

# 학술정보 관리 및 유통을 위한 에이전트 시스템 설계에 관한 연구

A study of design the agent system  
for management and service of S&T information

이석형, 윤희준, 윤화목  
한국과학기술정보연구원 시스템개발팀

Lee Seok-Hyoung, Yoon Hee-Jun, Yoon Hwa-Mook  
KISTI

## 요약

본 논문에서는 다양한 학술정보들을 수집, 분석, 처리 및 관리가 이루어지는 학술정보유통 에이전트 시스템을 제안한다. 일반적으로 학술정보를 수집하여 이를 서비스하기 위한 절차는 대부분 유사하게 처리되나, 데이터의 특성과 운용하고 있는 하드웨어와 소프트웨어 종류에 따라 같은 내용임에도 다른 내용으로 가공되기도 한다. 또한, 정보를 가공하는데 있어 오프라인 작업이 발생하는 경우 처리 인력 및 시간에 많은 소요가 되기 때문에 최신의 정보를 빨리 제공하는데 어려움이 있다. 이에, 데이터에 독립적이면서도 모든 데이터를 수용할 수 있는 학술정보유통프레임워크의 설계와 개발이 중요한 요소가 되며, 또한 프레임워크에 적용된 에이전트 시스템은 인력 및 시간 절감뿐만 아니라 시스템의 확장성 및 유연성을 보장하며, 양질의 정보서비스 제공이 가능할 것으로 기대한다. 학술정보유통 에이전트 시스템은 원시자료형식 등록 에이전트(Resource Format Register Agent), 자료 수집 에이전트(Data Collector Agent), 자료 분석 에이전트(Data Analyzer Agent), 자료처리 에이전트(Data Processor Agent), 데이터베이스적재 에이전트(MetaDB Creator Agent)의 멀티에이전트로 구성된다. 본 논문은 에이전트 시스템이 수행하는 필수적인 기능과 요소에 대해 설명한다.

## Abstract

In this paper, we suggest the Agent System for management and service of science and technology information that performs the collection, analyzing, processing and managing the information. Generally, any system can process the collection and service the information as usual method, but that mean can't be clear because of the data process method on different systems of hardware and software. So, it's very important point that the design and development of S&T information flow framework that can accept the various data, and we expect that the agent system adapted in framework guarantee the reduction of man power and process time and the enlarged system. Agent System consists of five multi agents, Resource Format Register Agent, Data Collector Agent, Data Analyzer Agent, Data Processor Agent and MetaDB Creator Agent. We describe the main roles and elements of agent systems.

## I. 서론

연구자나 과학기술자들이 제공하거나 작성하는 논문, 연구보고서, 뉴스 정보 등의 과학기술정보는 또 다른 연구결과를 낼 수 있는 중요한 정보이다. 정보 이용자 또는 연구자의 입장에서 연구 업적을 평가하고 재활용하기에 충분한 데이터를 얻을 수 있고, 진행 중인 연구들의 추세를 확인하여 연구방향 설정, 연구과제 도출업무 등에 필수적인 정보를 얻을 수 있다 [1,2].

이러한 정보의 가공 형태는 일반적으로 데이터 작성기관에서 제공하는 정형/비정형 형식의 서지정보를 수집→등록→분석→재가공→DB적재→서비스 하는 절차로 이루어진다. 그러

나, 비슷한 성격의 신규 학술정보가 발생하였을 경우 위의 절차에 따라 가공이 가능하도록 신규 정보에 맞게 학술정보유통 시스템 내에 추가 모듈을 개발해야 하므로 불필요한 시스템 개발 비용이 추가된다. 더욱이, 오프라인으로 수집되는 서지정보를 등록하고 가공하기까지, 혹은 가공 후 데이터베이스에 적재하기까지, 관리자가 그 내용을 인지하는데 까지 불필요한 시간이 걸리기 때문에, 가공 효율성이 많이 저하된다.

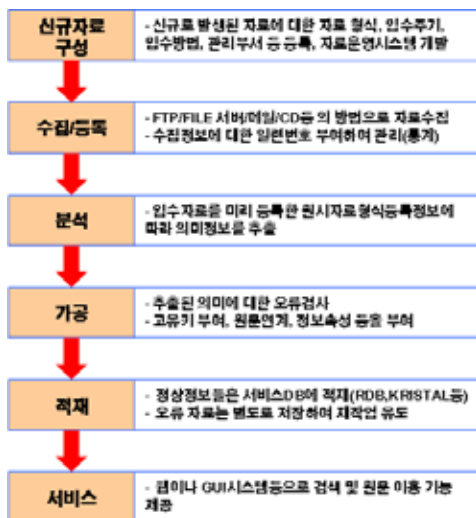
이에 본 논문에서는 학술정보유통 프레임워크를 위한 학술정보유통 에이전트 시스템을 제안한다. 학술정보유통 에이전트 시스템은 다양한 학술 정보를 수집, 등록, 분석, 처리 및 관리를 위한 멀티 에이전트 기반의 시스템으로 에이전트 서버 내

에 존재하여, 파일 서버나 데이터베이스 서버를 원격으로 접근하여 실시간으로 데이터를 처리한다. 또한, 학술정보유통을 위한 모든 가공 및 처리 방법을 분석, 대부분의 기능을 지원하도록 설계하였다. 본 시스템은 크게 원시자료형식 등록 에이전트(Resource Format Register Agent), 자료 수집 에이전트(Data Collector Agent), 자료 분석 에이전트(Data Analyzer Agent), 자료처리 에이전트(Data Processor Agent), 데이터베이스적재 에이전트(MetaDB Creator Agent)로 구성된다.

본 논문의 구성은 다음과 같다. 2장에서는 과학기술정보유통을 위한 절차를 국내의 대표적인 과학기술정보제공기관인 한국과학기술정보연구원의 yeskisti.net을 참고하여 분석한다. 3장에서는 학술정보유통 에이전트 시스템의 구성요소를 설명하고 5개 멀티에이전트의 기능에 대해 기술하고, 4장에서 결론을 맺는다.

## II. 학술정보유통 현황 분석

일반적으로 과학기술정보를 수집하는 절차는 텍스트나 XML 등 정형 혹은 비정형 구조의 파일로 이루어진 정보를 FTP나 Mail 혹은 DB에 저장된 정보를 가져오는 것부터 시작된다. 수집된 정보는 등록 절차를 거쳐, 서지 정보의 의미추출을 위한 파싱 작업 후 의미에 대한 검증과 재가공을 거쳐 데이터베이스에 저장하여 서비스 하는 식으로 이루어진다. 그림 1은 학술정보유통 흐름도를 설명한 것이다. 본 장에서는 학술정보유통 에이전트 시스템의 필수 모듈을 구성하기 위해, 국내의 대표적인 과학기술정보제공기관인 한국과학기술정보연구원의 yeskisti.net[3]의 분석, 동향, 논문, 연구보고서, 특허, 인력 및 기타 7개 분류, 22종의 데이터베이스, 35개 입수처를 대상으로 서지정보 형식, 수집형태, 수집방법 및 가공방법을 분석하였다.



▶▶ 그림 1. 학술정보유통 흐름도

### 1. 원시자료형식

한국과학기술정보연구원의 yeskisti.net에서 서비스하고 있는 정보는 크게 4가지 원시자료형태로 구분되며, 세부적으로 8가지 원시자료형태로 나눌 수 있다. XML과 HTML 등 정형 포맷으로 구성된 형태와 Tagging 된 텍스트 포맷, MARC 포맷, 그리고 데이터베이스에 저장되어 있는 형태 등 4가지 형태로 구분된다. 또한 Tagging 된 텍스트 포맷은 레코드 구분자와 라인 당 구성 항목에 따라 type-1부터 type-5까지 5개의 세부 포맷으로 구분된다. 원시자료 형식은 표 1과 같으며, 오른쪽의 건수는 해당 원시자료형식을 채택하고 있는 입수처 건수를 나타낸다.

[표 1] 원시자료형식

원시자료형식	건
XML/HTML	7
TEXT Type1-라인당 하나의 레코드로 구성되며, 특정 구분자로 필드가 구분	1
TEXT Type2-라인당 하나의 필드명 및 값으로 구성되며, 레코드구분자 존재	6
TEXT Type3-라인당 하나의 필드값만으로 구성되며, 레코드구분자 존재	1
TEXT Type4-라인당 하나의 필드명 및 값으로 구성되며, 레코드구분자 미존재	12
TEXT Type5-라인당 하나의 필드값만으로 구성되며, 레코드구분자 미존재	0
MARC	2
DATABASE	5

### 2. 원시자료저장형식

원시자료의 저장형식은 표2 와 같이 크게 다섯 개의 형태로 구분하였다. 오른쪽의 건수는 해당 자료의 저장형식을 채택하고 있는 입수처 건수를 나타낸다.

[표 2] 원시자료저장형식

원시자료저장형식	건
XML/HTML	7
TEXT	20
MARC	2
DATABASE	5

### 3. 수집방법

일반적으로 파일형태의 서지정보를 수집하는 방법은 표3과 같이 크게 3가지로 구분된다. 오른쪽의 건수는 해당자료의 수집방법을 채택하고 있는 입수처 건수를 나타낸다.

[표 3] 수집방법

수집방법	건
FTP 및 FILE 서버등 대용량 서버에 저장	12
Local PC에 저장	17
DATABASE에서 레코드 추출	5

#### 4. 자료검증방법

입수된 정보를 데이터베이스에 저장하기 위해서는 데이터의 검증 절차가 필요하다. 검증은 데이터의 의미의 검증이 아니라 단지 전산으로 처리가 가능한 항목 누락이나 중복자료의 추출 등의 기능만을 포함한다[4]. 또한 데이터의 유형과 자료의 성격에 따라 검증하는 내용이 상이하니, 그 범주는 크게 넓지 않는 것으로 파악되었다. 일반적으로 자료검증 내용은 다음과 같다.

- 항목검증내용
  - 제목 누락여부
  - 권호페이지정보 누락여부
  - 발행정보(발행년도, 저널명등) 누락여부
- 중복검증방법
  - 제목+발행년도
  - 제목+저자
  - ISSN+VOLUME No.+Page No.
  - 입수처 Primary Key
- 기타 자료 처리내용
  - 원문정보 추가(디렉토리명, 파일명)
  - Object Identity 부여(KO,DOI)
  - 초록, 키워드등 부가 서지정보 추가
  - 자료 속성 부여(초록여부, 유료여부, 과금여부, 소장여부)

위와 같이 자료검증이 완료된 정보는 검색 데이터베이스나 관리형 데이터베이스에 저장되어 최종 이용자에게 서비스 된다.

### III. 학술정보유통 에이전트 시스템

2장에서 살펴본 바와 같이 학술정보유통 흐름도는 일정한 절차와 방법에 의해 수행되고 있으며, 수집, 가공되는 원시 자료의 형식 및 저장형식, 수집방법, 가공방법은 2장에서 살펴본 내용에 크게 어긋나지 않는 것으로 파악된다. 학술정보유통 흐름도 및 수집 가공 현황을 바탕으로, 본 장에서는 학술정보유통 에이전트 시스템에 대해 설명한다.

#### 1. 시스템 구성요소

학술정보 에이전트 시스템은 크게 원시자료형식 등록 에이전트(Resource Format Register Agent), 자료 수집 에이전트(Data Collector Agent), 자료 분석 에이전트(Data Analyzer Agent), 자료 처리 에이전트(Data Processor Agent), 데이터베이스 적재 에이전트(MetaDB Creator Agent)로 구성된다. 각 에이전트는 등록된 파일서버와 FTP서버, 데이터베이스 서버와 연계되며, 내부적으로 중계기 에이전트를 두어, 에이전트 간에 발생하는 모든 작업내용을 제어하고, 모니터링하는 역할을 하도록 한다. 또한 에이전트가 접근할 파일서버 및 FTP 서버, 데이터베이스 서버의 정보를 저장하는 서버정보 등록기(Server Information Registry, SIR)과 원시자료의 형식과 속성, 처리방법들을 등록하는 자료 처리 등록기(Data Process Method Registry, DPMP), 원시자료의 처리과정을 저장하는 자료 수집 등록기(Data Collection Registry, DCR)을 두어 각 에이전트에서 관련 내용을 참조할 수 있도록 한다. 그림 2는 학술정보유통 에이전트 시스템을 보인 것이다.

#### 2. 에이전트

##### 2.1 원시자료형식 등록 에이전트

원시자료형식 등록 에이전트는 별도로 구성된 자료 처리 등록기(Data Process Method Registry, DPMP)와 연동되어, 자료관리자가 새로운 신규자료에 대한 자료형식과 속성을 DPMP에 등록할 수 있도록 도와주는 역할을 수행한다.

##### 2.2 자료 수집 에이전트

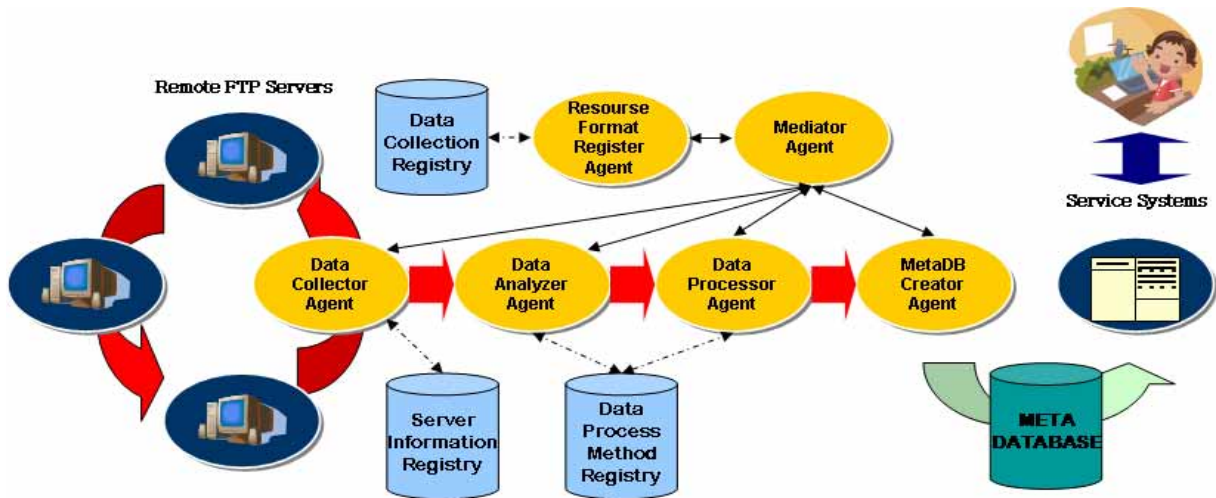
자료 수집 에이전트는 SIR에 저장된 입수정보의 수집 주기를 파악하여, 해당 서버에 신규 파일 정보가 접수되었을 경우, 입수정보를 등록하고, 파일을 자료 처리 에이전트에게 전송하는 역할을 수행한다. 자료 수집 에이전트를 통해 등록하는 입수정보는 입수일자 및 회차정보, 입수 정보 건수 등이다.

##### 2.3 자료 분석 에이전트

자료 분석 에이전트는 DPMP에 저장된 원시자료형식을 바탕으로 서비스하고자 하는 항목에 값을 매칭시키는 작업을 수행하는 에이전트로, 작업결과를 XML문서로 작성하여 자료 처리 에이전트로 전송한다. 이때, 발생한 오류내용은 중계기 에이전트로 전송하여 관리하도록 한다.

##### 2.4 자료 처리 에이전트

자료 분석 에이전트에서 넘겨진 결과를 참고하여 2장에서 기술한 자료검증절차를 자료처리 에이전트가 수행한다. 자료검



▶▶ 그림 2. 학술정보유통 에이전트 시스템 구성도

증내용은 DPMR에 등록된 자료검증내용 및 방법을 참고하여 이루어지며, 검증 결과 정상 및 오류코드, 오류내용을 포함하여 XML로 구성된 데이터메시지를 데이터베이스 적재 에이전트로 전송한다.

### 2.5 데이터베이스 적재 에이전트

SIR에 기술된 데이터베이스 서버의 해당 테이블로 정상 혹은 오류정보가 적재된다. 적재 과정에서 발생하는 오류사항이나 적재 건수 등은 별도의 자료수집등록기에 저장된다.

### 2.6 중개기 에이전트(Mediator Agent)

중개기 에이전트는 원시자료형식 등록 에이전트, 자료 수집 에이전트, 자료 분석 에이전트, 자료 처리 에이전트, 데이터베이스 적재 에이전트 등이 실행과정에서 발생하는 모든 실행정보와, 처리 현황 및 결과, 오류사항 등을 종합적으로 처리, 분석, 관리가 가능한 에이전트로, 하나의 원시자료가 수집되어 서비스 되는 일련의 과정을 모니터링하거나 그 결과를 열람할 경우에도 본 에이전트가 수행된다.

## 3. 시스템 흐름

본 에이전트 시스템과 2장에서 기술된 학술정보유통 흐름도는 다음과 같이 매핑된다.

- 신규자료 구성: 원시자료형식 등록에이전트(자료 처리 등록기, 서버 정보 등록기)
- 수집/등록: 자료 수집 에이전트(자료 수집 등록기)
- 분석: 자료 분석 에이전트(자료 처리 등록기)
- 가공: 자료 처리 에이전트(자료 처리 등록기)
- 적재: 데이터베이스 적재 에이전트(자료수집등록기)

## IV. 결론

본 논문에서는 학술정보유통 프레임워크를 위한 학술정보유통 에이전트 시스템을 제안하였다. 본 시스템은 학술정보유통에 필요한 인력 및 시간 절감뿐만 아니라, 새로운 정보가 추가되었을 경우에도 학술정보유통 시스템의 확장성 및 유연성을 보장하며, 양질의 정보서비스 제공이 가능하다. 또한, 학술정보뿐만 아니라 인력정보, 생물정보, 멀티미디어 콘텐츠 정보 등 다양한 형태의 정보를 수집, 분석, 가공, DB적재, 서비스하는데 유용하게 이용될 수 있을 것으로 기대한다.

### ■ 참고 문헌 ■

- [1] 이상현, 김익철, "광기술을 이용한 연구보고서 관리시스템 구축", 정보관리학회지, 제 9권 1호. 1992.
- [2] 정동원, 신동길, 정은주, 김영갑, 이정욱, 백두권, "과학기술정보를 위한 메타데이터 레지스트리 관리시스템", 정보과학회논문지, 제 10권 제 56호. 2004.12
- [3] 이석형, 강남규, 윤화목, 예용희, 김현, "Implementation of the XML based Science and Technology Information System using KRISTAL", The 7th IASTED International Conference, IMSA 2003, 2003.08
- [4] 한국과학기술정보연구원, "과학기술종합정보 서비스시스템 개발 및 운영", 2004.