

## RSS를 이용한 과학기술정보 연계시스템 설계 및 구현

The Design and Implementation of connection system of  
Science and Technology Information using RSS

이민호, 박상배, 주원균, 양명석, 최기석  
한국과학기술정보연구원

Lee Min-Ho, Park Sang-Bae, Joo Won-Kyun,  
Yang Myung-Seok, Choi Ki-Seok  
KISTI(Korea Institute of Science and Technology  
Information)

### 요약

정부 출연 연구기관에서는 수행하고 있는 연구과제에 대한 산출물뿐만 아니라 세미나/일정·연구 성과·사업공고·최근 연구동향 등 다양한 정보를 생산해 내고 있으며, 이러한 정보들은 과제관리 기관·연구회 등의 기관에 수집되고 가공되어 더 가치있는 정보로 거듭나게 된다. 이러한 정보의 흐름 속에서 정부 출연 연구기관과 정보 수집 기관의 정보 시스템들 간 연계는 필수적이라 할 수 있다. 기존의 정보 연계 시스템들은 자체적으로 정의된 프로토콜을 통하여 단단하게(tightly) 연계하거나 웹을 통한 HTML 파싱을 이용해 연계할 수밖에 없었다. 하지만, 이러한 방식은 새로운 시스템을 추가하거나 기존 시스템을 변경할 경우 빠르게 대처하기가 어려운 문제가 있다. 최근 들어 사용자가 원하는 시점에 원하는 정보만을 제공받을 수 있는 풀(Pull) 방식의 표준화된 포맷인 RSS[1]가 각 분야에서 다양하고 활발하게 연구 및 활용되고 있는데, 본 논문에서는 RSS를 이용하여 과학기술 정보를 연계/수집하는 시스템을 설계 및 구현한다. 또한 기존 시스템의 연계방식과 RSS의 장단점에 대해서 알아보고, 이것을 기초기술연구정보망 시스템에 적용한 사례도 살펴본다.

### Abstract

Research Institutes of Government has produced various information about research, seminar, recent research trend, and so on. This information flows into project management organization and institute management organization. It is necessary to connect information systems of these organizations and institutes. The existing systems connect with self-defined protocol between the organizations and the institutes or HTML parsing method. but, these methods have many drawback for expanding and changing problem. this paper designs and implements science and technology information connection system using RSS.

## I. 서론

정부 출연 연구기관(이하 출연(연))은 과제 관리기관으로부터 연구 사업에 대한 지원을 받아 연구 과제를 수행할 뿐만 아니라, 분야별 연구회에 소속되어 연구기관 운영에 대한 지원과 평가, 발전방향에 대한 지원 등을 받는다. 이러한 관계 속에서 출연(연)이 생산해내는 정보는 보고서, 논문, 특허 등의 연구 과제 산출물과 세미나 일정, 연구 성과, 사업공고, 최근 연구동향, 뉴스정보 등 매우 다양하다.

출연(연)이 생산하는 정보들은 과제관리기관이나 연구회와 공유되는 것들이 상당히 많이 있는데, 그러므로 과제관리기관 및 연구회와 출연(연)간 정보 시스템들의 연계는 매우 중요하다. 기존 정보 시스템들의 연계는 시스템들 간에 자체적으로 정

의한 프로토콜과 데이터 구조를 통하여 단단하게(tightly) 연계하거나 웹을 통한 HTML 파싱을 하는 방법으로 구축하여왔다.

하지만 이러한 방법들은 표준적인 방법이 아니라 연구회가 새롭게 관리하는 출연(연)이 생길 경우 시스템 호환 문제가 생기기도 하고, 기존 웹 결과 화면이 변경되었을 경우에도 HTML 파싱 모듈을 수정해 줘야하는 등 확장 및 유지관리에 어려움이 발생한다. 또한, 이러한 확장 및 변경에는 과제관리기관, 연구회, 출연(연) 등 관련 기관들이 서로 긴밀하게 지속적으로 시스템 변경에 대한 협력이 이루어지지 않으면 안된다.

최근 들어 사용자가 원하는 시점에 원하는 정보만을 제공하고 받을 수 있는 표준화되고 확장이 용이한 포맷인 RSS가 각 분야에서 다양하고 활발하게 연구 및 활용되고 있다.

본 논문에서는 RSS를 이용하여 출연(연)과 연구회 시스템 간 과학기술 정보를 연계/수집하는 시스템을 설계 및 구현한다. 다음 장에서는 기존 연계 방법과 RSS를 이용한 방법을 설명하고 장단점에 대해 알아본다. 3장에서는 정보 연계 시스템에 대한 설계 및 구성을 살펴보고, 4장에서는 기초기술연구정보망 시스템에 적용한 사례를 설명한다. 마지막 장에서는 향후 시스템 발전 방향을 결론으로 끝을 맺는다.

## II. 기존방법과 RSS의 장단점

### 1. 기존방법

기존 시스템에서의 정보 연계는 자체적으로 정의한 전용 프로토콜로 데이터를 전송하는 방법과 HTML 파싱을 이용하는 방법을 사용하였다.

전용 프로토콜을 사용하는 방법은 시스템 간에 사용할 프로토콜을 정의하고 이에 맞는 정보 전송 모듈을 구현하는 방식이다. 이 방법은 특수한 형식의 데이터를 보내거나 기밀성이 요구되는 정보를 전송하기에 적합하지만, 이미 구축된 시스템을 변경하고 방화벽을 여는 등 기관간의 긴밀한 협조가 필요하다.

HTML을 이용하는 방법은 표준 프로토콜인 HTTP를 사용하므로 기존에 사용하던 웹을 그대로 사용할 수 있으며, 저작권 문제를 제외하고는 방화벽 문제나 시스템 변경과 같은 복잡한 기관간의 협의는 필요가 없다. 하지만 다양한 웹 결과를 처리해야 하므로 HTML 파싱 모듈이 복잡해지며, 정보를 제공하는 웹 사이트가 변경되었을 경우 파싱 모듈을 계속 수정해 줘야 하는 등 유지, 보수가 어렵다. 또한, 기밀성을 유지하여야 하는 민감한 정보는 추가적인 보안 대책이 필요하다.

### 2. RSS

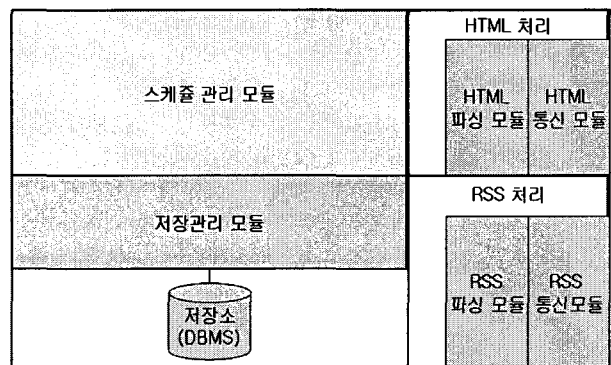
RSS는 RDF Site Summary, Rich Site Summary, Really Simple Syndication 등의 약칭으로 뉴스, 블로그와 같이 콘텐츠 업데이트가 자주 일어나는 웹사이트에서, 업데이트된 정보를 자동적으로 쉽게 사용자들에게 제공하기 위한 서비스로 개발되었다. 사용자가 선택적으로 구독할 수 있으며 자동화 연동이 매우 쉽게 구조화된 XML 포맷을 사용하므로 콘텐츠 재사용성이 높다. 또한, 사용자가 많아 널리 이용되고 있는 방식이므로 RSS 리더나 개발도구가 손쉽게 구해서 사용할 수 있다. 일반적으로 http 프로토콜 위에서 동작하므로 출연(연)의 방화벽 문제가 없으며 미리 정의된 아이템들만 RSS 콘텐츠 생성 모듈이 계속 제공해준다면 결과 출력 웹 화면이 변경되더라도 RSS 수집에는 영향을 끼치지 않아 유지 보수가 쉽다는 장점이 있다. 또한, 새로운 시스템을 추가할 경우에는 RSS 생성 모듈만 붙이면 되므로 확장이 용이하다. 하지만 RSS 정보연계 초

기에는 RSS 생성 모듈을 설치해야하며, 정보 배포가 주된 목적으로 설계되었으므로 보안에 민감한 정보는 사용하기 어렵다.

RSS는 크게 RDF 계열인 RSS 0.91, 1.0, 1.1 버전과 비RDF 계열인 RSS 0.92, 0.93, 0.94, 2.0 등이 있으며 PIE, ATOM, OPML, mbox 등의 또다른 방식의 RSS가 있다. 각 방식에 따라 RSS 형식이 조금씩 다르므로 특정 방식에 맞게 구독 모듈을 만들던가 모든 형식을 지원하는 구독 모듈을 구현해야 하는 불편이 있었다. 하지만 2005년 IETF 표준 초안으로 ATOM이 제출되어 표준화 작업에 들어갔으며, 형식이 크게 다르지 않으므로 요즘에는 모든 형식을 지원하는 모듈들이 많이 개발되고 있다[2].

## III. 시스템 설계 및 구현

설계한 시스템은 크게 RSS 배포 모듈, RSS 파싱 모듈, 기존 HTML 정보 수집을 위한 HTML 파싱 모듈, HTML 통신 모듈, 정보 수집 간격을 조정하고 수행하는 스케줄 관리 모듈, 수집된 정보를 저장하는 저장관리 모듈로 구성되어있다.



▶▶ 그림 1. 과학기술정보 연계시스템 구성도

구현된 시스템은 Java 기반의 공개 라이브러리인 ROME을 기반으로 배포 모듈과 RSS 파싱 모듈을 구현하였으며, 저장소는 상용 DBMS를 사용하였다. 스케줄 모듈과 저장관리 모듈 등은 Java로 직접 구현하였다. 배포/수집하는 정보는 과학기술 콘텐츠에서 많이 사용되는 형식으로 제목, 내용 연결 URL, 배포일자로 이루어져 있다. 현재는 단순하지만 추가 확장이 용이한 RSS의 특성상 필요할 경우 매우 쉽게 수정될 수 있게 구현되었다. RSS 및 HTML 파싱으로 수집된 정보는 UTF-8 코드로 변환되어 저장된다.

ROME은 Java기반의 공개 라이브러리로 RSS 0.9, 0.91, 1.0, 0.92, 0.93, 0.94, 2.0, Atom 0.3, 1.0 등 다양한 형식을 지원하고 있으며 RSS/ATOM 파서와 생성기 모두 구현되어 있고, 사용하기 쉬운 인터페이스를 가지고 있다[3].

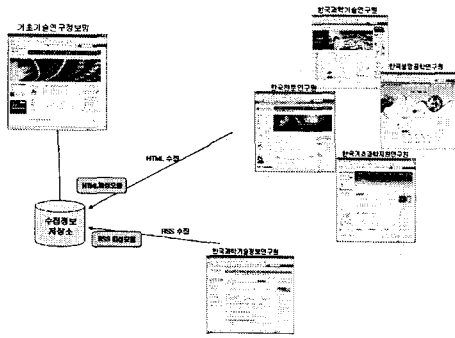
### IV. 적용사례

본 논문에서 구현한 시스템은 기초기술연구회와 소관 연구기관간의 정보 연계에 사용되었다. 기초기술연구회는 기초기술 분야의 정부출연 연구기관을 지원/육성하고 체계적으로 관리함을 설립목적으로 둔 조직으로 한국과학기술연구원, 한국생명공학연구원, 한국기초과학 지원연구원, 한국천문연구원, 한국한의학연구원 등 5개 정부출연 연구원을 소관기관으로 두고 있다.

기초기술연구회의 기초기술연구정보망 시스템은 소관 출연(연)에서 생산하는 정보 뿐만 아니라 기초기술과 관련된 정보들을 여러 매체에서 수집하여 소관 연구기관의 연구원및 대국민 서비스를 하고 있다.

서비스되는 정보는 소관기관의 공지사항, 발표뉴스, 세미나/일정, 연구성과, 사업광고, 채용광고 등의 뉴스정보와 정보통신, 나노소재, 생명공학, 천문, 기초과학 지원 관련 최근 연구동향 정보 등으로 수집/제공하고 있다. 그 외에 소관기관의 경영정보, 지적재산권 등 연구성과정보, 기관평가 정보등이 있다.

서비스되는 정보 중 뉴스정보와 최근 연구동향 등이 자동 연계하여 수집되는 정보이며 기존에는 뉴스정보는 각 소관기관의 홈페이지로부터 수집하고 최근 연구동향은 한국과학기술정보연구원과 인터넷 신문 등 각종 미디어로부터 html 파싱을 통하여 수집하고 있었다.

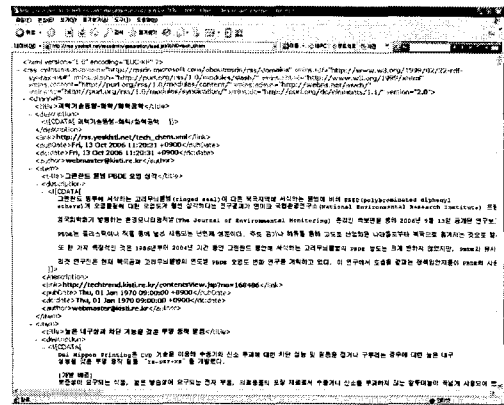


▶▶ 그림 2. 기초기술연구정보망 정보수집체계 구성도

뉴스정보는 각 출연(연)이 제공하는 웹 화면을 HTML 파싱하여 그대로 사용하면 되었으나, 최근 연구동향 정보는 정보를 제공하는 인터넷 매체를 찾는 것부터 어려웠으며, 제공한다 해도 여러 정보가 통합되어 기초과학 분야만 분리해내기 어려웠다.

한국과학기술정보연구원은 과학기술정보 전문 유동 기관으로 국내 및 해외의 최근 연구 동향을 신속하고 체계적으로 잘 분류하여 제공하고 있다. 따라서 제대로 정제되지 않은 인터넷 매체로부터의 정보 수집 통로를 없애고 한국 과학기술정보연구원으로부터 최근 연구동향 정보를 수집하도록 변경하였다. 또한 한국과학기술정보연구원은 RSS를 통한 정보 배포 서비스를 하고 있었기 때문에, RSS를 이용하여 수집하도록 구현한

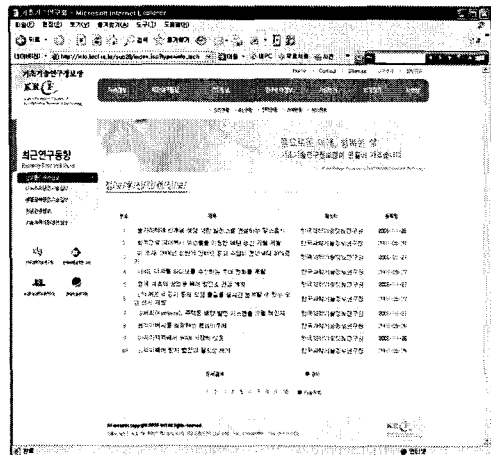
과학기술정보 연계시스템을 적용하였다.



▶▶ 그림 3. 한국과학기술정보연구원의 과학기술동향 RSS 서비스

적용한 결과로 한국과학기술정보연구원에서 체계적으로 분류되는 정보는 기초기술연구정보망에서 제공해야 하는 정보통신, 나노 소재 등의 분류체계에 적용하기가 무척 쉬워졌으며, RSS를 이용한 연계이므로 정보 제공 소스의 웹 사이트 변경에 얽매이지 않게 되었다. 본 시스템의 기초기술연구정보망 시스템에 적용된 이후, 한국과학기술정보연구원의 과학기술정보 제공 사이트가 변경된 적이 있었으나 기초기술연구정보망 시스템의 수집 모듈에는 영향을 끼치지 않았다. 또한, 제목과 생성일자 등 필요한 콘텐츠 외에 화면 디자인이나 내용 등과 같은 직접적으로 필요 없는 정보는 받지 않아도 되므로 수집 모듈이 간단해졌으며, 시스템에 주는 부하도 줄어드는 효과가 있었다.

최근 연구동향을 제외한 뉴스 등 다른 정보들은 아직 출연(연)들이 RSS를 통한 정보를 제공하지 않아 기존과 같은 HTML파싱을 계속 수행하고 있다. 향후 출연(연)에 RSS 배포 모듈을 설치하여 모든 정보를 RSS를 통한 방법으로 연계할 예정이다.



▶▶ 그림 4. RSS를 통해 수집된 기초기술연구정보망 시스템의 최근연구동향 정보

## V. 결론

정부 출연 연구기관에서 제공하는 다양한 정보는 과제관리 기관, 연구회 등 관련기관의 시스템과 연계되어 서비스되어야 한다. 지금까지는 자체적으로 정의한 프로토콜이나 출연(연)이 제공하는 HTML 결과를 파싱하여 연계하여왔다. 이러한 방법은 새 시스템의 추가와 같은 확장을 어렵게 하고 기존 시스템 변경에 대한 대처가 어려웠다.

RSS는 자주 갱신되는 정보를 자동적으로 쉽게 사용자들에게 제공하기 위해 개발된 것으로 표준화되어 있으며 간단하여 시스템 연계가 매우 쉽고 구조적인 XML 방식으로 콘텐츠를 표현하므로 시스템 변경에 효과적으로 대처할 수 있다.

본 논문에서는 이러한 RSS를 이용하여 과학기술정보 연계 시스템을 설계하고 기초기술연구정보회의 연구정보망 수집 시스템에 적용하였다. 이렇게 함으로 RSS의 장점인 확장성과 변경에 대한 대처의 용이성이라는 두 가지 효과를 거두었다.

향후 연구과제로는 보안에 민감한 평가정보, 연구성과 정보 등에 대해 RSS에 보안 모듈을 추가하여 연계하는 방법을 고려해보아야 한다. 또한, RSS 배포 모듈을 출연(연)에 설치하여야 하는데 출연(연)과 연구회의 밀접한 협력 방안에도 생각해 보아야 할 것이다.

### ■ 참고 문헌 ■

- [1] 김중태, “웹 2.0시대의 기회-시멘틱 웹”, 디지털미디어 리서치, 2006
- [2] 권이남, 김재수, 신동구, 전성진, “RSS 기반 과학기술정보 배급 표준시스템 구현에 관한 연구”, 한국콘텐츠학회 추계종합 학술대회 논문집, Vol3, No.2, 2005
- [3] “<http://www.xml.com/pub/a/2006/02/22/rom-parse-publish-rss-atom-feeds-java.html>”, ROME