

국가 과학기술 R&D 기반정보 온톨로지 구축을 위한 URI 관리 및 서비스 시스템 구현

구희관⁰ 정한민* 강인수* 성원경* 이승준** 심빈구**

과학기술연합대학원대학교 응용정보과학

한국과학기술정보연구원 NTIS사업단*

옴롬정보 지식기반구축사업부**

{hkoo⁰, jhm, dbaisk, wksung}@kisti.re.kr, {dino, bgshim}@orom.com

An Implementation of URI Management and Service System for the Construction of National R&D Reference Information Ontology

Hee-Kwan Koo⁰, Hanmin Jung*, In-Su Kang*, Won-kyung Sung*, Seung-june Yi** Bin-Gu Shim**

Practical Information Science, UST

NTIS Division, KISTI*

Knowledge Base Development Department, OromInfo**

요 약

전체적인 관점으로 국가 과학기술 R&D를 조망하는 국가 과학기술 R&D 기반정보 온톨로지는 인력정보, 과제정보, 성과정보를 포함하며 여러 URI들과 연계된다. 본 논문에서는 기존의 단일 URI 체계가 아닌 시스템적이고 다층적인 국가 과학기술 R&D 기반정보 온톨로지의 인스턴스 레벨에 필요한 다중 URI 관리 및 서비스 시스템을 구현한다. 본 시스템은 관리적인 측면에서 웹 인터페이스 기반 URI 등록과 조회를, 서비스 측면에서 URI 질의중심의 웹서비스를 제공한다. 기존 정보(KISTI의 KOI, 학술진흥재단의 부서 및 기관, 국가과학기술인력 종합정보시스템의 인력DB)와 연계를 기반으로 가용성 및 재사용성을 높이고 신규정보를 위한 URI 생성규칙을 제공함으로써 URI 관리 및 서비스 시스템 완성도를 높였다.

1. 서 론

인터넷 발달로 급속하게 증가하는 정보에 대한 URI 기반 식별 체계 요구와 이를 지원하는 다양한 서비스가 증가하고 있다. 지속적으로 정보를 관리해야 하는 단점을 지닌 기존 URL에 의한 정보접근방식에 비해 URI 기반 정보접근방식은 정보에 일관된 식별자 부여 및 관련된 각종 메타데이터를 운영하고 관리하는 장점을 지닌다.[1][2]

또한, 빠르게 발전하는 국가과학기술에 대응하기 위해서 시스템적이고 다층적인 접근을 하는 국가 과학기술 R&D 기반정보 온톨로지는 기존의 단일 URI가 아닌 다중 URI로 구성이 된다. 본 논문에서는 국가 과학기술 R&D 기반정보 온톨로지의 인스턴스 레벨에 필요한 다중 URI 관리 및 서비스 시스템을 구현한다. 이 URI 관리 및 서비스 시스템은 사용자에게 일관된 정보접근방식을 제공하여 정보 신뢰성을 향상 시킬 수 있다.

URI 관리 및 서비스 시스템은 웹인터페이스와 웹서비스를 제공하여 사용자 접근성을 증가시키고자 하였으며, URI로 구축된 정보자원은 국가 과학기술 R&D 기반정보 온톨로지 구축의 목적뿐만 아니라 외부 시스템에 대한 접근허용을 통한 정보 자원의 재사용성 증가도 기대된다.

2. 관련연구

이상환[2]은 과학기술분야 디지털콘텐츠에 대한 식별체계인 KOI를 개발하고, 이를 기반으로 신동구[1]는 이 식별체계와 다른 식별체계간의 연계시스템을 구축하였다. 이는 디지털콘텐츠에 대한 식별자를 대상으로 구성하는 단일 URI 시스템이며,

국가 과학기술 R&D 기반정보 온톨로지에 필요한 여러 URI의 조합적인 개별정보를 구성하지 못한다. 이는 국가과학기술인력 종합정보시스템의 인력DB도 마찬가지로 한계를 가지는 데, 국가 R&D를 전체적인 관점에서 조망하는 조합적인 단위정보로 구성할 수 없다.

Andreas Harth[3]는 분산되어 있는 온톨로지 인스턴스 레퍼지토리들을 대상으로 사용자 질의에 대한 결과를 생성할 때, URI 기반의 병합으로 응답질의를 생성하였다. 이런 접근방식은 URI를 체계적으로 관리하지 못하는 단점을 가지며, 이렇게 생성된 URI는 단지 응답데이터만을 생성하는 데만 이용하기 때문에 URI 재사용에 대한 방법은 제시하지 못했다.

3. 국가 과학기술 R&D 기반정보 온톨로지 URI

그림 1은 국가과학기술 R&D 기반정보 온톨로지를 구성하고 있는 인력정보, 과제정보, 성과정보와 성과정보에 포함되어 있는 논문정보, 보고서정보, 특허정보를 보여주며 객체 URI, 기관 URI, 부서 URI, 인력 URI, 출처 URI 및 과제 URI와의 관계를 표현한다. 성과정보 중 하나 인 보고서를 예로 들어 이들 URI의 구성을 설명하면 보고서는 그 보고서를 유발하게 하는 과제 URI, 보고서 자체의 원문에 대한 객체 URI, 그리고 보고서의 저자인 인력 URI들의 조합으로 구성된다.

객체 URI, 기관 URI, 부서 URI, 인력 URI들은 기존 구축된 정보와 호환성을 위해 객체 URI는 KISTI의 KOI와, 부서 URI 및 기관 URI는 학술진흥재단[5]의 부서 및 기관정보와, 인력 URI는 국가과학기술인력종합DB[4]와 연계할 수 있게 설계를 한다.

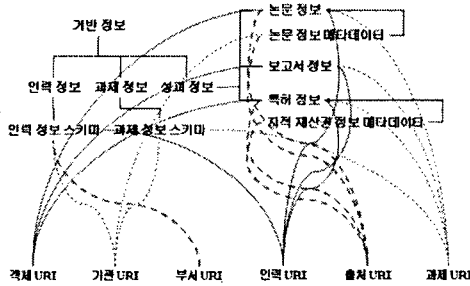


그림 1. 은론로지의 구성요소 및 URI 연관도

인력 URI를 예로 들어 URI의 형태에 대해 설명을 하면, URI는 "http://www.kisti.re.kr/isrl#PER:7010186243" 과 같이 네임스페이스와 자료유형과 일련번호로 구성된다.

객체 URI는 호환성을 고려해 KISTI의 KOI 생성 규칙을 간략화하여 사용하며, 생성형식은 "등록관리기관코드.자료유형코드.일련번호" 형태를 가지며 "등록관리기관코드"는 객체의 종류에 따라 학술지, 보고서, 개인공유자료, 기반정보 등으로 구분하며, "자료유형 코드"는 학술지, 학술대회, 학위논문, 특허, 연구보고서, 개인공유자료 등으로 구분한다. 마지막으로 "일련번호"는 7자리의 숫자를 사용한다.

기관 URI는 4~6자리의 영어 대문자와 숫자가 조합된 URI를 사용하며 기관을 추가하는 경우에는 6자리로 사용한다. 부서 URI는 4자리의 숫자로 된 URI를 사용하며 부서를 추가하는 경우에는 4자리를 유지하되, 첫 1자리를 "S"로 구성하고 나머지 3자리는 일련번호로 사용한다.

인력 URI는 국가과학기술인력종합정보시스템의 인력DB에 등록된 인력과 등록되지 않은 인력으로 구별하여 사용하며, 국가과학기술인력종합 DB에 등록된 인력은 10자리의 인력식별번호를 사용하고, 등록되지 않은 인력은 10자리를 사용하되, 첫 3자리를 "0"으로 설정하고 나머지 7자리에 대해서 일련번호를 부여한다.

출처 URI는 출처 식별 코드 3자리와 일련번호 6자리로 구성하며 출처 식별 코드는 논문지의 경우 "SOJ", 학술대회의 경우 "SOP", 학위논문의 경우 "SOT", 지적재산권의 경우 "SOI", 보고서의 경우 출처를 과제로 한다. 일련번호는 각 출처 별로 1부터 시작한다.

표 1. URI 예제

URI 예제	
객체URI식별자	KISTI.SOI.000001 지식기반구축 플랫폼연구 KISTI.PTN.000001 지식기반구축 방법론 특허
기관URI식별자	8A5327 한국정보과학회 9P9048 한국과학기술정보연구원 000002 오름정보
부서URI식별자	5459 전기.전자.컴퓨터공학 S001 NTS사업단
인력URI식별자	6410136403 성원경 7010186243 정한민 000000001 구희권
출처URI식별자	SOP000001 2005년 정보과학회 추계학술대회 SOJ000001 정보과학회 논문지, 시스템 및 이론 SOT000003 과학기술연립대학원대학교 박사학위 SOT000001 과학기술연립대학원대학교 석사학위
과제URI식별자	Z-2006-008819 해외과학기술정보 수집분석활용사업 Z-2006-008830 슈미컴퓨팅 인프라 기술 연구

과제 URI는 영숫자 혼용 형태로 KISTI의 과제번호 부여 체계를 따라 작성되고 외부 과제에 대해서는 "Z-연도-일련번호"로 구성하며, 연도는 네자리, 일련번호는 6자리로 구성한다.

4. URI 관리 및 서비스 시스템

4.1 URI 관리 및 서비스 시스템 설계원칙

기본적으로 URI 관리 및 서비스시스템은 크게 관리적인 측면과 서비스측면으로 분리하여 설계한다. 관리적인 측면의 설계는 기존 URI 데이터의 입력 및 확인작업을 위한 웹인터페이스로 작성된 URI 등록과 URI 조회로 구현이 되며, 서비스 측면 설계는 다양한 클라이언트에게 서비스를 제공할 수 있는 웹서비스로 구현된다. URI 등록에는 기존 정보를 일차적으로 검색하여 그 일치 여부에 의해 URI 등록이 결정되며, URI 등록 과정 중에 URI 정보에 대한 오류경사 및 중복 처리 등을 위한 검증의 단계가 포함이 된다. URI 웹서비스 인터페이스는 이들 URI에 대한 기본적인 URI요청, 확장, 취소 및 형식의 유효성 검사의 4가지 기본적인 서비스를 제공하며, 다종의 URI를 제공하는 웹서비스들을 생성한다.

4.2 URI 관리 시스템

그림 2는 URI 등록 관련 클래스 구조도를 보여준다. 표 1은 그림 2의 각 클래스의 기능예시표를 보인다.

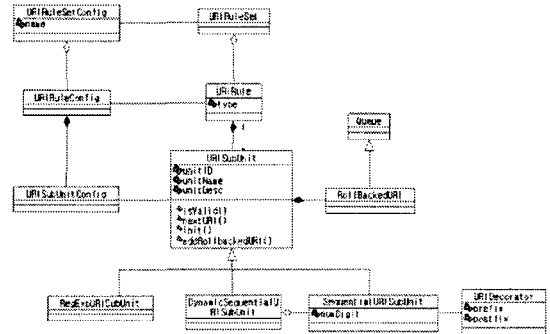


그림 2. URI 관리 시스템 클래스 구조도

표 2. URI 관리 시스템의 클래스 기능설명

클래스명	설명
URIRuleSetConfig	URIRule들의 집합에 대한 설정내용을 담는다.
URIRuleSet	URIRule들의 집합을 관리한다.
URIRuleConfig	특정 한 종류의 URI생성 용인 URIRule에 대한 설정내용을 담는다.
URIRule	특정 한 종류의 URI생성 룰을 관리한다.
URISubUnitConfig	URIRule의 하위 단에 대한 설정 내용을 담는다.
URISubUnit	URIRule의 하위 단 룰을 관리한다. 예를 들어, 인력 URI는 의미상으로는 한 종류이지만, 시스템적으로는, 인명사전에 존재되어 있는 사람과, 그렇지 않은 사람, 두 가지의 상세 룰이 존재한다. URISubUnit은 이런 상세 룰을 나타낸다.
RegExURISubUnit	정규식을 이용하여 URI를 검증하는 기능 제공.
SequentialURISubUnit	제한된 자리 수 내에서 시퀀스 값을 순차적으로 증가시켜 URI를 생성한다.
URIDecorator	SequentialURISubUnit이 만들어낸 시퀀스 값의 앞, 뒤에 미리 정의된 문자열을 붙여 붙여 맞는 URI를 만들어 낸다.
DynamicSequentialURISubUnit	SequentialURISubUnit을 이용하여, 여러 개의 시퀀스 값을 유지하면서, URI를 생성해 낸다. 예를 들어 과제 URI는 "Z[연도 4 자리][6 자리시퀀스]"와 같은 룰을 갖는다. 이때, prefix는 연도에 따라 바뀌게 된다.

URI 등록 시에 등록 요청, 서버응답, 사용자확인 의 3단계로 구성이 된다. 등록의 효율성을 증가시키기 위해 타임스탬프를 URI 등록요청 시에 함께 URI 관리 및 서비스시스템에서 생성한다. 한번 생성이 되고 나서 사용자확인 없이 정해진 시간을 넘은 URI는 생성대기 큐에 반납된다. 이를 통해 URI 값에 대한 관리성을 향상시킨다. 타임스탬프를 이용하여 생성대기 큐에 반납하는 시간은 URI 생성요구에 대해 대응이 가능하기 위해 변경이 가능하다.

OntoURI-K

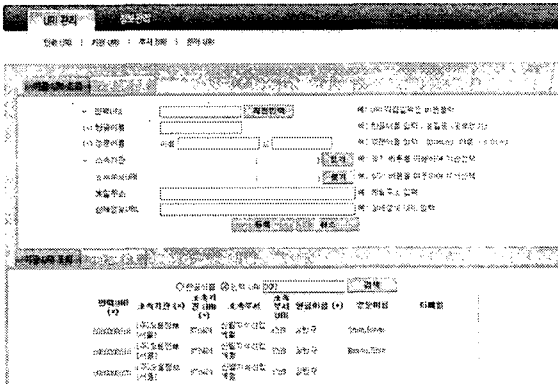


그림 3. 인력 URI 관리 화면 (조회 및 등록)

그림 3은 인력 URI 관리 화면을 보여준다. 입력정보 URI 등록 화면은 입력정보 URI를 직접 입력할 수 있는 인력 URI 필드와 한글이름필드, 영문이름필드, 소속기관 URI 필드, 소속부서 URI필드, 메일주소필드, 상세정보 URL의 입력필드로 구성된다. 등록화면 하단에는 기존에 입력된 입력정보를 리스트형태로 보여준다.

또한, 인명 등록 시에 소속기관 URI를 필수적으로 참조하도록 함으로써 데이터의 신뢰성을 증가시키고자 한다. 그리고, 생성규칙에 의거해 인력 URI를 자동생성 하거나 직접입력의 경우에도 생성규칙을 이용해 검증하도록 한다.

URI 관리 시스템은 직접적인 자료의 확인 및 입력은 웹인터페이스에서 동작한다. 인력 URI 관리화면을 예로 보여 설명했지만 그 외 다른 URI 관리 시스템의 서버메뉴도 등록 및 조회로 구성되어 있다.

4.3 URI 서비스 시스템

표 3. URI 웹서비스 이름 및 대상 URI

이름	대상 URI	영세
ArtCom_Service	객체 URI	requestURI,commitURI,
OrgURI_Service	기관 URI	rollbackURI,validate
DepURI_Service	부서 URI	URI를 4가지 기본 연
HumURI_Service	인력 URI	산 제공 및 조회결과
ConCom_Service	출처 URI	및 상세 결과조회 반
ProURI_Service	과제 URI	환의 세부 서비스로
		나누어져 구성.

표 3은 그림 1에서 보인 6개의 URI에 대해 각각의 웹서비스 명과 대상 URI를 보인다. 표 3의 영세에서 보여주는 4가지의 기본 요구에 대해 정의 하고 조회결과 및 상세결과에 관한 두 개의 연산을 제공하는 서버 연산으로 구성이 된다.

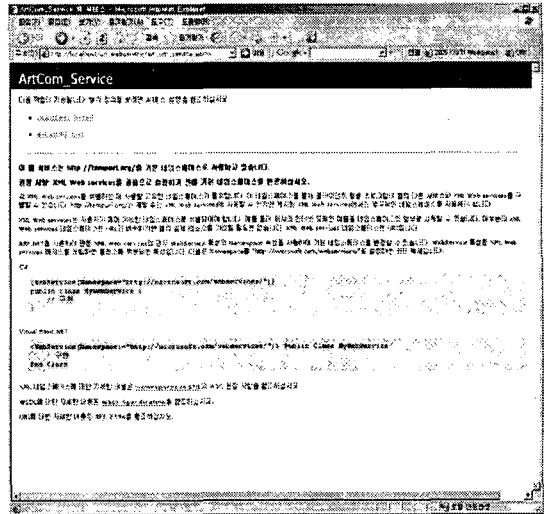


그림 4. 웹서비스 화면

그림 4는 URI 관리 및 서비스 시스템의 웹서비스 중 하나인 ArtCom_Service의 구동화면을 보여 주고 있다. 이 웹서비스는 객체 URI를 담당한다.

5. 결론

본 논문에서는 기존의 단일 URI 체계가 아닌 시스템적이고 다층적인 국가 과학기술 R&D 기반정보 온톨로지에 인스턴스 레벨에 필요한 URI들을 등록 및 조회로 관리하며, 이를 서비스하는 시스템을 설계하고 구현하였다. 관리적인 측면에서 웹인터페이스와 서비스측면에서 웹서비스의 2가지 인터페이스로 구축하여 재사용성을 높이고 가용성 및 유연성을 확보하였다. 향후 계획으로는 타 시스템과의 연계와 서비스의 제공범위를 넓히는 연구가 필요할 것이다.

참고문헌

- [1] 신동구, 김재수, 윤정모, 권이남, 전성진, 정택영, " 식별 체계 간 연계시스템 구축에 관한 연구", 정보과학회 2005 추계학술대회, 32권, 2호,pp. 895~897, 2005
- [2] 이상환, 신동구, 김재수, 최진영, 정택영, " 식별체계 기반의 과학기술분야 전자원문 연계시스템 설계 및 구현", 정보과학회 2004년 춘계학술대회, 31권, 1호,pp. 415~417, 2004
- [3] Ardreans Harth, "SECO: Mediation Services for Semantic Web Data", IEEE Intelligent Systems, 19권, 3호, pp.66~71, 2004
- [4] 국가과학기술인력 종합정보시스템 <http://www.hrst.or.kr>
- [5] 한국학술진흥재단 <http://www.krf.or.kr>