

# RSS 기반 과학기술정보 수집시스템의 개발과 활용방안에 관한 연구

## A Study on the Implementation and Service of the RSS Based Science and Technology Information Aggregation System

현미환, 이상환, 이태석, 예용희  
한국과학기술정보연구원

Hyun mi-hwan, Lee sang-hwan, Lee tae-seok,  
Yae yong-hee\*\*\*  
Korea Institute of Science and Technology  
Information

### 요약

웹 사이트에서 제공하는 콘텐츠의 일부 또는 전체를 다른 서비스에서 이용할 수 있도록 해주는 신디케이션 기술인 RSS(Really Simple Syndication)는 다양한 웹 사이트 상의 콘텐츠를 요약하고, 상호 공유하고 주고받을 수 있도록 만든 표준이다. 본 논문에서는 과학기술 연구자들이 웹상에 존재하는 신디케이션 된 과학기술 콘텐츠(또는 feed)를 효과적으로 이용할 수 있도록 주제를 중심으로 수집(aggregation)하여 저장(archive)하는 시스템을 구현하고 이의 활용방안을 제안한다. 또한 제안된 중앙의 수집기(aggregator) 시스템에 사용자가 직접 채널을 등록함으로써 다른 사용자에게 RSS 채널을 추천하고 이를 통해 연구자간의 지식 공유를 유도할 수 있다.

### Abstract

The RSS which is a kind of syndication technology enables a part or the whole of the content which it provides from web site to be able to be used in different service. It is protocol to summarize and share various contents on web site. This Paper propose the ways to implement system to aggregate RSS feed and archive it and the services using it. Also, the users register the channel RSS to the RSS aggregator system and recommend it. It can drive to share of knowledge between researchers.

## I. 서론

RSS는 정보서비스 기관과 이용자 간의 푸쉬(push) 메일 정보제공 방식 서비스를 대체할 수 있는 새로운 패러다임의 정보 배급(syndication) 기술으로써 이용자가 정보를 요청할 때 제공하는 이용자 측면의 풀(pull) 방식 기술이다. 이렇게 정보 제공자가 아닌 정보 이용자의 요구와 필요에 따라 정보의 선택권이 부여되는 정보서비스 환경의 변화에 따라 KISTI에서는 2005년부터 “과학기술 맞춤형정보”, “KISTI의 과학향기”, “해외 과학기술동향정보”, “과학기술동영상세미나정보”의 4가지 분야에 대해 RSS 배급시스템을 개발하여 이용자에게 RSS 콘텐츠 서비스를 제공하기 시작했다. 과학기술 분야의 지식 정보는 다양한 정보소스가 수시로 발생하며, 이에 따른 복잡하고 다양한 서비스는 상호 운용성을 유지하면서 유연하게 대응할 수 있어야 한다. 다양하고 전문적인 과학기술 분야 지식정보를 RSS 기반의 서비스로 제공하기 위한 시스템의 구현은 주제 분야별 과학기술 분류 및 정보의 생성주기, 과학기술정보 서비스의 특성을 고려해야 한다.

본 연구에서는 기존의 KISTI RSS 표준 배급시스템(STI-RSS)을 RSS 수집기(aggregator) 시스템과 연계하였고, 웹

기반의 RSS 구독기(reader) 모듈을 통합함으로써 RSS 배급, 수집, 구독의 전 과정을 하나의 시스템으로 수행할 수 있는 패키지 형태의 RSS 수집시스템을 개발하였다. 또한, 국내외에서 제공되고 있는 유용한 RSS 채널을 주제 중심으로 수집 및 분류하여 최신의 과학기술정보를 이용자들에게 제공하고 있는 RSS 기반의 과학기술정보 포털사이트(rss.yeskisti.net)의 활용 사례를 통해 RSS aggregator의 서비스 활용 방안을 제안하고, 사용자들에게 과학기술 RSS 채널 추천 기능을 제공함으로써 이용자의 참여를 유도하고 연구자간의 지식을 공유할 수 있는 방안을 제안하였다.

## II. 관련 연구

### 1. RSS 개요

RSS는 “RDF Site Summary”, “Really Simple Syndication”, “Rich Site Summary” 등의 약자로 사용되며, RDF(Resource Description Framework) 기반의 콘텐츠 배급 프로토콜이다. 즉, 다양한 웹 사이트 상의 콘텐츠를 요약하고, 상호교환 할 수 있도록 만든 표준 포맷이다[1]. RSS 표준

은 네스케이프사에 의해 처음 개발되어, 현재 블로그 커뮤니티에 가장 많이 적용되고 있으며, 가장 널리 활용되고 있는 공개된 웹 콘텐츠 신디케이션(Syndication) 방식이다. 이는 웹 정보 제공자 측에서 새로운 정보의 갱신 여부를 알려주는 용도로 쓰인다.

[표 1] RSS의 버전별 특징

버전	제작	특징
0.90	Netscape	RSS 초기버전
0.91	UserLand	간단한 구조
0.92	UserLand	0.91보다 풍부한 메타데이터 제공
0.93		
0.94		
1.0	RSS-DEV Working Group	RDF를 기반 모듈을 통한 확장성 단일 벤더에 의해 조정되지 않음
2.0	UserLand	모듈을 통한 확장성이 있고 0.9x버전에서 쉽게 이전 가능

현재까지 RSS 표준은 (표 1)에서와 같이 RSS 0.9, RSS 0.91, RSS 0.92, RSS 1.0, RSS 2.0 등의 총 7개의 버전이 있으며, 가장 많이 활용하고 있는 표준으로는 RSS 2.0, RSS 0.91, RSS 1.0 순이다. RSS 표준의 표현 방식은 Channel 아래에 title(제목), link(링크URL), description(간략내용), language(언어) 등의 여러 개의 item으로 구성되며 웹 사이트의 정보를 간략하게 표현할 수 있는 구조이다. RSS 파일만을 상호교환하게 되면 최신 정보의 신디케이션을 간단하게 구성할 수 있을 뿐만 아니라, RSS 구독기가 있는 웹 이용자라면 즐겨 찾는 사이트를 각각 방문하지 않고도 RSS 채널(Channel)의 URL 정보만을 등록하는 것으로 웹 사이트의 최신 정보를 하나의 구독기 화면에서 한꺼번에 확인할 수 있다. 여기서 RSS 채널은 RSS를 지원하는 사이트의 URL 정보이며, RSS 피드는 RSS 표준 형태로 제공하는 각종 정보를 말한다[2].

## 2. 웹 수집기

웹 수집기란 웹 문서 내부에 포함된 하이퍼링크를 따라 이동하면서 웹상의 정보를 수집하는 소프트웨어이다. 웹 수집기가 탐색하고자 하는 대상이 웹 전체가 아닌 특정 관심 분야에 대한 검색일 경우 문서에 나와 있는 모든 링크를 대상으로 탐색할 필요는 없다. 이 경우 특정 주제와 문서와의 연관성에 따라 탐색할 링크의 수를 줄여 나가는 수집기를 주제 기반 수집기(topic-specific crawler or focused crawler)라고 한다[3]. 주제 중심 수집기의 기본 알고리즘은 웹문서를 단어의 집합으로 보고 페이지와 주어진 특정 주제와의 연관성을 코사인 유사도 측정으로 계산해 내는 방식으로, 이러한 방식의 우수성은

이전 연구에서 밝혀졌다[4]. 이전의 주제 중심의 수집기에 대한 연구에서는 아직 방문하지 않은 웹 페이지들이 특정한 주제와 어느 정도 연관이 있는가를 식별하기 위하여 해당 웹 페이지의 내용만을 이용하여 분류를 수행하였고, 웹 페이지간의 링크 정보는 사용하지 않는다[5].

## III. RSS 기반 과학기술정보 수집시스템 개발 및 활용방안

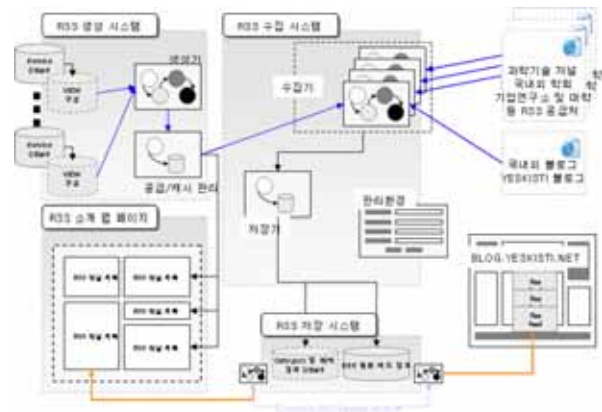
### 1. RSS 서비스 대상 콘텐츠 조사 및 분석

RSS aggregator를 이용한 RSS 기반의 과학기술정보 포털 사이트를 구현하기 위해 RSS 채널로 구성하여 RSS 서비스가 가능한 국내의 과학기술정보 관련 콘텐츠를 조사하였으며, 해당 채널을 그룹핑 하여 분류로 구성할 수 있는 분류체계를 조사하였다. 국내에서는 과학기술정보에 대해 RSS 서비스를 체계적으로 제공하는 곳이 없어 KISTI가 보유한 내부 콘텐츠 중에서 이용자에게 최신의 정보를 알리는 메일링서비스를 중심으로 RSS 형태로 배급되는 콘텐츠 체계를 보완하고, 정보를 접근하는 접근 채널로써 콘텐츠를 조사하였다. 또한, 외부 과학기술 관련 RSS 채널로부터 피드를 수집하고, 수집된 해당 피드에 분류를 부여하여 제공함으로써 정보의 유형과 접근성을 제고하고자 범용성과 표준을 감안하여 본 연구에서는 과학기술부에서 제시한 과학기술과 가장 관련성이 높은 과학기술표준분류로 수집할 정보의 주제 분류 체계를 선정하였다.

### 2. RSS 수집기(aggregator) 시스템 개발

#### 2.1 RSS 수집기(aggregator) 서비스 구성

RSS aggregator 시스템에 대한 서비스 구성도는 (그림 1)과 같다. RSS aggregator 시스템은 크게 수집기(aggregator), 저장기(repository), 관리환경을 설정하는 관리기(management system)의 크게 세 가지 부분으로 나눌 수 있다.



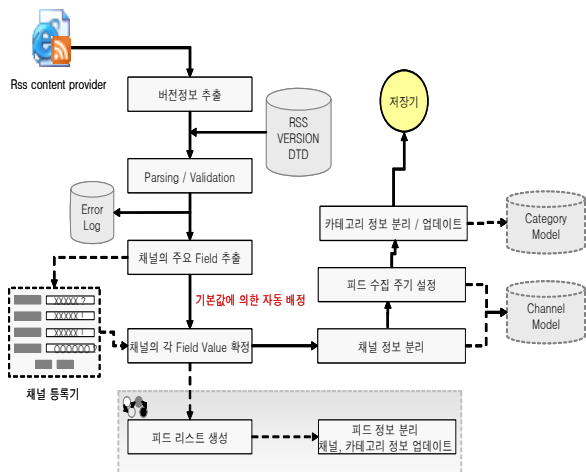
▶▶ 그림 1. RSS 수집기(aggregator) 시스템 구성도

수집기는 해당 RSS 채널의 스케줄링 정책에 따라 일정 시  
간에 주소를 방문하여 RSS 페이지를 읽은 후, RSS의 버전 정  
보를 체크하고, 주요 필드별 정보를 추출하는 기능을 담당하  
다. 저장기는 수집 RSS 채널에 대한 정보, 카테고리 정보, 디  
렉토리 정보를 저장 또는 변경하는 기능을 담당한다. 마지막으  
로 관리기는 채널에 대한 정보를 저장 및 삭제할 수 있다.

2.2 주제중심 RSS 수집기(aggregator) 설계

주제 중심 RSS 수집기는 웹상에 존재하는 RSS 채널 주소  
를 찾아서 수집해 온다. 외부의 과학기술 관련 RSS 채널로부  
터 피드를 수집하여 과학기술표준분류체계에 따라 주제별로  
분류하고, 이를 웹상에서 디렉터리 서비스를 하기 위하여 데이  
터 모델링을 수행하였다. 데이터 모델링은 과학기술표준분류  
에서 사용하는 분류 정보를 담고 있는directory 테이블, 각 분  
류별 주제 분야의 정보를 담고 있는 category 테이블, 각 채널  
과 피드들의 분류 정보 담고 있는 테이블과 수집된 채널 정보  
와 피드 정보를 담을 수 있는 테이블이 필요하다.

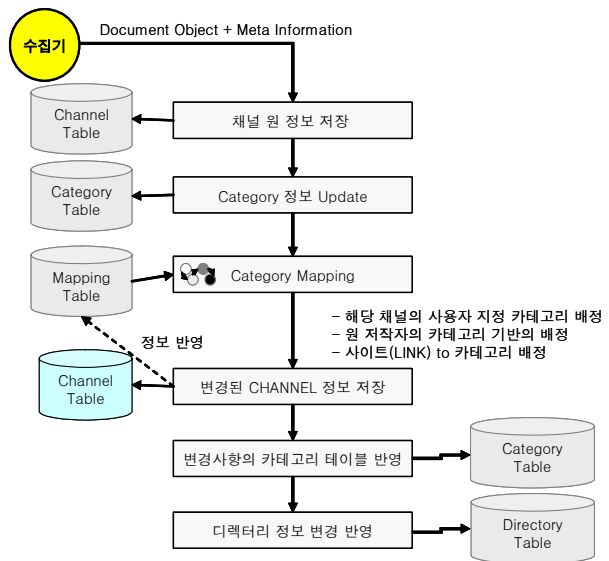
주제 중심의 RSS aggregator는 (그림 2)에서 보듯이 외부  
의 RSS 제공 사이트로부터 콘텐츠를 읽어 버전 정보를 추출  
하고 XML 파일을 파싱 하여 버전, 제목, 링크 정보 등의 기본  
적인 채널 정보를 시스템에서 자동으로 배정한다. 여기서 채널  
의 자동 분류 정보를 보완하기 위해 채널 등록기가 사용된다.  
채널 등록기에서는 채널의 주요 정보를 확인하고 설명 정보나  
이미지 정보 등을 추가하거나 수정할 수 있다. 또한 채널의 업  
데이트 주기 설정과 주제 분류 정보의 변경이 가능하다. 각 채널  
정보의 배정이 끝나고 채널이 등록되면 피드 수집을 위한  
별도의 프로세스에 의해 피드가 수집된다. 해당 채널의 피드  
수집 주기 설정이 완료되고 채널의 주제 정보를 분리하면 DB  
에 직접 저장하지 않고, Memory상의 모델에 저장하여 다음  
단계인 저장기로 전달한다.



▶▶ 그림 2. RSS 수집(aggregation) 프로세스

2.3 RSS 저장기(repository) 설계

RSS 채널을 등록하고 등록된 채널의 피드를 주제 중심으로  
수집하여 저장하기 위한 프로세스는 다음 (그림 3)과 같다. 그  
림에서 보듯이 RSS 채널을 분류하는 방식은 3가지를 사용하  
다. 첫 번째는 해당 채널의 사용자 지정 카테고리 배정 방식으  
로 과학기술표준분류체계에 따라 채널 등록시 사용자가 입력  
한 정보를 해당 채널에 강제 배정하는 방식으로, 이것은 수집  
대상(source)이 된 채널의 카테고리 정보를 참조하지 않을 수  
있다. 두 번째는 수집 대상이 된 원 저작자의 카테고리 기반의  
배정 방식으로, 원본 채널 및 해당 채널 소속 피드에 포함된  
카테고리에 따라 [채널 : 피드] 정보를 반영하여, 새로운 채널  
의 카테고리를 배정한다. 이 중 반복적으로 읽어오는 피드 정  
보에 포함된 해당 채널의 카테고리 정보를 반영하는 것이 복  
잡한 문제가 될 수 있다. 세 번째는 사이트(LINK) to 카테고  
리 배정 방식으로 원본 피드, 채널에 속한 LINK 필드의 값을  
참조하여, 특정 사이트, 특정 주소는 정해진 카테고리를 자동  
적으로 배정 받도록 한다. 이 경우 키워드를 매핑하여 채널 등  
록자에게 적절한 카테고리를 추천할 수 있다.



▶▶ 그림 3. RSS 저장(repository) 프로세스

3. RSS Aggregator 서비스 활용방안

3.1 과학기술 RSS 포털서비스

KISTI에서는 RSS aggregator를 통해 과학기술 관련 최신  
정보를 제공하는 RSS 채널을 수집 및 분류하여 서비스를 제  
공하는 과학기술 RSS 포털서비스(rss.yeskisti.net)를 제공하  
고 있다. 이러한 서비스를 통해 현재 해외 과학기술분야 RSS  
채널 800여개, 220만 건의 피드가 수집되어 및 제공되고 있다.  
RSS Aggregator를 통해 수집된 해외 과학기술정보 관련

RSS 피드를 이용자들에게 제공하기 위한 서비스 화면은 (그림 4)와 같다. (그림 4)에서 보듯이 RSS 포털서비스의 초기화면에서는 RSS Aggregator를 통해 수집된 RSS 서비스 현황은 전체 채널 및 피드에 대한 개수, 그리고 최근 수집된 피드들의 개수를 보여준다. 또한 좌측 메뉴에서는 수집된 채널의 분석하여 과학기술부에서 제정한 국가과학기술표준분류에 의거하여 분류한 분류가 메뉴로 구성되어 이용자가 원하는 분야의 정보를 선택할 수 있도록 하였다. 또한, KISTI가 자체 수집 및 분류하여 제공하는 RSS 채널 외에도 과학기술정보와 유용한 채널을 이용자 추천할 수 있도록 추천기능을 제공하였다. 이용자의 의해 추천된 RSS 채널은 관리자가 채널을 조사 및 분석하고, 분석결과에 따라서 서비스가 가능한 채널로 등록할 수 있다.



▶▶ 그림 4. 과학기술 RSS 포털서비스 화면

### 3.2 개인화서비스의 활용

본 연구에서는 RSS 수집시스템에 대한 이용자들의 편의와 니즈를 반영하기 위해 개인화 된 서비스로 웹 RSS 리더기와 RSS 콘텐츠를 블로그 등의 개인화 된 저장 공간으로 복사할 수 있는 스크랩 기능 등의 개인화 기능을 제안한다. 웹 기반의 RSS 리더기는 개인화 페이지에 RSS 채널을 등록하여 등록된 채널의 정보를 볼 수 있는 기능을 API 형태로 개발하고, RSS 피드의 스크랩 기능도 마찬가지로 API 형태로 구현하여 다양한 웹 인터페이스를 통해 활용될 수 있도록 하였다. KISTI에서는 이러한 서비스를 과학기술정보 포털사이트(yesKISTI.net)에 적용함으로써 과학기술 관련 지식의 공유와 확산 등 사이트의 이용 활성화에 기여하였다.

다음 그림은 과학기술정보 포털사이트(yesKISTI.net)에서 이용자가 원하는 외부 RSS 채널을 “채널추가” 메뉴를 선택하여 해당 채널을 등록하면 개인화 페이지로 저장되고, 개인화

페이지를 통해 최근에 수집된 RSS 피드 내용을 볼 수 있는 활용 사례를 보여주고 있다. 또한, RSS aggregator를 통해 수집되어 yesKISTI.net을 통해 서비스되고 있는 과학기술 RSS 채널도 “MY RSS 저장” 메뉴를 통해 선별하여 저장이 가능하다. 이렇게 MY RSS 페이지에 등록된 채널은 별도의 RSS 리더기 없이 웹사이트에 로그인하여 최근 수집된 피드를 즉시 받아볼 수 있고, 해당 콘텐츠를 자신의 블로그로 스크랩 할 수 있어 개인의 정보 저장소로 활용이 가능하다.



▶▶ 그림 5. RSS 개인화 페이지 화면

## IV. 결 론

정보서비스 제공기관에서 RSS는 이메일 제공 방식의 푸시(push) 서비스에 대한 대체 수단으로 주로 연구가 진행되어 왔다. 즉, RSS 서비스는 공급자적인 측면의 정보 제공 방식에서 이용자에게 최신 정보 구독에 대한 선택권을 부여하는 방식으로 정보 선택의 주도권을 이동시키는 변화를 가져왔다. 이러한 변화에 따라 RSS 기반의 과학기술 콘텐츠를 제공하는 웹 사이트가 증가하였고, 이용할 수 있는 RSS 채널은 많아졌지만 연구자들이 원하는 RSS 채널을 이용하기 위해서는 각 사이트를 직접 돌아다니며 내용을 판단하고 RSS 채널 주소를 찾아내야 하는 문제점이 있었다.

본 연구에서는 과학기술정보를 제공하는 국내의 RSS 사이트에 대한 조사, 분석을 수행하였고, 주제 중심 RSS 수집기를 통해 RSS 채널을 자동 분류하여 등록하는 시스템을 구현하였다. 또한 이러한 RSS aggregator를 통해 수집된 RSS 채널을 과학기술표준분류체계로 분류하여 서비스하는 RSS 기반의 과학기술정보 포털사이트(rss.yeskisti.net)의 활용 사례를 제안하였다. 이러한 과학기술 RSS 포털사이트는 RSS 배급, 수집, 저장 시스템이 연계되어 KISTI에서 생산하는 RSS와 외부 사이트에서 수집된 RSS 콘텐츠를 망라적으로 제공하고

있다.

KISTI에서는 과학기술정보에 대한 RSS 서비스를 지원함으로써 사용자들에게 최신 과학기술정보를 손쉽게 배포할 수 있게 되었고, KISTI의 콘텐츠를 구독하는 RSS 구독자를 통해 보다 많은 웹 사이트 방문을 유도할 수 있게 되었다. 뿐만 아니라, 정보 이용자 입장에서는 일일이 각 사이트를 방문하지 않고도 RSS 포털사이트를 통해 최근 갱신된 RSS 채널과 피드 정보를 확인함으로써 최신의 과학기술정보를 RSS 형태로 구독할 수 있고, 이것을 개인화 된 페이지에 저장할 수 있게 함으로써 개인의 정보 저장소로도 활용할 수 있게 되었다. 또한 RSS 포털사이트에서 서비스하고 있지 않은 RSS 채널을 사용자가 직접 추천하여 등록하게 함으로써 연구자간의 효율적인 정보 공유가 가능하다.

본 연구를 통해 KISTI에서는 RSS 포털사이트를 통해 해외 과학기술정보와 관련된 RSS 서비스를 제공하는 채널을 수집 및 분류하여 이용자들에게 편의를 제공하고, 최신의 정보를 제공함으로써 이용자의 니즈와 만족도를 제고하였다. 현재까지는 뉴스 등 특정분야의 정보를 RSS 서비스로 제공하는 것이 많이 활성화되어 있으나, 과학기술 분야의 지식정보를 망라적으로 서비스 하는 사례는 많지 않으므로 RSS를 통한 서비스를 준비하는 정보서비스 기관에는 좋은 사례가 될 것이다.

#### ■ 참고 문헌 ■

- [1] 전중홍, “컨텐츠 신디케이션 표준화 동향”, IT Standard Weekly, 제40권, 2004.
- [2] 권이남, 김재수, 신동구, 전성진, 정택영, “RSS 기반 과학기술정보 배급표준 시스템(STI-RSS)의 개발과 활용방안에 대한 연구”, 정보관리연구, 제36권, 제4호, pp.175-198, 2005.
- [3] 이영석, 조정원, 김준일, 최병욱, “주제 중심 수집기를 이용한 RSS 채널 추천 시스템”, 전자공학회 논문지, 제43권, 제6호, pp.52-58, 2006.
- [4] Soumen Chakrabati, “Focused crawling a new approach to topic-specific web resource discovery” In Proc. of 8th International World Wide Web Conference, 1999.
- [5] 서혜성, 최영수, 노상욱, 최경희, 정기현, “연관도를 계산하는 자동화된 주제 기반 웹 수집기”, 한국인터넷정보학회, 제7권, 제3호, pp.155-167, 2006.