

# 시맨틱 웹 기술 기반 정보서비스 시스템 *OntoFrame*<sup>®</sup>

## *OntoFrame*<sup>®</sup>: an Information Service System based on Semantic Web Technology

성원경 · 이승우 · 한선화 · 정한민 · 김평 · 이미경 · 박동인

Won-Kyung Sung, Seungwoo Lee, Sun-Hwa Hahn, Hanmin Jung, Pyung Kim, Mikyung Lee and Dong-In Park

한국과학기술정보연구원 정보서비스연구팀(ISRL)

### 요 약

시맨틱 웹 기술을 기반으로 한 정보서비스 시스템인 *OntoFrame*<sup>®</sup>은 연구자가 원하는 정보를 쉽고 정확하게 제공하는 학술정보 분석·융합 서비스 프레임워크를 목표로 한다. 본 시스템은 현재 학술정보를 지식 체계로 표현한 온톨로지와 이에 맞게 학술정보를 관리하고 가공하여 지식화하는 *OntoURI*<sup>®</sup>, 이 지식을 기반으로 추론을 수행하고 고속의 탐색 기능을 제공하는 *OntoReasoner*<sup>®</sup>로 구성되어 있다. 단순 검색 서비스를 제공하는 기존의 검색 시스템과는 달리, 본 시스템은 시맨틱 웹 기술을 바탕으로 학술정보에 대한 연도별 토픽 경향, 연관 토픽, 토픽별 연구자와 연구기관, 연구자 네트워크, 통계 정보, 지역적 분포 등과 같은 보다 의미적인 분석 서비스를 제공한다.

키워드 : 시맨틱 웹, 온톨로지, 추론, 정보서비스

### Abstract

As an information service system based on semantic web technology, *OntoFrame*<sup>®</sup> takes aim at a framework for providing analysis and fusion services of academic information. It currently consists of three parts: ontologies representing knowledge schema derived from academic information, *OntoURI*<sup>®</sup> which makes academic information into knowledge, and *OntoReasoner*<sup>®</sup> which performs inference and search on the knowledge. Unlike existing search engines which provides simple search services, our system provides, based on semantic web technology, several semantic and analytic services such as year-based topic trends in academic information, related topics, topic-based researchers and institutes, researcher network, statistics and regional distribution of academic information.

Key Words : Semantic web, Ontology, Inference, Information service

## 1. 서 론

미국 NSF<sup>1)</sup>의 보고에 따르면, 연구자들이 전체 연구 개발 시간의 절반 이상을 정보를 획득하는 데에 소비하고 있는 것으로 나타났다. 이는 정보 획득에 소요되는 시간을 줄이는 것이 얼마나 중요한지를 단적으로 보여주며, 그렇게 함으로써 연구자들이 순수한 연구 개발에 더 많은 시간을 투자할 수 있는 것이다. 이를 위해서는, 기존의 단순 검색 기능을 뛰어넘는, 연구자들이 원하는 다양한 정보를 제공할 수 있는 새로운 정보서비스가 필요하며, 본 논문에서 소개하고자 하는 *OntoFrame*<sup>®</sup>이 바로 그러한 서비스 시스템 중 하나이다. *OntoFrame*<sup>®</sup>은 시맨틱 웹 기술을 바탕으로 연구자가 원하는 정보를 쉽고 정확하게 제공하는 학술정보 분석 및 융합 서비스 프레임워크를 목표로 하고 있다. 이 시스템은 현재 학술정보를 지식 체계로 표현한 온톨로지와 이 온톨로지에 맞게 학술정보를 관리하고 가공하여 지식화하는 *OntoURI*<sup>®</sup>, 이 지식을 기반으로 추론을 수행하고 고속의 탐색 기능을 제공하는 *OntoReasoner*<sup>®</sup>로 구성되어 있다 (그림 1). 이들 각각의 구성요소에 대해 다음 장에서 차례로 설명하고자 한다.

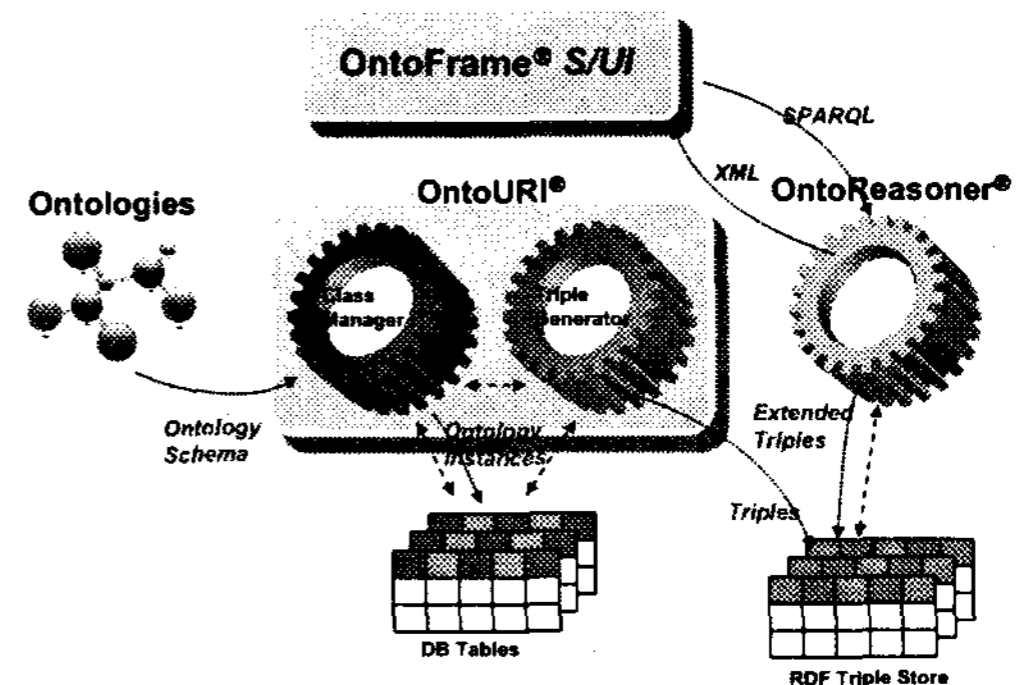


그림 1. *OntoFrame*<sup>®</sup> 전체 구조

## 2. 온톨로지

*OntoFrame*<sup>®</sup>은 세 가지 온톨로지를 기반으로 지식을 구축하고 있다. 하나는 "국가 과학기술 R&D 기반정보 온톨로지"[1]이고 나머지 둘은 시맨틱 포털 서비스를 위한 "응용 온톨로지"와 "개인화 온톨로지"이다. 기반정보 온톨로지는 연구자와 연구기관, 논문이나 저서, 특허, 보

1) <http://www.nsf.gov/>

고서 등의 성과물, 성과물의 주제, 기관이 위치한 지역 등의 학술 연구정보를 지식으로 모델링하고 있다. 연구 분야의 다른 온톨로지[2,3]와 달리, 이 온톨로지는 연구자와 연구자의 현재 소속기관 사이의 관계뿐만 아니라 특정 성과물을 저작할 당시의 소속기관과의 관계도 모델링하고 있는 것이 특징이다. 현재 26개의 클래스와 27개의 객체 속성, 83개의 데이터 타입 속성으로 구성되어 있다.

설계된 온톨로지에 맞게 온톨로지 인스턴스를 생성하는 역할은 *OntoURI*<sup>®</sup>가 수행한다. 기존의 레거시 DB 혹은 테이블 형태의 학술정보 데이터를 입력으로 받아 각 클래스에 속하는 인스턴스에 URI를 부여하여 RDF 트리플을 생성한다. 이 과정에서 동명이인 문제를 해결하기 위해 공저자 정보를 활용한다. 현재 CiteSeer OAI를 통해 2000년~2006년 사이의 논문 114,337건을 수집하여 6,378,572개의 RDF 트리플을 구축하고 있다.

### 3. 트리플 저장소와 추론

*OntoReasoner*<sup>®</sup>는 생성된 RDF 트리플을 저장하고 추론 및 탐색 기능을 수행한다. 대용량의 RDF 트리플을 효율적으로 저장하고 탐색할 수 있는 DBMS 기반의 트리플 저장소를 포함하고 있으며 RDFS 및 OWL 추론 규칙과 사용자 규칙을 통해 기존의 트리플로부터 새로운 트리플을 생성해내는 추론 기능도 포함하고 있다. 또한, 확장 클래스-속성 뷰(ECPV)를 도입하여 질의에 대한 빠른 응답을 제공하고 있다[4].

### 4. 서비스

*OntoFrame*<sup>®</sup>은 학술연구정보에 대한 의미적인 분석 및 융합 서비스를 제공하는 시맨틱 포털을 목표로, 기존의 학술정보검색 서비스와는 차원이 다른 서비스를 제공한다. 현재 Google Scholar<sup>2)</sup>, CiteSeer<sup>3)</sup>, YesKiSTi<sup>4)</sup>, NDSL<sup>5)</sup>, CSERI C<sup>6)</sup> 등 많은 학술연구정보 서비스가 활용되고 있지만, 이들은 모두 문자열 메타데이터와 키워드 기반의 검색 서비스에 의존함으로써 동명이인 해소나 식별체계 도입 등의 정교한 서비스 프레임워크를 갖추지 못하고 있다. 이에 비해, *OntoFrame*<sup>®</sup>은 시맨틱 웹 기술을 기반으로 URI 부여를 통해 개체를 식별하고 개별 정보를 의미적으로 융합하고 분석하여 연구자의 연구 개발 전주기를 지원하는 프레임워크를 갖추고 있다. 특히, 두 연구자 개체가 동일인으로 판단되면 owl:sameAs 관계를 이용하여 둘 중 어느 연구자를 통해서도 두 연구자에 관련된 통합된 정보를 제시할 수 있다. 또한, 본 서비스의 효용성과 사용성에 대한 평가를 지속적으로 수행하면서 개선해 가고 있다[5].

본 서비스는 메인 페이지와 몇 가지 개체 페이지로 구성되는데, 온톨로지에 설계된 토픽과 연구자, 논문 등이 개체에 해당한다. 사용자의 입력 질의에 따라 자동으로 해당하는 개체 페이지로 이동하게 되며 그 개체와 관련된 다양한 융합 및 분석 서비스가 제공된다(그림 2).

#### ● 학술연구토픽의 경향 분석

지정된 연도 구간 내에서 특정 연구 토픽과 관련된 토픽들의 경향을 분석하는 서비스로, 각 연구 성과물이 다루는 토픽을 연도별로 분석함으로써 얻어진다. Tag

Cloud 형식을 이용해 연구 경향을 한눈에 알아보기 쉽게 구성하였다.

- 토픽의 주요 연구자 파악  
특정 연구 토픽을 주로 연구하는 연구자들을 순위를 매겨 제시한다.
- 토픽의 주요 연구기관 파악  
특정 연구 토픽을 주로 연구하는 연구기관들을 순위를 매겨 제시한다. 이는 연구자가 성과물을 저작할 당시의 소속기관 정보를 활용하여 얻는다.
- 연구자 간의 연구자 네트워크  
연구자 사이의 공저자 관계와 인용/피인용 관계를 분석하여 특정 연도 구간 내에서의 연구자 네트워크를 구성한다.
- 토픽의 지역별 연구 분포 분석  
특정 연구 토픽을 수행하는 연구자나 연구기관, 지역의 분포를 제시하는 서비스로 구글맵과의 매쉬업을 통해 시각화한다.
- 학술연구정보의 통계 분석  
특정 연구 토픽이나 연구자의 연구 성과물에 대한 통계정보를 제공한다.

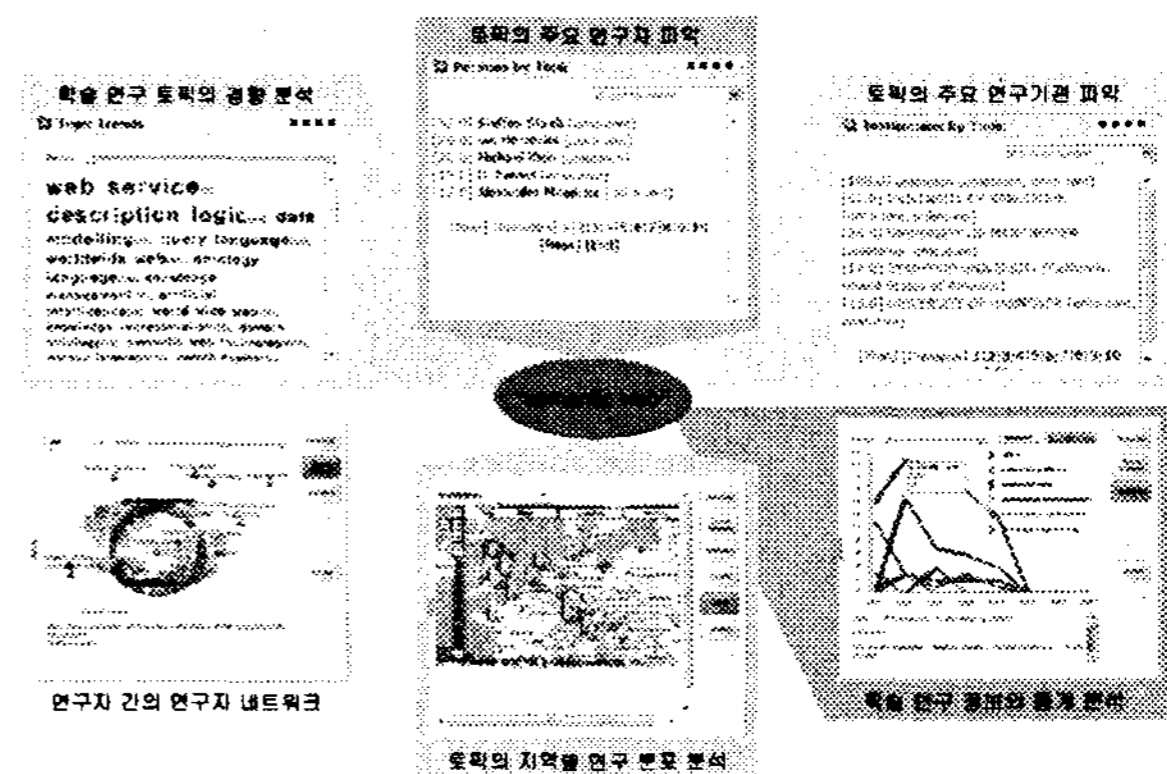


그림 2. *OntoFrame*<sup>®</sup>의 주요 서비스

그림 2는 “semantic web”이라는 토픽에 대한 이상의 서비스의 예를 보여준다. 이들 서비스 외에도 비슷한 연구를 수행하는 유사 연구자 분석이나 연구자의 상대적 연구 전문성 측정 등의 서비스들도 제공하고 있으며, 사용성 평가를 통해 새로운 서비스를 지속적으로 발굴해 나갈 예정이다.

### 참 고 문 헌

- [1] 강인수, 정한민, 이승우, 김평, 성원경, “국가과학기술 R&D 기반정보 온톨로지와 추론 모델링”, 2006 한국 컴퓨터융합학술대회 발표논문집, 2006.
- [2] V.R. Benjamins, D. Fensel, S. Decker and A. Gomez-Perez, “(KA)2: building ontologies for the internet: a mid term report”, International Journal of Human-Computer Studies, 51(3):687-712, 1999.
- [3] Y. Sure, S. Bloehdorn, P. Haase, J. Hartmann and D. Oberle, “The SWRC ontology - semantic web for research communities”, In Proceedings of the 12th Portuguese Conference on Artificial Intelligence, pp.218-231, 2005.
- [4] 이승우, 김평, 성원경, “확장 클래스-속성 뷰를 이용한 추론시스템의 응답 속도 개선”, 2007 한국콘텐츠학회 춘계융합학술대회 발표논문집, 2007.
- [5] 정한민, 이미경, 성원경, “반복적 사용성 평가를 통한 학술연구정보 서비스 구현”, 제34회 한국정보과학회 추계학술대회 발표논문집, 2007.

2) <http://scholar.google.com/>  
 3) <http://citeseer.ist.psu.edu/>  
 4) <http://www.yeskisti.net/>  
 5) <http://ndsl.or.kr/eng/newindex.html>  
 6) [http://cseric.cau.ac.kr/new\\_cseric/main.asp](http://cseric.cau.ac.kr/new_cseric/main.asp)