

연구성과물 통합 관리 시스템 구현

Implementing an Integrated System for R&D Results Management

신성호, 엄정호, 서동민, 이승우, 최성필, 정한민
한국과학기술정보연구원 소프트웨어연구실

Sungho Shin(maximus74@kisti.re.kr), Jungho Um(armian@kisti.re.kr),
Dongmin Seo(dmseo@kisti.re.kr), Seungwoo Lee(swlee@kisti.re.kr),
Sung-Pil Choi(spchoi@kisti.re.kr), Hanmin Jung(jhm@kisti.re.kr)

요약

연구개발을 통해 얻어진 결과물 자체가 잘 보존되고 관리되면, 기술이전 등 상업화에 유용하다. 그렇지만, 연구 현장에서 연구개발을 통해 얻어진 데이터셋이나 소프트웨어 등 성과물들은 개인별로 관리되고 있거나 제대로 관리되지 않아서, 재활용이 어려운 실정이다. 따라서 국가차원에서는 전체 연구성과물의 메타정보 수집을 통한 연구성과물 현황 관리와 이를 분석한 분석정보의 제공에 중점적으로 노력하고, 실제 연구개발을 수행하는 개별 연구원이나 기관들은 연구성과물의 재활용에 초점을 맞춘 성과물 관리가 필요하다. 이를 위해 본 연구에서는 연구성과물의 체계적인 관리를 위해서 연구성과물 등록, 저장, 조회/사용, 관리 등 연구성과물 관리를 위한 프로세스를 제시한다. 프로세스를 기반으로, 연구개발의 결과물을 관리하기 위한 시스템 아키텍처를 설계한다. 무엇보다 이질적인 형태의 스키마를 가지는 연구성과물을 통합 관리하기 위한 통합 데이터베이스를 어떤 모습으로 설계해야 하는가도 주요한 고려사항이다. 구현된 시스템은 사용자들에게 연구성과물에 대한 상세 정보를 보여주고, 사용자가 연구성과물 자체도 편리하게 얻을 수 있는 환경을 제공한다. 연구성과물관리를 위한 효율적인 프로세스 및 데이터베이스의 설계는 연구성과물의 재활용을 위한 비용을 절감할 수 있으며, 연구의 연속성을 확보하여 질 높은 연구성과를 창출할 수 있는 환경을 마련할 것으로 기대한다.

■ **중심어** : | 연구성과물 | 연구성과물 관리 프로세스 | 자원고유번호 | InSciTe |

Abstract

In case that R&D results from R&D projects are well managed and archived, the research institutes can transfer the valuable technologies related to R&D results with some costs to corporations. However, it is still difficult to maintain and reuse R&D results because they are managed by each person or each department and not integrated between R&D results. Therefore, the government should undertake to manage R&D results overall by collecting meta data and distribute analyzed information from meta data. Each researching institute also makes an efforts to manage R&D results focusing on their reusing. For this purpose, in this paper, we present a process to manage R&D results; insert meta data of R&D results to the system, upload files of R&D results to the database of the system, inquire, and use meta data of R&D results. Based on the process, we design a system architecture for managing R&D results. In addition, it should be mainly considered to design a global schema for integrating R&D results into one database. The system shows detailed information on R&D results and provides R&D results conveniently to users. We expect that we may reduce the cost of reusing R&D results and improve the quality of R&D results with designing efficiently a process and a global schema of R&D result management system.

■ **keyword** : | R&D Results | Process for R&D Results Management | Identified Resource Number | InSciTe |

I. 서론

연구성과는 연구개발을 통해 창출되는 특허·논문 등 과학기술적 성과와 그 밖에 유·무형의 경제·사회·문화적 성과를 의미한다. 이는 다시 단기·직접적으로 발생하는 1차적 성과와 장기·간접적으로 발생하는 2차적 성과로 구분할 수 있다[1]. 1차적 성과는 논문, 특허, 시제품 등 3 가지를 지칭하고, 2차적 성과는 연구결과를 활용하여 발생한 비용절감, 매출증대, 품질개선 등을 의미한다[그림 1].

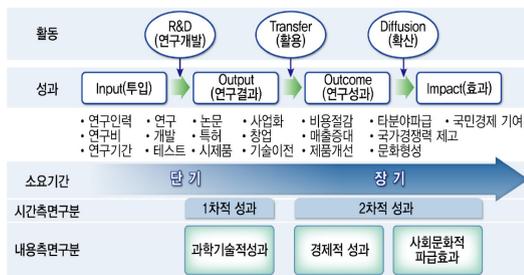


그림 1. 연구성과의 개념 및 분류

연구성과가 제대로 관리되면, 다음 연구를 위한 기반이 되고, 더 좋은 연구로 이어질 수 있다. 하지만, 현실은 연구개발을 통해 얻어진 데이터셋이나 소프트웨어 등 성과물이 제대로 관리되지 않아서 다음 연구를 위해 재활용되는데 어려움이 있다. 따라서 본 연구에서는 연구성과물 관리를 위한 시스템 기반의 관리 방안을 제시하고자 한다. 특히, 연구성과물 재활용을 위한 시스템 아키텍처를 제시한다.

II. 국내외 연구성과물 관리 현황

1. 국내 연구성과물 관리 현황

논문, 특허 등 연구성과물 관리시스템은 구축·운영되고 있으나, 부서 간 연계되지 않아 연구성과 정보의 통합 관리가 미흡한 실정이다. 논문, 특허 등은 교수업적 관리 차원에서 관리되고 있으며, 연구성과의 관리와 활용 업무가 연계되지 못하고 별도의 조직체계에 의해

수행되고 있다[2]. 성과정보시스템 간 연계비율은 대학보다는 출연연이 높았지만 미흡한 수준이며, 연구과제 관리시스템과 성과정보시스템간의 연계비율은 대학의 경우 약 35%에 불과하다[그림 2].

정부 연구개발 예산이 지속적으로 증가함에 따라 국가연구개발사업의 효율성제고 및 성과관리의 중요성이 증대되었다. 정부는 법적 근거를 토대로 성과중심의 평가제도를 도입하고 연구성과의 효율적 관리 및 활용을 촉진하고 있다[3]. 효율적인 실행을 위해 범정부 차원에서 1차, 2차 『연구성과 관리·활용 기본계획』을 수립하여 추진하고 있다[4].

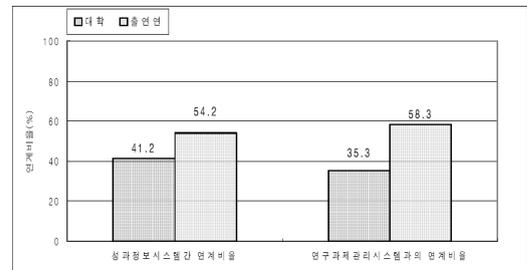


그림 2. 대학 및 출연(연)의 연구관리시스템 연계비율

2005년부터는 국가과학기술지식정보서비스(NTIS)를 구축하여 서비스하고 있으며, 이를 통해 범부처 차원의 성과관리를 위해 노력하고 있다. NTIS는 연구개발의 기획에서 성과 활용에 이르기까지 전 주기에 걸쳐 연구개발의 효율성을 높이기 위한 목적으로 구축된 지식포털로써, 과제정보, 인력정보, 시설장비정보, 성과정보 등 국가 차원에서 수행하는 R&D 사업정보를 수집·관리·서비스하고 있다[5].

2. 해외 연구성과물 관리 현황

미국의 베이-돌(Bayh-Dole) 법은 연방자금으로 지원한 모든 연구의 성과물에 대해 일반에 공개하는 것을 의무화 하고 있으며, 연구개발 종합정보시스템인 RaDiUS4와는 별도로 i-Edison이라는 연구개발 성과관리종합시스템을 구축하여 운영하고 있다[6].

i-Edison은 연방자금이 지원된 연구성과를 일반에게 공개함으로써 상업화의 기회를 증진시키는데 그 목적

이 있다. 미국정부는 연방정부의 예산이 투입된 연구개발성과를 상업적 성과로 전환하는 것을 각 연구기관의 기본적인 책임이자 의무로 규정하고 있다.

NSF(National Science Foundation, 미국국가과학재단)으로부터 지원받고 있는 UCI(University of California, Irvine)에서 구축 및 서비스하고 있는 Machine Learning Repository는 연구개발을 통해 얻어진 데이터셋과 소프트웨어들을 체계적으로 관리하여 사용자들에게 제공하고 있다[7].

이스라엘은 1950년대부터 과학기술의 상용화에 중점을 두어 대학과 연구소를 중심으로 기술 이전이 이루어져 왔다. MAGNET 프로그램을 통하여 경쟁 이전 단계 기반기술과 연구개발을 대학 및 연구기관이 반드시 민간기업과 컨소시엄을 구성할 것을 의무화 함으로써 산학연 연계를 통한 기술혁신의 상업화를 촉진하고 있다 [6].

일본은 1998년에 대학 등의 기술에 관한 연구성과의 민간 사업자에게로의 이전 촉진에 관한 법률을 제정하고, 산업 활성화와 학술 진흥을 위해 대학의 기술이나 연구성과를 민간 기업으로 이전하는 중개역으로 TLO(기술이전기관)의 활동을 정부가 지원할 것을 명시하고 있다[8]. 비교적 최근에는 지적재산기본법을 제정하여 국가, 지방공공단체, 대학 등과 사업자의 책무를 명확히 하는 동시에 지적재산의 창조, 보호, 활용에 관한 추진계획의 작성과 지적재산전략본부의 설치를 통해 지적재산에 관한 시책을 집중적으로 추진하기 위해 노력하고 있다.

이와 같은 해외 사례를 통해, 정부주도의 연구성과 관리·활용 정책 및 활성화가 필요하고, 이를 위해 연구성과 관리 기관 중심의 연구성과 관리·활용이 활성화되어야 함을 알 수 있다[8]. 이처럼 연구성과의 제대로 된 활용(상업화)을 위해서는 연구성과물에 대한 체계적인 관리가 선행되어야 한다. 연구성과물에 대한 체계적인 관리를 위해서는 연구성과물에 대한 메타정보 관리 뿐 아니라, 성과물 자체의 관리가 필요하다.

연구개발을 통해 얻어진 결과물 자체가 잘 보존되고 관리되면, 이를 활용한 기술이전 등 상업화에 유용하다. 국가차원에서는 전체 연구성과물에 대한 메타정보 수집을 통한 연구성과물 현황에 대한 종합적인 관리와 이

를 분석한 분석정보의 제공에 중점적으로 노력하고, 실제 연구개발을 수행하는 개별 연구원이나 기관들은 연구성과물 자체를 잘 관리하여야 할 것이다.

III. 연구성과물 관리를 위한 아키텍처

연구성과물 자체를 잘 관리하기 위해서는 연구성과물을 관리하기 위한 체계적인 시스템이 갖추어져야 한다. 이는 기존에 존재하던 단순 연구성과물 메타정보관리시스템과는 달리, 연구성과물 자체를 저장·관리할 수 있는 기능들을 갖추어야 한다. 연구 결과물이 정보 시스템 또는 서비스인 경우, 대표적인 연구성과물들은 소프트웨어, 데이터셋 등이 될 수 있다. 이러한 자원들을 관리하기 위한 시스템을 위해서는 메타정보 관리와는 다른 아키텍처 설계가 필요하다.

본 연구에서는 한국과학기술정보연구원에서 2010년부터 연구·개발하여 활용하고 있는 시맨틱 기술 기반의 InSciTe 서비스 개발 과제의 연구성과물 관리를 위한 자원통합관리시스템의 시스템 아키텍처 설계 방안을 제시하고자 한다. InSciTe 서비스는 연구자들의 연구 생산성을 극대화하고 정책 수립을 위한 의사 결정에 필요한 서비스를 개발하기 위해 기존 학술정보 검색 서비스와 정보분석 도구의 장점을 가지는 테크놀로지 인텔리전스 서비스이다[9]. 2011년에는 기술 예측 서비스를 위해 서비스 모델을 다양화 하고, LOD(Linked Open Data) 등 외부 데이터와의 연계를 통해 InSciTe Advanced 서비스를 제공하고 있다[10].

1. 프로세스

자원통합관리시스템의 프로세스 설계 시, 자원의 관리적인 측면과 관리된 자원의 재활용을 위한 활용의 측면으로 나누어 고려해야 한다. 먼저 자원의 관리를 위해서는 자원의 등록, 저장, 관리 등 자원등록부터 관리까지의 Workflow가 필요하다. 자원의 관리를 위한 Workflow의 설계를 위해, MicroSoft의 Software Asset Management(SAM)의 프로세스를 참고하였다[11]. MicroSoft의 SAM은 소프트웨어 현황 조사, 설치된 소

프트웨어를 구매한 라이선스와 일치 시키기, 정책 및 절차 검토, SAM 계획 수립 및 문서화의 4단계로 구성된다. MS의 SAM을 기초로 InSciTe 연구성과물에 맞게 수정하여 아래와 같이 성과물관리의 프로세스를 설계한다[그림 3].

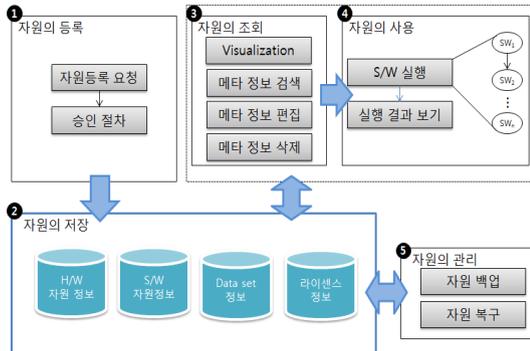


그림 3. 자원(성과물) 통합관리 프로세스

- ① 자원의 등록 : 각 자원에서 정의된 메타 정보를 관리자의 승인 절차를 거쳐 플랫폼에 등록
- ② 자원의 저장 : 등록되는 자원을 통합 자원 관리를 위한 DB에 저장
- ③ 자원의 조회 : 자원의 메타 정보를 사용자 권한에 따라 UI를 통해 조회, 검색, 편집 및 삭제
- ④ 자원의 사용 : 플랫폼을 통해 등록된 S/W를 웹 등을 통해 사용자가 실행 및 결과 확인
- ⑤ 자원의 관리 : 자원의 백업 및 복구에 대한 관리

2. 메타 항목 및 서비스

소프트웨어 자원은 사용자가 소프트웨어의 메타 정보를 조회하고, 소프트웨어의 소스코드 또는 실행 파일을 다운로드 받을 수 있으며, 웹 기반으로 소프트웨어의 기능을 테스트해 볼 수 있도록 시스템을 설계한다. InSciTe 서비스는 데이터를 기반으로 한 분석 및 예측 서비스이기 때문에, 원시 데이터로부터 시맨틱 데이터의 추출과 관련된 것들이 주된 소프트웨어 자원들이다. 따라서, 입력 또는 출력 데이터들이 XML, HTML, TXT, XLS, 트리플 등의 형태를 가진다. 이러한 소프트웨어의 기능을 사용자가 웹화면에서 실행시키기 위해

서는 테스트 대상 데이터와 파라미터를 입력할 수 있도록 UI 설계가 필요하다. 각 소프트웨어마다 단순 데이터 입력이 아니라, 데이터 입력 시 파라미터값을 요구하는 경우도 있을 수 있기 때문이다. 소프트웨어 기능의 테스트는 학술적인 목적이고, 실제 작업용도는 아니므로, 시스템에 부하가 가지 않도록 제약사항을 설정하는 것도 필요하다. 소프트웨어의 메타 정보 조회 서비스는 한국IBM에서 구축해 놓은 SolutionFinder가 참고 사례가 될 수 있다[12]. 기능 테스트는 Google에서 제공하는 번역 서비스가 참고 사례가 될 수 있다[13].

데이터셋 자원은 사용자가 데이터셋의 메타정보를 관리하는 기능, 데이터셋을 다운로드 받을 수 있으며, SQL 질의어를 던져서 간단한 결과를 제공하는 UI를 중심으로 설계한다.

이와 같은 성과물(자원)의 관리를 위한 주요 메타항목은 [표 1]과 같다. 메타정보의 관리를 통해 자원의 검색, 자원관리의 연속성, 자원의 공유 등 자원의 재활용을 위한 기반이 마련될 것으로 기대한다. 아울러 메타정보와 실제 자원의 통합 관리를 통해, 한번 생산된 자원들을 재생산하기 위한 비용을 절감할 수 있으며, 각 단위 연구소나 기관별로 연구의 연속성이 확보되어 질 높은 연구성과를 창출할 수 있는 환경이 마련될 것으로 기대한다[14].

표 1. 자원별 관리 대상 메타 항목

성과물(자원)	메타항목
소프트웨어	기본정보(유형, 명칭, 관리자정보, 설명), 설치위치, 개발언어, 사용방식, I/O포맷, 라이선스, 버전정보
데이터셋	기본정보(명칭, 관리자정보, 설명), 용도, 위치, 스키마 정보, 버전정보

3. 저장 구조

관리 대상 자원 중 소프트웨어는 중간 결과물이 없거나 최종 결과물만 관리하여도 문제가 발생하지 않지만, 데이터셋은 최종 결과물만 저장하면, 다음 연구를 위해 해당 데이터셋의 이전 버전이 필요하거나 최종 결과물로 결정한 데이터셋에 오류가 있을 때, 조치가 어렵다. 따라서 데이터셋의 경우 최초 원시 데이터의 수집부터 최종 결과 데이터셋이 만들어지는 각 단계별로 데이터

를 관리할 필요가 있다.

이를 위해 먼저, 데이터셋 구축을 위한 프로세스를 정의하고, 각 단계별 데이터베이스의 설계가 필요하다. 데이터셋이 만들어지는 프로세스는 데이터셋의 특성과 종류에 따라서 달라질 수 있기 때문에, 여러 종류의 데이터셋에 공통적으로 적용할 수 있는 표준화 프로세스가 필요하다.

표준 프로세스를 정의할 때, 개별 데이터셋들이 가지는 모든 중간 단계를 포함하여 표준 프로세스를 정의하는 것은 비효율적이다. 즉, 개별 데이터셋에서 한 번이라도 필요한 모든 단계를 포함하는 포괄적인 프로세스는 중간 단계가 너무 많아서 올바른 데이터베이스 설계라 할 수 없다. 데이터셋의 관리를 위한 프로세스 설계 목적이 향후 필요할지 모를 중간 단계의 데이터셋을 관리하고, 오류 발생 시 역추적하여 조치를 하기 위한 목적이기 때문에, 해당 목적을 달성할 수 있는 적절한 프로세스를 정의한다.

InSciTe 서비스를 위해 관리 대상이 되는 데이터셋에는 시소러스(Thesaurus), 온톨로지(Ontology), 사전(Dictionary), 코퍼스(Corpus), 관계형 지식(Triples) 등 다양하다. 이중 관계형 지식은 다양한 소스로부터 수집된 원시데이터를 가공하여 의미있는 개체를 추출하고, 개체 간 관계정보를 저장한 데이터셋이다[15]. 따라서 개별 원시 데이터마다 가공을 위한 프로세스가 존재하고, 각 단계별로 중간 결과물이 발생한다. 관계형 지식을 추출하기 위한 원시데이터에는 논문 및 특허의 초록 정보, 뉴스 및 매거진 등 웹사이트의 기사(원문) 등이 있다. 이러한 원시 데이터를 가공하여 최종 관계형 데이터셋을 구축하는 구조는 [그림 4]와 같다.

그림에서 전처리기에 의해 수행되는 전처리 단계는 파싱, 데이터 포맷변경 등 세부 단계를 거칠 수 있다. 또한 개체 추출이나 관계 추출 단계는 검증을 거치기 때문에 최종 데이터를 확보하기까지 여러 세부 단계들을 거치게 된다. 위에서 언급한 것과 같이, 세부 단계별로 데이터를 모두 저장하는 것은 비효율적이므로, 전처리의 최종결과물, 개체추출의 최종결과물, 관계추출의 최종결과물 데이터만 저장하는 것으로 데이터베이스 구조를 설계한다.

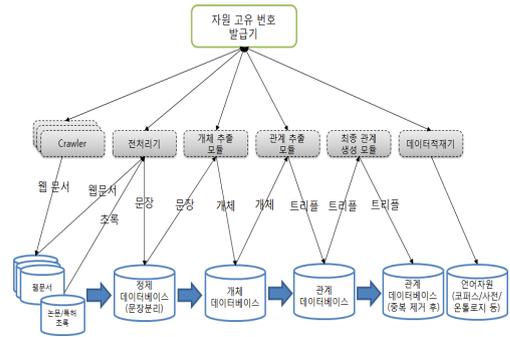


그림 4. 관계형 지식 생성 과정

데이터셋 구축을 위한 프로세스에서 각 단계별 데이터 저장소는 서로 다른 스키마를 가지게 되기 때문에, 물리적으로 서로 분리된 공간에 위치하여야 한다. 물리적인 분리를 위해 데이터 저장소마다 테이블스페이스 (Table Space, DB 계정)를 달리해 주거나, 하나의 테이블스페이스 내에서 각 단계의 데이터를 테이블로 분리할 수도 있다. 본 연구에서는 데이터 건수와 향후의 확장성을 고려하여, Crawler를 통해 수집된 웹문서, 논문/특허 초록, 정제된 데이터까지 하나의 테이블스페이스로 하고, 개체 데이터와 관계 데이터를 하나의 테이블스페이스, 언어자원을 하나의 테이블스페이스로 분리하여 저장 구조를 설계한다[그림 5].

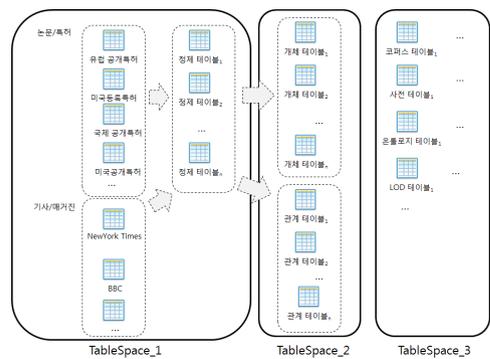


그림 5. 데이터 저장소 구조

4. 고유 번호 발급 체계

소프트웨어, 데이터셋 등 자원들은 개별 인스턴스(개별 소프트웨어, 데이터 레코드)를 가지며, 인스턴스별

로 중복 입력의 방지를 위해 고유하게 식별해 줄 필요가 있다. 인스턴스들이 고유하게 식별되지 않으면, 다른 값을 가지고 있는 인스턴스들이 동일한 데이터로 인식되어 이전에 수집하거나 저장한 데이터를 잃어버린다.

자원의 인스턴스의 중복 입력 방지를 위해, 개별 인스턴스에 대해서 고유 번호를 발급한다. 인스턴스에 대한 고유 번호 발급은 순차적인 아이디를 부여하는 것처럼 단순하지 않다. 순차적인 아이디를 부여하면, 최초 고유 번호 발급 때는 중복 문제가 발생하지 않겠지만, 동일한 데이터가 재수집 되면, 중복 문제가 발생한다. 이러한 고유 번호 발급의 문제점을 해결하기 위해 고유 번호 발급 체계의 설계가 필요하다.

고유 번호의 발급은 고유 번호 발급, 중복 체크, 기 발급 번호 관리 등의 업무로 나눌 수 있다. 고유 번호 발급은 기 정의한 번호 발급 체계에 의해 번호를 발급하는 것이고, 중복 체크는 기 발급 번호 정보들을 비교하여 동일 데이터에 대해서는 동일 고유 번호를 발급해 주고, 신규 데이터는 신규 고유 번호를 발급해 주는 업무이다. 기 발급 번호 관리는 기준에 발급된 고유 번호에 대해서 해당 인스턴스가 삭제되거나 수정되면, 해당 고유 번호의 발급 정보들에 대해서도 수정이나 삭제를 해 주는 작업이다. 이와 같은 고유 번호 발급 체계를 [그림 6]과 같이 도식화하였다.

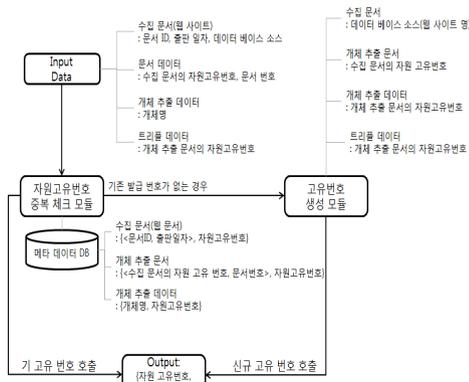


그림 6. 자원 고유 번호 발급 체계

고유 번호의 번호 체계에 대한 고려도 필요하다. 고유 번호를 의미없는 숫자만으로 구성하기보다, 고유 번

호 자체에 의미를 부여함으로써, 고유 번호만 보아도, 해당 인스턴스에 대한 특징을 알 수 있도록 번호 체계를 설계하는 것도 필요하다. 본 연구에서는 고유 번호 체계를 [그림 7]과 같이 설계한다.



그림 7. 자원 고유 번호 체계

소스는 가공하기 전의 원시 데이터를 의미하며, 원시 데이터의 종류에 따라서 각각 번호를 부여한다. 예를 들어, IDC Press Release 사이트로부터 수집된 웹문서인 경우 소스 코드를 IDC로 할당하고, The NewYork Times로부터 수집된 웹문서인 경우 소스 코드를 NYT로 할당한다. 각 소스별로 할당된 소스 코드(3자리)와 해당 소스 정보를 저장하고 있는 별도의 코드 테이블을 둔다. 저장소는 크게 수집DB, 정제DB, 개체DB, 관계DB, 언어자원DB 등으로 나누어진다. 따라서 수집DB를 SD(Seed DB), 정제DB를 CD(Cleansing DB), 개체DB를 ED(Entity DB), 관계DB를 RD(Relation DB), 언어자원DB를 LD (Linguistic DB)로 정의한다. 마지막으로 레코드 번호는 해당 인스턴스 데이터에 대해서 순차적으로 부여한다.

표 2. 자원 코드 테이블

구분	풀네임	코드명(약어)
소스	IDC Press Release	IDC
	The NewYork Times	NYT
	Wikipedia	WKP
	BBC	BBC
	Fox News	FOX
	USA TODAY	UST
	CNN	CNN
	EtnEws.com	ETN
	Techneworld	TNW
저장소	Seed DB	SD
	Cleansing DB	CD
	Entity DB	ED
	Relation DB	RD
	Linguistic DB	LD

이미 소스와 저장소 번호를 통해 해당 데이터의 특성을 반영하였기 때문에, 레코드 번호를 별도의 의미를 가지도록 설계할 필요는 없다. 코드 테이블은 [표 2]와 같다.

IV. 연구성과물 관리 시스템 구현

지금까지 연구성과물 관리 프로세스, 메타 항목 및 서비스 설계, 저장 구조 설계, 고유 번호 발급 체계 설계 등 연구성과물 관리를 위한 아키텍처를 제시하였다. 설계된 연구성과물 관리를 위한 아키텍처를 기준으로 시스템을 구현한다.

JAVA(version 1.6)와 JSP를 사용하여 구현하였고, 개발 서버의 구성은 WAS tomcat5.5 이상으로 하였고, DBMS는 MsSql2005를 사용하였다.

[그림 7]은 소프트웨어 자원 관리 화면이다. 소프트웨어명, 소프트웨어유형, 담당자, 프로그램등록여부, 메뉴얼보유여부, 개발언어, 사용방식, 개발년도 등의 정보를 등록할 수 있다. 개발년도, 소프트웨어명, 소프트웨어유형, 담당자 별로 검색도 가능하다.

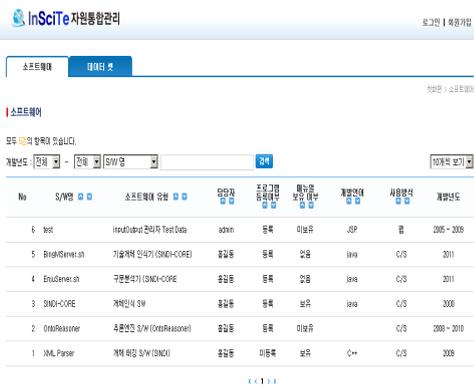


그림 7. 소프트웨어 자원 관리 화면 UI

소프트웨어 자원별 소프트웨어 자원의 개별 인스턴스에 대한 상세정보도 조회 가능하다[그림 8]. 상세정보 조회 화면에는 해당 소프트웨어에 대한 보다 자세한 정보를 확인할 수 있다. 그리고 소스코드를 내려받을 수 있고, 소프트웨어와 연관된 데이터셋에 대한 정보도 확

인 가능하다.



그림 8. 소프트웨어 자원 상세정보 조회 UI

소프트웨어 자원 등록 시, 연관된 데이터셋을 등록하기 위해, 사전에 등록 대상 데이터셋을 등록해 주어야 한다[그림 9]. 이는 데이터셋파일관리 화면에서 수행할 수 있다. 데이터셋파일관리 화면에서는 데이터셋명칭, 데이터 위치, 데이터건수 등을 등록하여, 연관된 데이터셋을 목록에서 쉽게 찾을 수 있도록 하였다.



그림 9. 연관된 데이터셋 등록 화면 UI

[그림 10]은 데이터셋 자원 관리 화면이다. 목록에서는 데이터셋명칭, 담당자, DB저장소, 건수, 최종수정일, 원시데이터 대상년도, 구축데이터 대상년도 등의 정보를 확인할 수 있다. 데이터셋명칭과 담당자별 검색도 가능하도록 구현하였다.



그림 10. 데이터셋 자원 관리 화면 UI

데이터셋 자원의 개별 인스턴스에 대한 상세정보 조회도 가능하다[그림 11]. 상세정보 조회 화면에서는 해당 데이터셋의 저장 위치, 테이블스페이스 정보, 스키마 정보를 알 수 있다. 간단한 샘플 데이터를 내려받을 수 있도록 구현하였다.



그림 11. 데이터셋 자원 상세정보 조회 화면 UI

V. 결론

일반적으로 연구 현장에서 연구개발을 통해 얻어진 데이터셋이나 소프트웨어 등 성과물들은 개인별로 관리되고 있거나 제대로 관리되지 않아서, 재활용이 어려운 실정이다. 본 연구에서는 연구개발을 통해 얻어진 결과물이 잘 보존되고 관리되어 재활용 되도록, 실제 연구개발을 수행하는 개별 연구원이나 기관들에서 연구성과물을 관리하기 위한 시스템적인 아키텍처를 제시한다. 이를 위해 국내·외 주요 국가들의 연구성과물

관리 현황을 분석하고, 바람직한 연구성과물 관리 사례를 제시한다. 아울러 MicroSoft의 SAM을 기반으로, KISTI의 InSciTe 서비스 개발의 결과물을 관리하기 위한 시스템 아키텍처 방안을 연구한다. 연구성과물의 체계적인 관리를 위해서 연구성과물 등록, 저장, 조회, 사용, 관리 등 연구성과물 관리를 위한 프로세스 마련이 필요하고, 개별 연구성과물에 대해서 어떤 메타항목들을, 어떤 저장 구조에 따라 관리할 것인지도 결정하여야 한다.

본 연구를 바탕으로, 향후에는 데이터셋, 소프트웨어 이외의 성과물 관리에 필요한 시스템 아키텍처적인 요소에 대한 연구가 진행되어야 하고, 연구소나 기관들이 공동으로 활용할 수 있는 성과물 관리 시스템을 구축하기 위한 표준 아키텍처의 설계도 필요할 것이다. 또한 연구성과물의 내용기반 검색 및 연관검색이 가능하도록 설계를 보완하는 것도 필요하다.

참고 문헌

- [1] 한국과학기술기획평가원, “정부 R&D성과 관리·활용 체계 현황진단과 시사점,” 2012.
- [2] 한국과학기술기획평가원, “공공연구기관의 연구 성과 관리·활용 현황 및 활성화 방안,” 2007.
- [3] 국회, “국가연구개발사업 등의 성과평가 및 성과 관리에 관한 법률,” 2005.
- [4] 한국연구재단, “성과관리업무 매뉴얼,” 2011.
- [5] <http://www.ntis.go.kr/ThIntroduction.do>
- [6] 한국과학기술정책연구원, “정부연구개발사업의 성과관리 현황과 향후 과제,” 2007.
- [7] <http://archive.ics.uci.edu/ml/index.html>
- [8] 한국과학기술기획평가원, “주요국 연구성과 관리·활용 제도 조사 분석,” 2011.
- [9] 이미경, 정한민, 김평, 성원경, “연구개발 전략 수립 지원을 위한 테크놀로지 인텔리전스 서비스,” 정보과학회논문지, 제17권, 제5호, pp.337-342, 2011.
- [10] H. Jung and W. K. Sung, “Towards discovering and predicting technical opportunities and technology trends,” 2011 3rd International

Conference on Data Mining and Intelligent Information Technology Applications, pp.210-212, 2011.

[11] <http://www.microsoft.com/sam/en/us/default.aspx>

[12] http://www.solutionfinder.co.kr/introduction/01_1.html

[13] <http://translate.google.com/?hl=ko&tab=0T#>

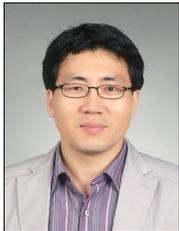
[14] 신성호, 엄정호, 최성필, 정한민, “연구성과물 재 활용을 위한 성과물관리시스템 아키텍처,” 2012 한국컴퓨터종합학술대회 논문집, 제39권(1D), pp.311-313, 2012.

[15] 신성호, 최윤수, 송사광, 최성필, 정한민, “기술 지식 자동 추출을 위한 테스트 컬렉션 구축,” 한국콘텐츠학회 논문지, 제12권, 제7호, pp.463-472, 2012.

저 자 소 개

신 성 호(Sungho Shin)

정회원



- 2000년 : 경북대학교 경영학과 (학사)
- 2002년 : 경북대학교 경영학과-경영정보시스템(석사)
- 2002년 ~ 현재 : 한국과학기술정보연구원 선임연구원

<관심분야> : 정보추출, 지식공학, 시맨틱웹, MIS

엄 정 호(Jungho Um)

정회원



- 2006년 : 전북대학교 컴퓨터공학과(석사)
- 2011년 : 전북대학교 컴퓨터공학과(박사)
- 2011년 ~ 현재 : 한국과학기술정보연구원 선임연구원

<관심분야> : 데이터베이스, 정보추출, 지식공학, 분산 병렬 처리

서 동 민(Dongmin Seo)

정회원



- 2008년 : 충북대학교 정보통신공학과(박사)
- 2008년 ~ 2010년 : 한국과학기술원 전산학과 연수연구원
- 2010년 ~ 현재 : 한국과학기술정보연구원 선임연구원

<관심분야> : XML, 시맨틱웹, 이동객체 데이터베이스, 센서 네트워크

이 승 우(Seungwoo Lee)

정회원



- 1999년 : 포스텍 컴퓨터공학과(석사)
- 2005년 : 포스텍 컴퓨터공학과(박사)
- 2006년 ~ 현재 : 한국과학기술정보연구원 선임연구원

<관심분야> : 자연어처리, 정보추출, 시맨틱웹, 정보분석, 빅데이터

최 성 필(Sung-Pil Choi)

정회원



- 1998년 : 부산대학교 전자계산학과(석사)
- 2012년 : 한국과학기술원 정보통신공학과(박사)
- 1998년 ~ 현재 : 한국과학기술정보연구원 선임연구원

<관심분야> : 기계학습, 정보검색, 자연어처리, 정보추출, 텍스트마이닝

정 한 민(Hanmin Jung)

정회원



- 2003년 : 포항공과대학교 컴퓨터공학과(박사)
- 2004년 ~ 현재 : 한국과학기술정보연구원 책임연구원
- 2004년 ~ 현재 : 과학기술연합대학원대학교 겸임교수

<관심분야> : 시맨틱웹, 정보검색, 자연어처리, HCI