

데이터 상호운용을 활용한 연구보고서 연계에 관한 연구

박민우*, 최기석*, 김재수*, 이홍로*, 주원균*

*한국과학기술정보연구원

e-mail:{pminwoo,choi,jaesoo,hongro,joo}@kisti.re.kr

A Study for R&D Reports Connection Using the Data Interoperability

Min-Woo Park*, Gi-Seok Choi*, Jae-Soo Kim*
, Hong-Ro Lee*, Won-Kyun Joo*
*KISTI

요 약

많은 연구기관에서 연구성과물로 연구보고서, 논문, 특허등의 산출물을 생산하고 있다. 이러한 데이터는 기관의 특성과 자체적으로 운영되고 있는 정보서비스의 성격에 따라 고유의 데이터 스키마와 형태를 가지고 있다. 또한 연구성과물(논문, 연구보고서)을 체계적으로 관리하고, 데이터의 부가가치를 높이기 위해 서로의 데이터의 연계와 융합을 통해 새로운 정보서비스가 요구되고 있다. 본 논문에서는 서로 다른 네트워크간 에 있는 데이터를 연계하기 위해 데이터 상호 운용을 활용하고, 이를 위한 방안으로 MAS(Multi Agent System)기반의 연계프레임워크를 통해 연구보고서의 체계적인 수집 및 서비스를 위한 데이터의 융합을 위한 방안을 제시하고자 한다.

1. 서론

이전의 인터넷이 활성화 되어 있지 않은 상황에서는 많은 연구기관에서 연구결과물로 발생하는 연구보고서, 논문, 특허등을 자체의 시스템을 통해 내부에서 성과를 관리하고 있었다. 이러한 각 기관에서는 연구성과물에 대한 정보서비스가 미흡하게 이루어졌었고, 일부 연구성과물을 대외 서비스하는 기관들은 정보서비스의 특성과 예산에 따라 연구성과물의 서지정보만 제공되고, 원문이 제공되지 않는 경우도 많았다. 이러한 정보환경에서는 새로운 연구를 하는 사람들이 기존의 연구성과물을 활용하기 어려웠다.

이러한 정부의 막대한 예산이 투입되어 생산된 R&D 연구성과물을 다양한 학자 및 연구자들이 공동활용 할 수 있는 기반이 필요하게 되었고, 이를 위해 정부에서는 연구성과물에 대하여 체계적으로 수집 및 유통할 수 있도록 8개의 유형으로 분류하여 성과관리전담기관을 지정하였다. 이중 연구보고서, 논문에 대한 전담기관을 한국과학기술정보연구원[1]으로 지정하여 이를 체계적으로 관리, 유통할 수 있도록 하였다.[2] 특히 연구보고서는 메타정보와 원문데이터의 DB구축 프로세스는 수집→가공→관리→배포의 과정으로 이루어진다. 이때 많은 연구기관수에 따라 수집 및 배포에 어려움이 많다. 또한 각 연구기관에서 생산되는 연구성과물은 연구기관의 예산과 정보서비스의 특성에 따라 고유의 데이터 스키마와 형태로 설계 되어 있어 상호간에 데이터를 연동시켜 서비스를 하는데 어려움이 있다.

본 논문에서는 이러한 문제점을 해결하는 방안으로 MAS(Multi Agent System)기반의 연계프레임워크를 활용하여 연구보고서의 메타정보 및 원문데이터를 상호연계시키는 방안에 대하여 논의 하고자 한다.

2. MAS(Multi Agent System)

에이전트는 어떤 환경에 위해 있는 컴퓨터 시스템으로서 자신의 설계 목적에 부합하기 위하여 이러한 환경에서 자발적인 행위(autonomous action)를 할 수 있고 그 에이전트는 환경으로부터 센서 입력을 취하여 상호 작용이 계속적으로 진행된다.[3] 하나의 에이전트는 독립된 기능을 할 수 있고 이러한 독립된 기능이 가능한 여러 에이전트들로 이뤄진 시스템을 멀티 에이전트 시스템이라 한다. 이들 에이전트들은 각각의 목표와 전체 시스템의 목표를 위해 각자 서로의 정보를 주고 받고, 상호 협조하고, 조화를 이루거나 협상을 한다.[4]

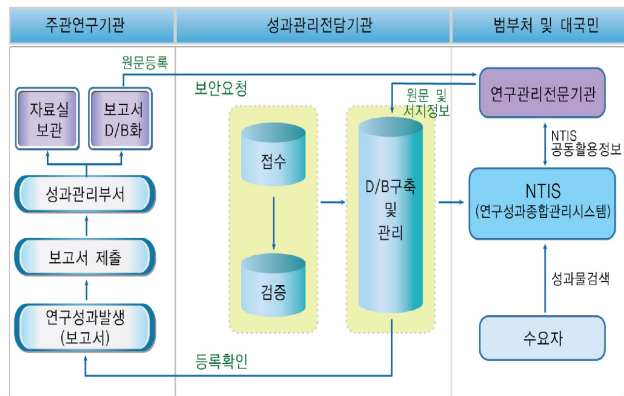
이러한 상호연계를 위해 멀티 에이전트 시스템의 지원하는 프레임워크 모델들이 다양하게 연구되어 졌고 개발되었다. 표준 모델로는 FIPA(Foundation for Intelligent Physical Agents) 모델이 있다. FIPA 모델에서 제안하는 에이전트 플랫폼은 ACC(Agent Communication Channel), ANS(Agent Name Server), DF(Directory Facilitator), AMS(Agent Management System)으로 구성되어 있다. ACC는 에이전트 사이에 메시지를 전달하며, ANS는 전역적인 에이전트 이름과 지역적인 전송 주소 사이의 매핑을

저장하며, DF는 에이전트의 기능과 그들이 제공하는 서비스를 저장하고, AMS는 에이전트의 생성, 삭제, 정지, 복구, 이동 등을 관리한다.[3][4]

이러한 멀티에이전트 시스템은 분산된 환경의 데이터를 상호연계 시키는데 효율적이고 자율성을 가지고 있는 에이전트를 추가함으로써 확성에 용이한 장점을 가지고 있다.

3. 연구보고서 연계를 위한 방안

현재 연구보고서의 관리프로세스는 연구보고서 메타정보와 원문데이터를 연구기관에서 WEB서비스를 통해 입력시스템을 통해 제공받거나 오프라인을 통해 데이터를 수집한다. 이를 연구보고서 전달기관의 서비스를 위한 스키마에 맞추어 메타정보를 가공하고 PDF 원문데이터로 DB를 구축하여, NTIS[7]와 NDSL[8]로 통해 연구보고서 정보서비스하고 있다. 전달기관에서 생성된 메타정보와 원문데이터는 필요로 하는 해당기관에 연구보고서 시스템을 통해 파일 형태로 제공되고 있다. 수집되는 메타데이터 형태는 RDBM, XML, 엑셀등로 다양하게 이루어 있고, 원문데이터는 HWP, DOC, PDF, TIF등 다양한 파일 포맷을 가지고 있었다. 이를 연구기관에서 수집된 데이터를 NDSL, NTIS 서비스와 연계를 위하여 데이터 표준화 하여 메타데이터를 RDBMS와 XML로 DB를 구축하고, 원문데이터는 PDF 파일로 구축되어 있다.



<그림 1> 보고서 정보 등록절차 및 관리체계

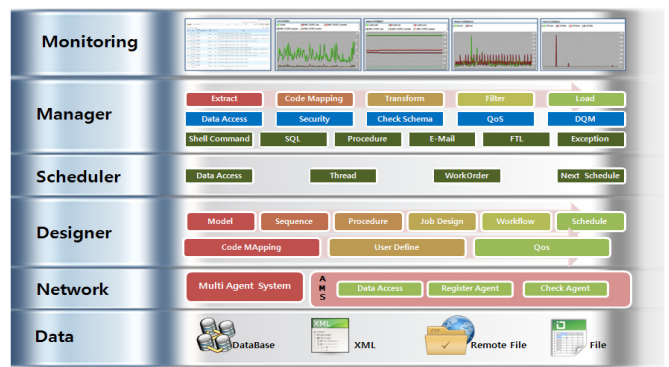
현재 성과관리 전담기관은 연구기관으로부터 수집된 데이터를 기초로 하여 고품질의 메타정보와 원문DB를 구축하고 있다. 이를 각 연구기관에서 공동활용하기 위해서는 분산되어 데이터의 관리적 측면과 업무별 다른 데이터 형태에 따라 어려운 점이 있다. 이전까지는 입력을 통한 DB구축에 중점을 두었지만 새로운 시스템에서는 연구보고서의 구축률을 높이고, 구축된 연구보고서를 공동활용할 수 방안 필요하게 되었다.

이를 위하여 수집을 위한 방안과 공동활용을 위한 방안으로 데이터 상호운용을 통해 시스템을 설계하고자 한다. 시스템 설계의 필수 요구사항은 안정적인 데이터 전송, 데이터 매핑, QoS(Quality of Service)를 보장해 주어야 한다. 또한 기존의 연계기관으로부터 데이터를 자동으로 수집할

수 있어야 하고, 자동 배포가 가능하여야 한다. 이를 위해서 기존의 수행하여 만들어진 멀티에이전트 기반의 연계프레임워크를 확장하여 데이터를 연계 시키고자 한다.

4. 연계프레임워크 설계 방안

연계프레임워크는 분산되어 있는 데이터의 안정적으로 상호연계를 위해 멀티 에이전트를 통신 기반으로 하여 AMS(Agnet Management System)에 등록된 Agent 정보를 이용하고 Agent간의 정의된 업무 규칙을 Message Protocol에 준하여 정보 전달을 수행한다. DB Adapter Agent를 통해 상호 접속할 대상의 DB 및 서버를 연결하고 Select, Update, Insert, Delete등의 Query 메시지를 전송하여 데이터를 제어하게 된다. 이렇게 멀티 에이전트 시스템은 데이터를 수집하고 네트워크 연결을 유지 할 수 있는 프레임 역할을 한다.



<그림 2> 연계·프레임워크 계층 구조도

위 그림2와 같이 연계 프레임워크 기능은 멀티에이전트 기반위에 크게 독립적으로 수행이 가능한 Designer, Scheduler, Manager, Monitoring으로 구성되어 진다. Designer는 GUI형태로 전체적인 수행할 작업을 만들고, 수행할 작업을 관리하는 기능을 수행한다. 하나의 WorkFlow를 생성하여 데이터매핑을 위한 설계를 하고, 데이터 매핑을 위한 룰을 생성 한다 생성된 WorkFlow에 수행을 위해 Job, 스케줄을 설정한다. Scheduler는 Designer를 통해 정해진 스케줄에 따라 실제적으로 수행을 한다. Manager는 Designer를 통해 설계된 Job을 실제적으로 수행한다. 데이터를 실제적으로 ETL(Extraction Transformation Load) 기능을 수행하여 두 상호간의 데이터를 매핑 시킬 수 있다. 데이터 상호운용을 위해서는 데이터의 품질과 시스템의 안정성이 필요하다. 이를 위해 Manager에서는 데이터 전송에 특정 수준의 성능을 보장하기 위한 QoS(Quality of Service), 보안을 위한 Security, 연결한 대상의 DB형태를 제어하는 Data Access등을 제공한다. Monitoring은 데이터의 전송상태, 데이터 전송 Log를 관리기능을 수행하고, 각 연구기관의 데이터 상호연계의 통해 전송된 데이터의 이력을 관리 할 수 있도록 통합관리시스템으로 로그정보를 보내준다.

또한 데이터 상호운용을 위한 연계프레임워크를 이용하여 연구보고서의 효율적인 데이터 수집 및 배포를 위해

서는 기존의 입력시스템과 상호간에 인터페이스를 제공해야 한다. 메타데이터 및 원문 DB를 새롭게 구축하는 경우는 기존의 입력시스템을 통해 데이터를 입수하고, 연구기관에서 직접 메타데이터와 원문을 가공하여 만드는 경우는 수집과정에서 연계프레임워크를 통해 입수하여, 기존의 입력시스템과 연계를 통해 메타정보입력을 최소화하고, 원문의 경우는 스케줄에 따라 자동적으로 연계프레임워크의 FTP 전송을 통해 가지고 올 수 있도록 한다. 이를 통해 연구보고서 DB 구축률을 향상 시킬 수 있을 것이다.

공동활용 측면에서 보면 전담기관에서 구축된 메타데이터 및 원문을 각 연구기관에 맞는 데이터 형태로 변환하여 연계하고, 원문데이터를 활용한다면 연구보고서의 공동활용도를 높이고, 중복으로 DB를 구축하는 비효율성을 줄일 수 있을 것이다. 또한 전담기관에서는 연구보고서의 고부가가치 서비스를 위해 서지정보 및 전문을 XML화 시켜 DB를 구축하고 있다. 각 연구기관에서 이러한 데이터를 활용한다면, 각자의 연구기관에 맞는 효과적이고 다양한 정보서비스 구축에 활용 될 수 있을 것이다.

5. 결론

연구보고서의 데이터의 효율적인 데이터 수집 및 공동활용을 위해서는 각 연구기관의 데이터 상호연계가 필요하다. 상호연계를 위해 멀티에이전트 기반의 연계프레임워크를 이용하는 방안을 제시하였다. 이는 국가R&D 성과물 활용을 활성화 시키고, 확산 시킬 수 있는 체계를 만들어 국가의 R&D 사업의 연구효율성 및 생산성을 극대화 할 수 있을 것이다. 본 논문에서는 요구사항 측면에서 분석하고 이를 위해 연계프레임워크를 설계방향을 제시하였다. 향후 이를 위해서는 구체적인 기능적 상세 설계 및 기능 구현이 필요할 것이다.

참고문헌

- [1] 한국과학기술정보연구원, <http://www.kisti.re.kr>
- [2] 국가연구개발사업의 관리 등에 관한 규정 제24조 제13항
- [3] Michael Wooldridge, An Introduction to MultiAgent Systems, Wiley, 2002
- [4] R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi, Reasoning about Knowledge, MIT Press, Cambridge, MA, 1995
- [5] MultiAgent Systems , <http://www.multiagent.com/>
- [6] FIPA(Foundation for Intelligent Physical Agents) , <http://www.fipa.org>
- [7] NTIS , <http://www.ntis.go.kr>
- [8] NDSL , <http://www.ndsl.kr>