때 용이한 DOM(Document Object Model)을 사용하여 파싱한다. DOM을 이용해 만들어진 DOM Document를 이용하여 트리 형태의 인터페이스로 변환하여 사용자에게 문서상의 내포관계를 구조적으로 보여줄 수 있도록 구현했다. 추가적으로 RDF와 RDFS, 그리고 RDF에 삽입되어 사용되는 더블린 코어(Dublin Core), RSS(RDF Site Summary)의 Element Set을 트리 기반 편집기에서 노드를 추가하는 형태로 제공해 줌으로써 사용자에게 보다 직관적이고 효율적으로 RDF 문서를 작성하는 기능을 제공한다.

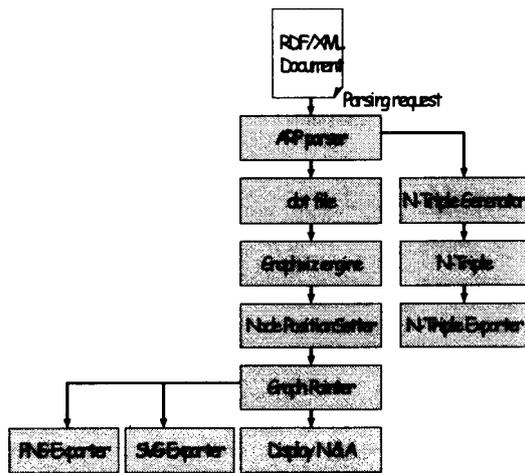
<그림 7>은 RDF 문서를 텍스트 기반의 편집 화면에서 트리 기반의 메타데이터 편집 인터페이스를 전환하여 RDF 문서를 트리의 형태로 편집 가능하게 변환해주는 실행 화면을 나타낸다.



<그림 7> 트리 기반 편집 모듈의 실행화면

4.4 Grape 기반 편집 모듈

RDF는 가장 일반적인 XML 표현법 외에도 노드 앤 아크 다이어그램으로 표현할 수 있다. 노드 앤 아크 다이어그램에서 타원형으로 표현되는 노드가 자원을 표현하는 것이고, 아크는 이름이 부여된 특성을 표현한 것이다. 문자열 리터럴을 표현한 노드는 직사각형으로 표현된다. 노드 앤 아크 다이어그램은 N-Triple의 Subject, Predicate, Object(Literal)를 구성요소로 가진다. 이러한 노드 앤 아크 다이어그램은 자원간의 관계를 분석하는데 용이하게 사용할 수 있다. 본 연구에서는 RDF 문서를 분석할 수 있는 노드 앤 아크 다이어그램을 그려주는 그래프 기반의 편집 모듈을 제공한다.

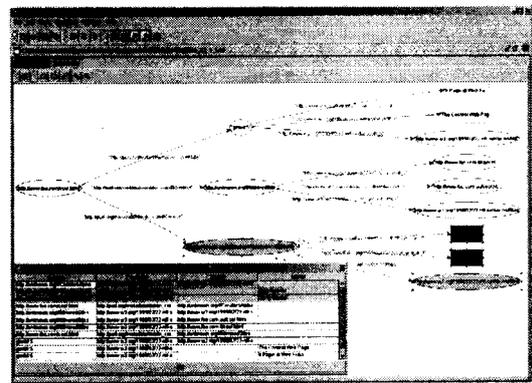


<그림 8> 그래프 기반 편집 모듈의 흐름도

그래프 기반의 메타데이터 편집 모듈을 제공하기 위해서는 먼저 파싱의 과정을 거쳐야 한다. RDF 문서를 노드 앤 아크 다이어그램으로 변환하기 위해서는 RDF 문서의

파싱 요청에 의해 ARP(Another RDF Parser) 파서가 RDF 문서 전체를 파싱하게 된다. ARP 파서의 파싱으로 생성된 dot 파일(dot file)은 N-Triple 형태의 Subject, Predicate, Object(Literal)를 가지며 노드에 대한 포지션(position) 값은 없는 상태이다.

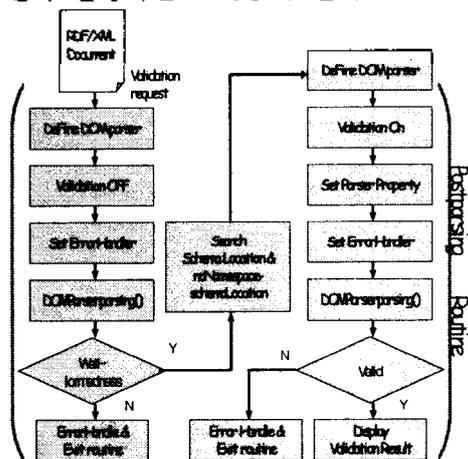
포지션 값은 Graphviz를 통해 노드 앤 아크 다이어그램의 각 노드의 포지션 값을 할당 받게 된다. 각각의 노드에 대한 포지션 값이 할당되면 Graph Painter가 RDF 노드 앤 아크 다이어그램을 그려준다.



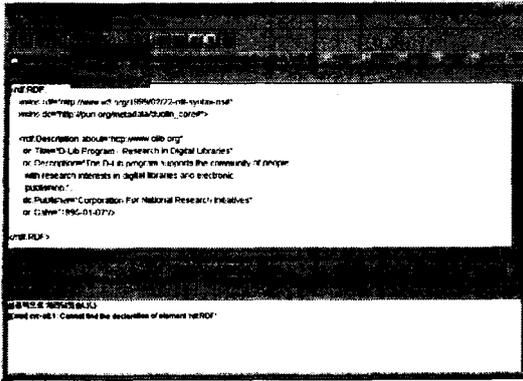
<그림 9> 그래프 기반 편집 모듈의 실행화면

4.5 유효성 검증 모듈

유효성 검증 모듈은 파서가 수행했던 RDF/XML 문서의 Well-formedness는 물론이고 문서에 대한 DTD와 스키마에 기술된 구문 정의대로 문서가 작성되어 있는지에 대한 유효성을 검사할 수 있는 구문 분석기를 의미한다. 문서의 구문 구조와 유효성 검사를 위해서 이를 처리하기 위한 처리 루틴과 실행화면은 다음과 같다.



<그림 10> 유효성 검증 루틴의 흐름도



<그림 11> 유효성 검증 모듈의 실행화면

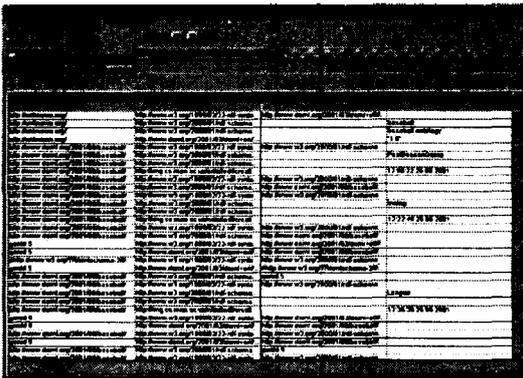
4.6 N-Triple 생성 모듈

웹 상의 복잡한 메타데이터를 생성하고 편집하기 위해서는 구체화 과정이 필요하고 이 과정을 위해서 RDF는 Triple을 사용하고 있다. Triple은 object-attribute-value의 구조를 가지며 A(O,V)와 같이 표현할 수 있다.

본 논문에서 웹 상에서 사용 가능한 메타데이터를 분석하고, 생성 및 편집을 위해서 Triple 기능을 구현하였다.

N-Triple 생성기를 통해 처리된 결과는 <그림12>와 같다.

ARP 파서를 통해 파싱된 RDF 문서를 N-Triple 생성 모듈을 통해 Subject, Predicate, Object(Literal)에 따라 N-Triple 테이블에 저장한다.



<그림 12> N-Triple 생성기 실행 화면

5. 결론 및 향후 과제

본 연구에서는 RDF/XML 문서의 저작 환경으로 텍스트, 트리, 그래프 기반 편집을 지원하고 있다. 다양한 저작 환경을 통해 다양한 유형의 웹 자원을 쉽고 빠르게 저작

할 수 있도록 하였다. 이 저작도구는 노드 앤 아크 다이어그램을 통해 그래프 형태로 문서를 저작할 수 있고 저작된 다이어그램을 SVG 이미지로 변환하여 관리할 수 있으므로 다양한 도메인에서 사용될 수 있다는 장점을 가지고 있다. 한편, Application 형태로 구현된 RDF 저작도구를 Servlet과 Applet으로 변환하여 웹 상에서도 RDF/XML 문서의 저작이 가능하도록 하는 추후연구가 필요하다.

참고문헌

- [1] 김홍기, 지식 컴포넌트 구축과 활용을 위한 시맨틱 웹(Semantic Web) 기술의 응용, 한국경영정보학회 추계학술대회 논문집, 2001.
- [2] 전자상거래 표준화 통합포럼, "시맨틱 웹 기술을 적용한 전자상거래 표준 운용체계", 2002.
- [3] W3C, Resource Description Framework(RDF) Model & Syntax Specification, 1999, <http://www.w3.org/RDF/TR/1999/REC-rdf-syntax-19990222/>.
- [4] RDF Vocabulary Description Language 1.0: RDF Schema, 2002, <http://www.w3.org/TR/2002/WD-rdf-schema-20020430/>.
- [5] Apache XML Project, August, 2002, <http://xml.apache.org/xerces2-j/>.
- [6] Hewlett-Packard, The jena semantic web toolkit, July, 2001, <http://www-uk.hpl.hp.com/people/bwm/rdf/jena/index.htm>.
- [7] Jeen Broekstra, Michel Klein, and Stefan Decker, Enabling knowledge representation on the Web by extending RDF Schema, In Proceedings of the 10th World Wide Web conference, pg. 467-478, Hong Kong, China, May 1-5, 2001.
- [8] Olivier Corby, Rose Dieng, and Cedric Hebert, A Conceptual Graph Model for W3C Resource Description Framework, ICCS2000, Darmstadt, Germany, August 14-18, 2000.

- 저자약력

2002년 한남대학교 컴퓨터공학과(공학사)

2002년 - 현재 한남대학교 대학원 컴퓨터공학과 석사과정

- 공동저자

이무훈 : 한남대학교 대학원 컴퓨터공학과 석사과정

안병규 : 한남대학교 대학원 컴퓨터공학과 석사과정

김동혁 : 한남대학교 대학원 컴퓨터공학과 박사과정

백주현 : 국방과학연구소 연구원

고병오 : 공주교육대학교 컴퓨터교육과 조교수

최의인 : 한남대학교 컴퓨터공학과 부교수

조현규 : 한국전자통신연구원 팀장

송병렬 : 한국전자통신연구원 연구원

- 관심분야

XML, RDF, Semantic Web

- 연락처

주소 : 대전광역시 대덕구 오정동 한남대학교 컴퓨터공학과 데이터베이스 연구실

우편번호 : 306-791

전화 : 042)629-7864

e-mail : miniyoo@dblab.hannam.ac.kr