RSS기반 과학기술정보 배급 표준시스템 설계에 관한 연구

권이남*, 김재수*, 신동구*, 전성진*, 정택영*, 박병희**
*한국과학기술정보연구원 표준화기술지원실,
**(주)에즈넷
e-mail:ynkwon@kisti.re.kr

A Study on the Design of Standard system for the Syndication of Science and Technology Information based on RSS

Lee-Nam Kwon*, Jae-Soo Kim*, Dong-Gu Shin*,
Sung-Jin Jhun*, Tack-Young, Chung*, Byong-Hee Park**
*Technical Support for Standardization Dep., Korea Institute of Science and Technology Information,
**ASNET Co.

요 약

최근 RSS(Really Simple Syndication)를 응용한 서비스가 해외는 물론 국내에서도 각 분야에서 다양하고 활발하게 연구 개발되고 있다. 국내외 RSS서비스 현황 및 사례조사를 통해 민간과 공공부문의 RSS 서비스 현황을 분석하고, 과학기술분야 지식정보의 특성을 고려하여 다양한 정보와 환경에 유연하게 대응할 수 있는 과학기술 지식정보 서비스를 위한 RSS기반 정보배급 표준시스템의 설계 방안을 제시하였다.

1. 서론

최근 2005년 9월 멕시코 노티맥스 통신은 인터넷 회사인 아이언포트 시스템보고서를 통해 한국이 미국과 중국에 이어 스팸 메일 발송국 세계 3위임을 알렸다. 이제 국내에서도 정보이용자의 의지와 상관없이 발송되는 각종 스팸 메일을 통한 공해는 심각한 수준에 이르렀음을 알 수 있다.

기존 정보제공기관 중심의 이메일 푸시(Push)방식 서비스의 문제점을 해결하기 위한 대안으로서, 이용자가 각 웹사이트를 방문하지 않고도 원하는 시점에 원하는 정보만을 한 화면에 제공받을 수 있는 풀(Pull)방식의 표준화된 포맷의 RSS(Really Simple Syndication)서비스가 부각되고 있다.

과학기술분야 지식정보는 다양한 정보 소스가 수시로 발생하며, 이에 따른 복잡하고 다양한 서비스는 상호운용성(Interoperability)을 유지하면서 유연하게 대응할 수 있는 시스템을 필요로 한다.

만약, 각각의 다양한 정보 채널별로 RSS 배급시스템이 표준화(Standardization)되어있지 않다면 과학기술분야의 다양한 환경에 따른 서비스별 중복 개발의 문제점 뿐만 아니라, 각 서비스별 콘텐트의 통합과 연계 및 상호호환, 채널별 통계의 통합 관리를 위해서 많은 시간과 노력을 투자해야만 하는 문제가 발생하게 된다.

다양하고 전문적인 과학기술분야 지식정보를 RSS 서비스를 통해 배급(Syndication)하기 위해서는 이러한 과학기술분야의 특성을 고려하여 보다 정교하고 표준화된 시스템 설계가 필요하다.

본 연구의 구성은 2장에서 대표적인 국내외 RSS 서비스 현황 및 사례조사로서 국내외 주요기관 및 민간포털 및 공공부문의 RSS 서비스 현황을 살펴보았으며, 3장에서는 RSS기반 과학기술정보 배급 표준시스템에 대한 설계방안을 제시하였다. 마지막으로 4장에서 결론을 제시하였다.

2. RSS 서비스 사례

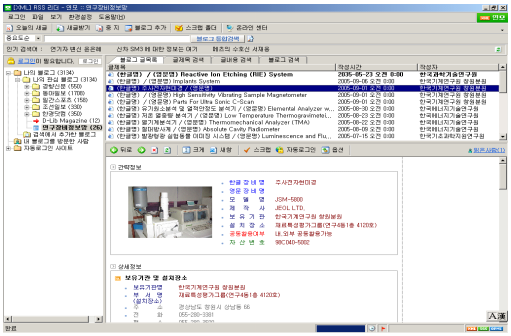
RSS를 이용한 서비스는 해외는 물론 최근 국내에서도 각 분야에서 다양하고 활발하게 연구 및 응용되고 있다. 최근 국내 RSS서비스로 대표적인 민간포털과 공공부문, 과학기술분야의 사례와 해외 주요 기관별 RSS서비스 동향과 RSS 적용 사례를 들고자 한다.

국내에서는, 초기에 포털 사이트 내에서 블로그끼리 정보교환을 위한 포맷으로 RSS를 활용하는데 그쳤으나, 2004년부터 뉴스 분야 사이트를 중심으로 RSS 정보배급 서비스를 본격적으로 시작하였다[1].

2005년에는 보다 다양한 부분에서 RSS를 응용하여 시범 적용하고 있다. 야후코리아의 '야후 피플링'은 RSS기능을 이용해 여러 블로거들이 자신이 쓰고 있는 글을 주제별로 따로 분류해 등록하고 해당 주제의 글만 한꺼번에 읽을 수 있도록 지원하고 있다. '다음 RSS넷'은 RSS기능을 활용해서 정보를 편집하고 구독할 수 있는 일종의 마이페이지 서비스를 제공하고 있으며, SK커뮤니케이션즈의 '통'도 주제에 따라 유사한 관심사를 가진 블로그들을 연결해 주는 RSS서비스의 일종이다[7]. 이는 블로그끼리 정보를 교환하고 연결하기위한 포맷으로서 RSS를 활용하고 있는 사례이다.

또 다른 RSS 응용 사례로는 RSS 피드를 특정 포털 내에서 만이 아닌 국내 전체 RSS 피드를 대상으로 검색해주는 '드림위즈 RSS 검색' 서비스가 있다. 이는 RSS피드들을 서비스 콘텐트로 규정하고 RSS 피드의 검색 및 신규 피드 정보를 제공해주는 일종의 RSS 포털 사이트라고 볼 수 있다.

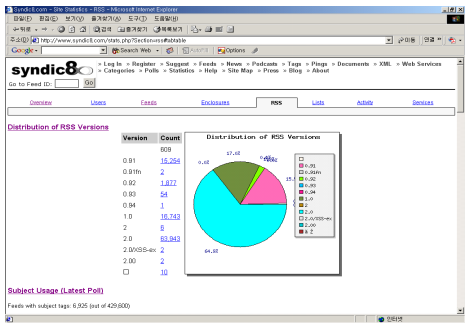
국내 공공부문에서의 RSS서비스로는 최근 한국기초과학지원연구원(www.keol.net)의 최신 연구장비 정보의 RSS피드 제공사례를 들 수 있다(그림 1).



(그림 1) 공공부문 RSS서비스 사례[3]

국내 과학기술분야 지식정보를 종합적으로 생산·유통하고 있는 한국과학기술정보연구원(KISTI)은 정보 소스가 다양하고 각 분야의 서비스별로 다양한 기술 분류를 활용하고 있는 특성 등을 감안하여 2004년부터 표준화된 RSS 배급시스템 개발에 관하여 연구하였으며, 2005년에 RSS기반 과학기술정보 배급 표준시스템(STI-RSS)을 개발중이며 이를 다양한 정보 서비스에 활용할 예정이다.

해외에서는, 이미 뉴스 분야 뿐만 아니라 다양한 분야에서 RSS서비스를 보편적으로 제공하고 있으며, RSS피드 및 이용자층에 대한 전문적인 분석을 제공하는 사이트(Syndic8, NewsIsFree, UserLand 등)는 물론 전문적인 RSS 피드 검색엔진(Feedster)을 제공하고 있다[2, 4].



(그림 2) RSS피드 버전별 분석사례(Syndic8.com)

즉, 세계적인 RSS피드의 현황 정보(피드 작성언어 분포, 접속 순위, 인기순위, RSS 버전 사용분포, 총 피드수, 총 이용자수)를 전문적으로 분석하여 통계 그래프를 제공하여 서비스하고 있다(그림 2).

마이크로소프트(MS)는 내년에 출시되는 윈도 차기 버전인 '비스타'에 RSS 지원기능을 추가할 계획을 발표하였다. 무엇보다, 기존 RSS버전에서 정의된 아이템(item)외에 <msdn : headlineImage>, <category domain>, <msdn : contentType> 등 MS만의 독자적인 아이템을 정의하여 익스플로러 화면에서 RSS 구독기(Reader) 없이도 피드별로 정의해놓은 카테고리 정보나 콘텐트타입 등을 볼 수 있도록 하고 있다.

과학기술분야의 전문 정보도 RSS를 이용하여 보다 다양하게 응용하여 서비스하고 있다.

중국은 중국 최대 과학기술 전문 정보 사이트인 '중국 과학기술 정보 네트워크(china.gov.cn)'와 중국내 최대 RSS 콘텐츠 및 서비스 제공 사이트인 '간천하(看天下)'가 RSS 검색 엔진 보급, 서비스 제공 분야에서 장기적인 협력관계를 구축하여 본격적인 RSS 서비스를 준비하고 있다[6].

해외 과학기술분야의 대표적 출판사인 Nature는 새로운 뉴스, 저널 기사 정보를 RSS포맷으로 제공하는 방식은 기본이고, 최신 저널의 목차(TOC)정보

를 관련 기관 및 이용자에게 RSS로 웹 및 모바일(mobile)통신으로도 서비스하고 있다. 또한, Nature도 기존의 표준 RSS 요소(item)외에 DC(더블린코어)와 PRISM 메타데이터 요소를 RSS 포맷에 추가하여 서비스하고 있다.

Nature외에도 저널의 TOC정보를 RSS로 서비스하고 있는 과학기술분야의 대표적인 출판사로는 BioMed(BMC), Extenza, OUP, American Journal of Neuroradiology, AIP(American Institute of Physics), BMJ(British Medical Journal), Medscape 등이 있다[5].

이처럼, RSS 이용이 확산되면서 RSS 포맷의 발전 방향에 대한 논의가 활발하게 진행중이며, 국제적인 표준 제정의 필요성이 증대되면서 'Atom'이라는 프로젝트로 구체화되었다. 현재 IETF의 AtomPub WG에서 출판 프로토콜과 신디케이션 포맷의 확장에 대한 표준화 작업이 진행 중이며, 'Atom 1.0'은 RFC 인터넷 표준이 되기 전 마지막 단계인 "제안 표준(proposed standard)"으로 인정받은 상태이다[8].

3. RSS기반 과학기술정보 배급 표준시스템 설계

RSS기반 과학기술정보 배급 표준시스템이라고 함은 정보이용자들에게 다양한 채널에서 제공되는 정보를 이용자가 일일이 웹사이트를 방문하지 않고 채널주소만 등록해서 언제든지 새로운 정보를 선택적으로 하나의 S/W로 다양한 채널과 함께 받아볼 수 있는 풀(Pull)방식의 정보제공시스템이며, 과학기술정보의 다양한 형태와 서비스 환경 변화를 유연하게 받아들여 이 표준시스템을 통해 언제든지 RSS로 배급 가능한 시스템을 말한다.

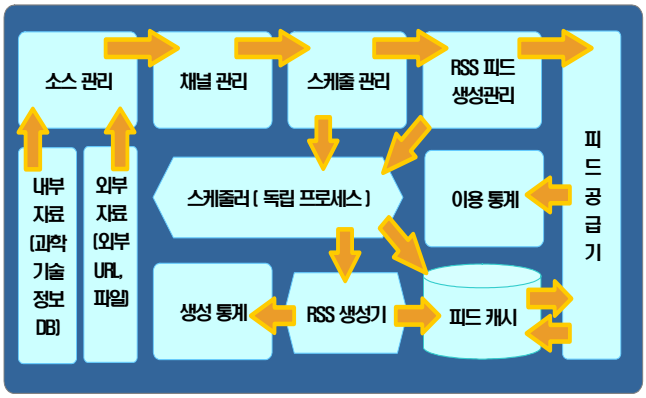
즉, 시스템 관리자가 언제든지 RSS 피드의 배급 시간이나 배급량, RSS버전, 해당 채널, 해당 소스, 채널 관리자 등을 구성할 수 있는 유연한 표준화된 시스템이 요구된다.

RSS기반 과학기술정보 배급 표준시스템(STI-RSS)의 설계 방향은 다음과 같다.

첫째, 과학기술정보의 다양한 정보 소스를 수용할 수 있어야 하고, 둘째, 무수히 많은 과학기술 분류를 수용하여 처리할 수 있어야 하며, 셋째, 세계적인 표준화의 흐름에 맞춰 채널별로 RSS의 다양한 버전으로 생성할 수 있어야 한다. 넷째, 피드의 배급 시간과 배급량을 설정 관리할 수 있어야하며, 다섯째, 피드의 생성과 이용에 관한 통합적인 통계정보를 제공하고 관리할 수 있어야 한다. 여섯째, 채널별 관리 뿐만 아니라 여러 채널의 통합적인 관리가 가능해야 한다.

따라서, 시스템 구성은 기본적으로 각 기능별 모듈화된 방식으로 설계하고, 시스템 운영시 발생하는 모든 중간 생성물을 XML형태로 정의하여 변환과 활용을 극대화하며, 다양한 운영환경에서 호환성이 보장되도록 설계되어야 한다.

이러한 주요 설계 방향을 근거로 시스템을 설계하였으며, (그림 3)은 이러한 요구조건을 반영한 RSS기반 과학기술정보 배급 표준시스템 구성도이다.



(그림 3) RSS기반 과학기술정보 배급 표준시스템 구성도

과학기술정보의 다양한 정보 소스를 수용하기 위한 '소스설정'은 과학기술정보의 입력 자료를 지정하는 것으로 과학기술정보 데이터베이스(오라클, My SQL 등), 파일, 외부URL, 외부RSS 채널 등 다양한 형태의 내부 및 외부 정보 소스를 수용하도록 설계하였다. 특정 조건만을 만족하는 피드를 생성하기 위해 피드 생성 조건 설정을 통해 동일한 소스더라도 해당 조건을 만족하는 정보만을 선별하여 피드로 생성할 수 있도록 하였으며, 설정 파일의 유효성 검사를 위해 '미리보기' 기능을 설계에 포함하였다.

계층적이며 세분화된 과학기술 분류를 수용하기 위해서는 무엇보다 기술 분류별 채널의 자유로운 구성과 채널 통합이 가능해야 한다. '채널관리'는 다양한 분류별로 유연하게 구성하기 위하여 개별 채널 생성 뿐만 아니라 단위 채널들을 그룹 채널로 묶어서 그룹단위로 한꺼번에 설정하고 관리할 수 있도록 설계함으로서, 다양한 과학기술의 주제별, 기술 분야별 그룹핑이 가능한 맞춤 채널을 구성할 수 있게 된다. 또한, 세계적인 표준화의 흐름에 맞춰 채널별로 다양한 RSS버전으로 생성할 수 있어야 하므로, '채널관리'는 새로운 채널 생성 단계에서 채널별로 채

널관리자 및 채널별 RSS버전(0.91, 0.92, 1.0, 2.0, ATOM 등)을 선택할 수 있도록 한다. 과학기술정보 RSS 피드가 한꺼번에 많은 량이 발생할 수 있으므로 피드의 생성과 배급량을 조절하기 위해, 피드의 생성시간, 피드 생성량 및 캐시운영(캐시기간 및 캐시량) 방법 등의 설정이 가능하다.

보다 정확한 피드의 생성과 배급을 위해 '스케줄 관리'를 통해 채널별, 그룹채널별, 피드 단위별로 피드 생성 스케줄을 만들고 관리할 수 있도록 하고, 개별 채널이나 그룹채널의 스케줄 외에 보다 정교한 생성 스케줄의 보완을 위해 피드단위 스케줄을 추가 배정이 가능하다.

'스케쥴러'는 채널별 그룹별 설정된 시간에 따라 피드 생성 신호를 'RSS생성기'에 보내고, '피드 생성관리'의 피드캐시 운영조건에 따라 캐시의 피드 삭제를 위한 신호를 보낸다. 'RSS생성기'는 이 스케쥴러의 생성 신호에 따라 각종 조건에 따른 피드를 생성하여 피드캐시에 저장한다.

피드의 보관을 위한 '피드 캐시'는 '스케줄 관리'를 통해 생성된 스케줄에 따라 미리 생성된 피드와 서비스 유효기간내의 피드를 보관하는 임시저장소의 역할을 한다.

생성된 피드가 제대로 배급되었는지와 캐시가 너무 많을 경우를 고려하여 '생성관리'에서는 피드의 생성과 배급의 운영 현황을 모니터링하며, 생성된 피드의 배급을 위해 캐시량을 조정할 수 있다.

'통계관리'는 단위 채널별, 그룹 채널별, 또는 전체 통계를 통합하여 제공하며, 피드의 생성과 배급・이용에 관한 통계정보도 제공이 가능하도록 설계하였으며, 각각을 기간별로 구분하여 그래프로 표현하고 엑셀파일로 다운로드 할 수 있다.

'권한 관리'는 그룹과 개별 채널에 대하여 메뉴별 읽기, 쓰기, 삭제, 수정 권한을 각각 부여할 수 있도록 하였고, 그룹별 관리, 단위 채널 관리, 또는 통합하여 관리할 수 있도록 관리자별 그룹 및 권한을 등록하여, 소속 그룹에 따라 권한을 부여하고, 관리자 개인에게 배정된 추가 권한이 있는 경우 해당 권한을 부여하도록 하였다.

앞에서 제시한 설계 방향에 따라 내・외부 다양한 정보 소스를 고려하였고, 계층적이고 세분화된 기술 분류별로 효율적인 서비스를 하기 위해 그룹핑이 가능한 그룹채널, 피드 생성기간과 배급양의 조절을 위한 피드 생성 스케줄, 피드의 생성과 배급 및 피드캐시의 모니터링, 채널별 또는 그룹채널별, 전체 통합관리가 가능한 통계관리, 다양한 관리자 권한 관리 등을 고려하여 설계를 하였다.

4. 결론

본 연구는 과학기술분야 지식정보의 다양한 정보 소스와 환경을 고려한 RSS기반의 표준화된 정보배급시스템에 관한 설계로서 과학기술분야의 다양한 서비스별 중복 개발 가능성을 미연에 방지할 수 있고, 무엇보다 표준화된 포맷을 사용함으로서 얻게 되는 콘텐트 재사용 및 시스템간 상호운용성, 통합관리 측면의 효과에서 매우 의미가 있다.

다양한 환경에도 유연하게 적용 가능한 엔진 역할을 하는 표준화된 시스템의 개발은 과학기술분야 지식정보에만 필요한 것은 아니라고 본다. 향후에는 좀더 다양한 분야에서 RSS를 기반으로 한 응용시스템 연구가 필요하다. 해외의 사례처럼 피드와 이용자에 대한 전문적인 분석과 함께 전문정보에 대한 RSS 피드도 정교하게 검색가능 한 한국형 RSS피드 검색엔진에 관한 연구도 필요하다. 또한, 과학기술분야의 다양한 국내외 RSS피드들을 수집하여 재가공하고 이를 통합적으로 제공해주는 서비스에 대한 연구도 필요하다.

참고문헌

[1] 권이남, 김명일, 김재수, 신기정 "효율적인 맞춤정보 서비스를 위한 RSS기반 개인화 지원 연구" 『KOSTI Workshop 2004 - 과학기술정보』, p.69-82.

[2] 권이남, 김재수 "학회정보 개인화 지원을 위한 RSS기반 유통 모델 연구", 『2004 한국콘텐츠학회 추계종합학술대회』, p.545-553.

[3] 구중억, "RSS 메타데이터 기반 SDI 시스템 구축방안 연구", 『제12회 한국정보관리학회 학술대회』, 2005, p.107-115.

[4] Ben Hammersley, "Content Syndication with RSS", O'Reilly & Associates, Inc, 2003.

[5] Tony Hammond, Timo Hannay, Ben Lund, "The Role of RSS in Science Publishing", 『D-Lib Magazine December 2004』, vol 10, no 12.

[6] http://www.chinainfo.gov.cn/data/200505/1_20050510_109960.html

[7] http://www.etnews.co.kr/news/detail.html?id=200502100049

[8] http://www.ietf.org/html.charters/atompub-charter.html