

## OntoFrame-K: 연구자 간 협업 지원 서비스를 위한 시멘틱 웹 기반 정보 유통 플랫폼

정한민<sup>○</sup> 이미경 성원경 박동인  
한국과학기술정보연구원 NTIS사업단  
{jhm<sup>○</sup>, jerryis, wksung, dipark}@kisti.re.kr

### OntoFrame-K: Semantic Web-based Information Dissemination Platform for Collaboration Support Service between Researchers

Hanmin Jung<sup>○</sup>, Mikyoung Lee, Won-Kyung Sung, Dong-In Park  
NTIS Division, KISTI

#### 요약

연구자 간 협업을 지원하기 위한 정보 유통 플랫폼은 통제 가능한 검증된 정보의 유통을 보장하여야 한다. 그렇지만, 기존의 정보 유통 관련 연구들은 어떠한 유형의 정보를 유통시킬 것인가, 또는 어떤 명세로 플랫폼을 설계할 것인가에만 국한되어 왔다. 본 연구에서는 기존 정보 유통 연구들이나 P2P 시스템들에서 다루지 못한 정보 검증과 추적에 초점을 맞추고 있으며, 연구자들을 위한 추론 서비스 제공까지도 가능한 플랫폼의 설계와 구현을 기술한다. 본 정보 유통 플랫폼을 통해 제공되는 서비스는 연구자들의 자발적 정보 교류를 지원하는 정보 유통 서비스와 단순 정보 검색이나 질의 응답으로 제공하기 힘든 지식 서비스로 구성된다. 지식 서비스는 다시 온톨로지를 이용한 지식화 과정과 시멘틱 웹 기반 추론 엔진을 이용한 추론 서비스로 나누어진다. 특히, 본 연구에서는 기존의 시멘틱 웹 기반 시스템들이 보여주지 못한 동적인 지식 추가와 이를 반영한 추론 과정을 연구자 간 협업 지원 서비스 시나리오에 맞추어 구현하였다. 국가 R&D 기반 정보 모델링을 위해 온톨로지를 구축하였으며, 한국과학기술정보연구원 내부 성과 정보 2,300여 건으로부터 얻어진 37,656 RDF Triple들을 이용하여 연구자 네트워크, 연구자 추적, 연구 맵 등 세 가지 추론 서비스를 제공할 수 있도록 하였다. 본 연구는 실제적인 시멘틱 웹 기반 정보 유통 플랫폼의 실현 가능성을 보여주었다는 데 그 중요성이 있다.

#### 1. 서론

현재와 같이 정보가 대량으로 무절제하게 생성되는 환경에서는 지식 관리 시스템(KMS; Knowledge Management System)이나 콘텐츠 관리 시스템(CMS; Contents Management System) 등을 통한 체계화된 정보 유통의 중요성이 크게 부각되고 있다. 그렇지만, 정보를 어떻게 지식화(Knowledgization)하고 검증할 것인가에 대한 연구는 정보 유통 이외의 분야, 특히 시멘틱 웹 관련 분야,에서만 독립적으로 이루어져 왔다. 이러한 연구 영역 간 분리는 결국 시멘틱 웹 기반 서비스들에서 정보의 생성, 저장, 관리, 검색, 유통의 동적 변화를 고려하지 못하게 만들었다. CAS, SEAL과 같은 대표적 서비스들도 배치 작업을 통해 생성된 정보들만을 디렉토리로서 동적 정보에 기반한 정보 유통 플랫폼에 바로 적용하기에는 어려움이 있다.

이에 본 연구는 정보의 흐름을 관리하는 정보 유통 플랫폼 상에서 지식화와 추론 서비스를 포함하는 지식 서비스를 제공할 수 있도록 시멘틱 웹 기술을 접목하고자 한다. 온톨로지를 통해 메타데이터를 지식화하고, 이들을 이용하여 연구자 네트워크, 연구자 추적, 연구 맵의 세 가지 추론 서비스를 제공하고, 정보 컨테이너(Information Container) 개념을 도입하여 검증된 정보만 유통될 수 있는 플랫폼 환경을 구축하고자 한다. 정보 유통과 시멘틱 웹의 결합은 결국 즉각적이고 동적인 지식화된 정보 교류를 가능하도록 하며, 본 연구에서 대상으로 삼고 있는 연구자 간 협업을 원활하게

지원할 수 있도록 한다.

#### 2. 기존연구

기존의 P2P 시스템을 포함하는 정보 유통 시스템들에서의 정보 검증은 문서에 대한 식별자(Identifier)를 부여하는 수준에 그치고 있다. [이원주 외 2005]는 미디어 콘텐츠 파일의 헤더 정보를 이용하여 ID를 부여하고 같은 파일에 대해 동일한 ID를 부여하는 방법을 사용하였다. 그렇지만, 사용자가 임의로 헤더 정보를 변경하는 경우와 문서 이력 추적이 필요한 경우에 대한 대처 방안을 가지고 있지 못하다. 또한, 온톨로지를 활용할 수 있는 추가적인 메타데이터를 고려하지 못함으로써 연구자 간 협업을 지원하기 어렵다. 이와 같이 사용자가 특정한 정보의 관리 및 이력 추적이 제대로 이루어지지 않는다면 정보 유통 플랫폼 상에서 유통되는 정보에 대한 신뢰를 떨어뜨리는 결과를 초래할 수 밖에 없다. [이혜주 외 2004]와 [최성원 외 2004] 역시 디지털 콘텐츠에 초점을 맞추어 다른 사용자에 의한 정보 갱신 가능성과 이전 정보와의 연계를 고려하지 못했다. 본 연구에서는 이러한 문제점들을 해결하고자 문서와 메타데이터 및 관리 정보를 포함하는 정보 컨테이너 개념을 도입하여 문서 접근 권한을 제어하고, 이력을 추적할 수 있도록 하여 문서 검증을 용이하게 한다. 또한, 연구자 간 협업에 필요한 메타데이터를 추가로 입력하도록 함으로써 온톨로지를 이용한 추론 서비스가 가능하도록 한다. [조성훈 외 2004]는 RDF 문서를 검증하는 방안을 제시하고 있다. RDF 파싱 루틴을 거쳐 DTD와 Schema에 대

한 유효성을 검증하도록 하고 있다. 우리는 RDF 유효성 검증 부담을 없애고자 사용자 입력 인터페이스를 온톨로지 기반으로 설계하여 메타데이터 입력 단계에서 RDF 유효성을 별도로 검증하지 않고도 자동적으로 보장될 수 있도록 한다.

현재까지의 시멘틱 웹 기반 서비스들은 국내의 경우 아직까지 실험용 데이터가 아닌 실세계 데이터를 구축한 예가 없으며, 해외의 경우에도 AKT (Advanced Knowledge Technologies)의 CAS (CS AKTive Space), SEAL (Semantic portAL) 등을 제외하고는 실세계에 성공적으로 적용되어 운영되는 사례를 찾아보기 힘들다. 시멘틱 웹 기반 서비스들은 정적 지식 서비스를 위한 구조로 설계되었다 [CAS] [SEAL] [순주찬 2004]. 지식은 언젠지 추가되거나 삭제될 수 있으므로 이를 반영한 지식화 프로세스를 설계하여야 한다. 본 연구에서는 실시간으로 추가되는 정보를 바로 지식화할 수 있도록 DB-to-OWL Conversion을 추가하고, 동시에 추후 배치 작업을 통한 전체 온톨로지 로딩도 일관성을 유지할 수 있도록 기존 온톨로지에 해당 인스턴트들을 Individual들로서 추가한다.

### 3. 정보유통 플랫폼 OntoFrame-K

연구자 간 협업 지원 서비스를 위한 정보 유통 플랫폼 OntoFrame-K는 인력, 보고서, 논문 등을 포함하는 국가 R&D 기반 정보 (Resource Information)를 이용하는 연구자들 사이의 자발적인 정보 교류를 보장할 수 있도록 중앙 서버를 중심으로 연구자 클라이언트들을 배치하는 방사형 구조를 가진다. 시멘틱 웹 기반의 플랫폼 구현을 통해 정보를 체계적으로 지식화하고 이를 이용한 추론 서비스를 제공함으로써 연구 협업을 위한 최적 환경을 구축할 수 있도록 한다. OntoFrame-K는 연구자 네트워크 (COP; Communities Of Practice), 연구자 추적 (Research Tracing), 연구 맵 (Research Map)의 세 가지 서비스를 통해 즉각적이고 정확한 연구 정보 획득을 지원한다. 궁극적으로 하나의 서버에 연결된 연구자 그룹을 하나의 가상 연구 커뮤니티로 정의하며, 다른 커뮤니티와의 서버 간 연결을 통해 타 분야 지식 획득을 원활하게 하는 방향으로 발전할 것이다.

### 4. 연구자 간 협업 지원 서비스

정보 유통 플랫폼에 기반한 연구자 간 협업 지원 서비스는 자발적 정보 유통 서비스 측면과 지식 서비스 측면으로 나누어 볼 수 있다. 지식 서비스는 다시 지식화와 추론 서비스로 나누어진다. 정보 유통 서비스는 연구자들의 자발적이고 즉각적인 정보 교류를 보장할 수 있는 체제 구축을 목표로 하며, 지식 서비스는 기존의 단순 정보 검색이나 질의 응답에서 처리하기 힘든 서비스 영역에서 추론 서비스를 제공하는 것을 목표로 한다. 정보 유통 서비스 역시 정보 제공 기능을 포함하는 데 키워드 기반 문서 검색, 이력 추적 (Version Tracing)을 통한 연관 문서 검색 등이 여기에 해당한다.

#### 4.1 정보유통 서비스

정보 유통 플랫폼 상에서의 정보 유통 서비스는 CS (Client-Server) 기반에서 이루어진다. 연구자 간의 자발적인 협업이 이루어지고, 정보를 조직화하고 서비스할 수 있기 위해서는 중앙 지식 서버를 유지해야 한다. 사용자 클라이언트는 필요한 정보를 중앙 지식 서버에 요청하고 다운로드받아 처리하고 생성되거나 변경된 정보를 업로드하는 일련의 정보 순환 과정을 거치면서 자발적인 정보 유통 서비스를 실현한다.

(1) Document Creation: 연구자는 논문, 보고서, 특허 등의 문서를 MS 오피스, 아래한글 등의 문서 편집기를 이용하여 생성한다.

(2) Document Upload: 생성된 문서 또는 정보 컨테이너를 중앙 지식 서버에 등록한다. 단순 문서 정보 (생성일, 파일명, 작성자 등)만을 작성할 수도 있으며, 내용에 대한 상세 정보를 미리 정의된 메타데이터 스키마에 맞추어 작성할 수도 있다. 후자의 경우에는 작성된 메타데이터가 지식화되어 추론 서비스를 통해 연구자들에게 제공된다.

(3) URI Assignment for Document: 단순 문서 정보의 입력인 경우에도 파일 객체에 대한 고유한 URI (Universal Resource Identifier)를 할당받아 문서 ID로 활용하거나 메타데이터와 연동시킨다.

(4) URI Assignment for Metadata: 상세 정보 입력의 경우에는 온톨로지서 클래스로 정의된 모든 필드들에 대해 URI 서버로부터 URI를 부여받는다. URI는 각 클래스를 구별할 수 있는 Prefix를 가져야 하며, URI 요청 시에 부여받고자 하는 클래스 정보를 매개변수로 제공한다. 4.2장의 (1)과 같으며, 이후의 과정은 4.2장을 참조한다.

(5) Document Packing/Validation: 기존의 P2P (Peer-to-Peer) 기반 시스템들은 사용자가 등록한 문서에 대한 검증 기능이 약하여 불법 자료의 유통을 막는 데 한계를 가지고 있다. 또한, 문서의 변형이나 복제를 억제하기 위한 장치를 가지고 있지 못하다. 연구자 간의 협업이 원활히 이루어지려면, 검증되지 않은 정보의 유통을 방지하고 국가 R&D 기반 정보에 대한 추적 및 보호가 이루어져야 한다. 중앙 지식 서버 내에 등록된 문서들에 대해 이러한 요구 사항을 만족시키기 위해 정보 컨테이너를 이용하여 문서를 패키징한다. 정보 컨테이너는 단순 문서 정보, 메타데이터를 포함하여 최초 작성 정보 (URI 형태의 문서 ID)와 바로 이전 작성 정보 등 관리 정보를 가지고 있다. 사용자가 정보 컨테이너 내의 문서를 편집하는 경우에 주어진 권한에 따라서만 접근할 수 있으며, 편집이 종료되는 순간에 자동적으로 패키징이 다시 이루어진다. 문서 검증 대상은 문서 또는 정보 컨테이너이며, 후자에 대해서는 메타데이터와 작성 정보들을 문서 저장소 (Document Repository) 내의 대응 정보와 비교하는 과정을 통해 문서 검증을 수행한다.

(6) Document Download: 연구자는 정보 검색이나 추론 서비스를 통해 원하는 정보를 찾아 다운로드한다. 정보는 그 연구자에게 정보 컨테이너 형태로 제공된다.

(7) Document Editing: 다운로드된 문서를 연구자가 편집하고자 하는 경우 정보 컨테이너는 사용자 권한에 맞게 메뉴들을 활성화시킨다.

#### 4.2 지식화

지식화 (Knowledgezation)는 국가 R&D 기반 정보의 메타데이터를 온톨로지를 이용하여 체계화시키는 과정으로 정의된다. 사용자에 의해 등록된 메타데이터는 최종적으로 RDF (Resource Description Framework) Triple로 변환되어 추론 서비스에 이용된다. 지식화 과정을 살펴보면 다음과 같다.

(1) URI Assignment for Metadata: URI 할당이 필요한 메타데이터 필드들을 위해 URI 서버를 호출하는 과정이며, 4.1장의 (4)와 같다.

(2) Metadata Storing: 작성된 메타데이터를 RDBMS 기반의 메타데이터 저장소 (Metadata Registry)에 저장한다. DB 접속 API를 이용하여 직접 메타데이터를 정해진 스키마에 맞도록 저장한다.

(3) DB-to-OWL Conversion: 메타데이터를 OWL (Web Ontology Language) 형식에 맞도록 변환하여 OWL 문서를 생성한다. 이 과정은 서비스 시나리오에 의존적이며, 온톨로지의 변경에 영향을 받을 수 있다. 또 다른 방법으로는 OWL 문서 생성 과정을 거치지

않고 직접 RDF Triple Store로 저장하는 DB-to-Triple Conversion 방법이 있다. 그렇지만, 이 경우는 RDF 파싱 (RDF Parsing) 기능을 변환기가 대신 수행해야 하므로 시나리오와 온톨로지에 대한 의존도가 더욱 커지게 된다. 실시간 메타데이터의 지식화를 위해서 이 과정은 필수적이며 전체 온톨로지의 변환이 필요없으므로 온톨로지 로딩 시간이 단축되어 추론 서비스의 부담을 감소시킨다. 기존의 시멘틱 웹 기반 시스템들과 달리 사용자가 등록하는 정보를 즉각적으로 반영할 수 있도록 함으로써 실용적인 시스템 구현을 현실화시킬 수 있다.

(4) RDF Parsing: Jena의 RDF/XML Parser (ARP)를 사용하여 OWL 기반의 온톨로지를 해석한다. 온톨로지를 이용하여 사용자 입력에 의해 실시간으로 추가된 인스턴스들을 파싱 (Parsing)한다.

(5) Persistent Model Creation: Jena는 관계형 데이터베이스 (RDBMS)에 저장된 RDF Triple들을 이용하여 추론을 하는데, 이러한 형태로 구성된 데이터를 Persistent Model이라 일컫는다. 이 과정을 통해 OWL 기반 온톨로지는 RDF Triple로 변환되어 데이터베이스에 저장된다.

(6) Instance Appending to OWL-based Ontology: 배치 작업을 수행하거나 온톨로지 최적화 또는 검증용 하기 위해 사용자에게 의해 입력된 인스턴스들을 국가 R&D 기반 정보 온톨로지 (Resource Information Ontology)에 추가한다. RDF Triple Store를 초기화하는 경우에는 전체 온톨로지로부터 배치 작업을 통해 RDF Triple들을 재생산한다. 국가 R&D 기반 정보 온톨로지와 연결된 (4), (5)는 배치 작업 시에 이루어진다.

### 4.3 추론 서비스

연구자 간 협업을 지원하기 위한 정보 유통 플랫폼과 중앙 지식 서버는 정보 검색을 통해 직관적으로 찾을 수 있는 정보들을 제공하는 한편, 다음의 추론 서비스를 통해 간단한 연산으로 찾기 힘든 고급 정보를 효율적으로 제공한다.

(1) RDQL Creation: 미리 정의된 시나리오에 맞추어 RDQL (RDF Data Query Language) 램플릿들이 존재하며, 사용자에게 의해 입력된 데이터는 RDQL 내의 특정 영역에 삽입된다.

(2) Thesaurus Loading: 지식화된 정보들과 사용자가 사용하는 질의어 (예. 주제 명, 소속 기관 명, 학회 명 등)간의 어휘적 불일치 (Lexical Disagreement)를 해소하기 위해 시소러스를 활용한다. 현재는 주제 명 (Topic Name)에 대해서만 적용하고 있으며, 시소러스 상의 특정 개념 노드와 매칭되는 경우 하위 개념 명들을 이용하여 주제 명을 확장한다. 시스템 로딩 시에 웹 서비스를 이용하여 시소러스를 호출하고 이를 RDF Triples로 변환한다. 시소러스를 이용하여 확장된 주제 명들은 (3)에서 예시한 규칙에 의해 추론에 이용된다.

(3) Inference: Jena는 RDF Triple들과 추론 규칙들 (Inference Rules)을 이용하여 추론을 수행하며, 그 결과로 인스턴스들이나 Triple들을 출력한다.

(4) Inference Result Generation: 이 과정은 추론 결과를 시각화용 입력 데이터로 만드는 과정이며, 시나리오에 의존적인 규칙들 (예. 전문가 순위화 등)을 적용하게 된다. 또한, 시각화에 필요한 각종 통계 정보들을 생성하는 과정도 포함한다.

(5) Inference Result Visualization: 서비스 시나리오에 따른 추론 결과를 최적으로 연구자들에게 제공하기 위해 시각화 모듈을 통해 정보를 가공하여 보여준다. 플래시 (Flash) 등을 이용하여 다양한

시각적 효과를 보여줌으로써 연관 정보를 보다 쉽게 찾을 수 있도록 도와준다.

### 5. 실험

국가 R&D 기반 정보를 위해 행사 (Event), 조직 (Group), 지적 재산권 (Ownership), 과학 기술 인력 (Person), 과제 (Project), 출판물 (Publication), 연구 분야 (ResearchTopic)의 최상위 7개 클래스들을 포함하여 74 개 클래스들을 구축하였으며, 이들에 대해 75개 Property들을 정의하고 someValuesFrom, allValuesFrom, minCardinality 등을 포함한 제약들을 기술하였다. 한국과학기술정보연구원 내부 성과 정보 2,300 여건은 Jena ARP에 의해 37,656 RDF Triple들로 변환되어 추론을 위해 사용되었다. 범용 과학 기술 분야 시소러스는 전문 용어 생명 주기 기법으로부터 추출, 선정된 15,000 여 용어들을 포함하고 있다 [정한민 외 2005]. 연구자 네트워크, 연구자 추적, 연구 맵 세 가지 추론 서비스 시나리오에 기반하여 지식 서비스를 제공하고 있으며, 웹 기반 정보 유통 서비스를 동시에 제공하고 있다.

### 6. 결론

본 연구를 통해 연구자 간 협업 지원 서비스를 위한 정보 유통 플랫폼을 설계하고 구현하였다. 자발적이고 즉각적인 정보 교류를 위해 실시간 정보를 지식화할 수 있도록 시멘틱 웹 기반 정보 유통 플랫폼을 구성하였으며, 유통되는 정보를 보장할 수 있도록 정보 컨테이너 상에서의 권한별 접근 및 검증과 이력 추적이 가능하도록 하였다. 특히, 시멘틱 웹 기술을 이용한 의미 기반 온톨로지 추론 서비스를 통해 지능적 정보 획득이 가능해짐으로써 새로운 정보 유통 패러다임의 기반을 마련하였다고 할 수 있다. 향후 본 연구는 외부 성과 정보로 그 대상을 확대하고, 정보 유통 서비스와 지식 서비스 간의 상호 연동을 통한 지식 기반 융합 서비스를 제공하는 것을 목표로 하고 있다.

### 참고문헌

손주찬, 2004. 지능형 비즈니스 웹 플랫폼 기술. 제2회 지식정보처리와 온톨로지 워크숍.  
이원주, 김승연, 전창호, 2005. P2P 환경에 유효한 허위 데이터 감속 알고리즘. 한국컴퓨터종합학술대회.  
이혜주, 최병석, 홍진우, 석종원, 2004. 디지털 방송 콘텐츠 보호 유통 시스템 설계 및 구현. 한국정보처리학회논문지 C, 제11C권, 제6호.  
정한민, 구희관, 이병희, 성원경, 2005. 효율적인 자원 운영을 위한 전문용어 생명주기 관리 연구. 한국컴퓨터종합학술대회.  
조성훈, 송병열, 조현규, 최의인, 2004. 메타데이터의 생성 및 관리를 위한 RDF 문서 검증기와 N-Triple 생성기. 정보처리학회논문지 B, 제11-B권, 제5호.  
최성원, 이상환, 이상기, 조성남, 석종호, 2004. 식별체계와 공개기반의 디지털 콘텐츠 유통모델 연구. 한국정보과학회 추계학술대회.  
CAS. <http://www.triplestore.aktors.org/demo/AKTiveSpace/>  
SEAL. [http://www.aifb.uni-karlsruhe.de/Projekte/viewProjektenglish?id\\_db=58](http://www.aifb.uni-karlsruhe.de/Projekte/viewProjektenglish?id_db=58)