

RSS기반 과학기술정보 배급 표준시스템 구현에 관한 연구

A Study on the Implementation of Standard System for the Syndication of Science and Technology Information Based on RSS

권이남, 김재수, 신동구, 전성진
한국과학기술정보연구원 표준화기술지원실

Kwon Lee-Nam, Kim Jae-Soo, Shin Dong-Gu,
Jhun Sung-Jin
Technical Support for Standardization
Department,
Korea Institute of Science and
Technology Information

요약

최근 RSS(Really Simple Syndication)를 응용한 서비스가 해외는 물론 국내에서도 각 분야에서 다양하고 활발하게 연구되고 있다. 과학기술분야 정보의 RSS서비스는 다양한 정보소스와 서비스 환경에 유연하게 적용할 수 있어야 하며 이러한 특성을 고려한 표준화된 시스템을 필요로 한다. 본 논문에서는 현재 국내의 RSS 서비스 사례조사를 통해 RSS 서비스 현황을 분석하고 설계에 반영하였으며, 이를 토대로 과학기술정보의 특성을 고려한 RSS기반 과학기술정보 배급 표준시스템의 구현 방안을 제시하였다.

Abstract

Research on RSS(Really Simple Syndication) based application and services are being deployed both domestically and internationally. The RSS service in the area of science and technology should be flexible to the various sources of information as well as the various platform and environments that they are being serviced. It also requires a standardized system to meet the above needs. We have analyzed the use of RSS services overseas as well as domestic services to suggest the implementation of standard system for the syndication of science and technology information based on RSS.

I. 서론

최근 멕시코 노티맥스 통신(2005년 9월)은 인터넷 회사 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 방문

하지 않고도 원하는 시점에 원하는 정보만을 한 화면에 제공받을 수 있는 풀(Pull)방식의 표준화된 포맷의 RSS(Really Simple Syndication)서비스가 부각되고 있다. RSS서비스의 장점은 정보이용자가 자신이 관심 있는 홈페이지를 일일이 방문하거나 자신의 개인정보를 제공하지 않고도 여러 개의 관심 있는 홈페이지에서 발생하는 신규 정보를 하나의 화면에서 스팸 메일 없이 한꺼번에 볼 수 있다는 것이다.

과학기술분야 지식정보는 다양한 정보 소스가 수시로 발생하며, 이에 따른 복잡하고 다양한 서비스는

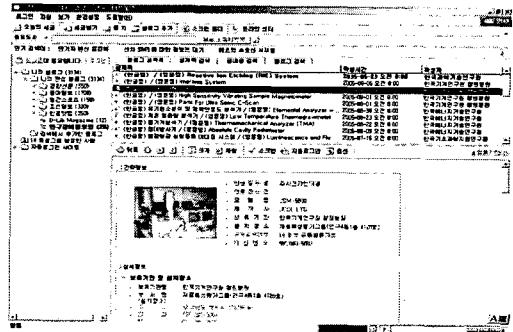
상호운용성(Interoperability)을 유지하면서 유연하게 대응할 수 있는 시스템을 필요로 한다. 만약, 각각의 다양한 정보 채널별로 RSS 배급시스템이 표준화(Standardization)되어있지 않다면 과학기술분야의 다양한 환경에 따른 서비스별 중복 개발의 문제점뿐만 아니라, 각 서비스별 콘텐츠의 통합과 연계 및 상호호환, 채널별 통계의 통합 관리를 위해서 많은 시간과 노력을 투자해야만 하는 문제가 발생하게 된다. 다양하고 전문적인 과학기술분야 지식정보를 RSS 서비스를 통해 배급(Syndication)하기 위해서는 이러한 과학기술분야의 특성을 고려하여 보다 정교하고 표준화된 시스템의 설계와 구현이 필요하다. 시스템 구현시에는 다양한 과학기술지식정보의 특성을 고려하여 단위채널을 그룹으로 묶어주는 그룹채널과 단위채널간의 종속관계가 피드를 생성하기 위한 스케줄 관리에서도 동일하게 적용되도록 유지하는 것과 스케줄에 따라 중복 생성될 수 있는 피드를 방지하기 위한 방안을 제시할 필요가 있다.

본 연구의 구성은 2장에서 대표적인 국내외 과학기술분야 RSS 서비스 사례조사로서 국내외 과학기술분야 주요기관 및 공공부문의 RSS 서비스 현황을 살펴보았으며, 3장에서는 RSS기반 과학기술정보 배급 표준시스템에 대한 설계를 하고 4장에서 구현 방안을 제시하였다. 마지막으로 5장에서 결론을 제시하였다.

2. 국내외 과학기술분야 RSS 서비스 사례

RSS를 이용한 서비스는 해외는 물론 최근 국내에서도 각 분야에서 다양하고 활발하게 연구 및 응용되고 있다. 여기에서는 최근 국내외 과학기술분야에서 대표적인 RSS서비스 사례와 표준화 동향을 살펴보았다. 국내에서는, 초기에 포털 사이트 내에서 블로그끼리 정보교환을 위한 포맷으로 RSS를 활용하는데 그쳤으나, 2004년부터 뉴스 분야 사이트를 중심으로 RSS 정보배급 서비스를 시작하였다[1]. 2005년에는 보다 다양한 부분에서 RSS를 응용하여 시범 적용하고 있는데 블로그끼리 정보를 교환하고 연결하기 위

한 포맷으로서 RSS를 활용하고 있거나, RSS피드를 서비스 콘텐츠로 규정하고 RSS 피드의 검색 및 신규 피드 정보를 제공해주는 일종의 RSS 포털 성격의 사이트도 생겨났다. 국내 과학기술분야 공공부문에서의 최근 RSS서비스로는 단일 정보 소스이기는 하지만, 한국기초과학지원연구원(www.keol.net)의 최신 연구장비정보의 RSS피드 제공 사례를 들 수 있다[그림 1].



▶▶ 그림 1. 공공부문 RSS서비스 사례[2]

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 개발하였고 이를 다양한 정보 서비스에 활용할 예정이다.

해외에서는, 이미 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[3]. 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[1].

해외의 과학기술분야 전문 정보 역시 RSS를 이용

하여 보다 다양하게 응용하여 서비스하고 있다. 중국 최대 과학기술 전문 정보 사이트인 '중국 과학기술 정보 네트워크'와 중국 내 최대 RSS 콘텐츠 및 서비스 제공 사이트인 '간천하(看天下)'가 장기적인 협력관계를 구축하여 본격적인 RSS 서비스를 준비하고 있다[5].

과학기술분야의 대표적 출판사인 Nature Publishing Group(NPG)은 새로운 뉴스, 저널 기사 정보를 RSS포맷으로 제공하는 방식은 기본이고, 최신 저널의 목차(TOC)정보를 관련 기관 및 이용자에게 RSS로 웹 및 모바일(mobile)통신으로도 서비스하고 있다. 또한, 기존의 표준 RSS 요소(item)외에 DC(Dublin Core)와 PRISM 메타데이터 요소를 RSS 포맷에 추가하여 서비스하고 있다.

NPG외에도 저널의 TOC정보를 RSS로 서비스하고 있는 과학기술분야의 대표적인 출판사로 BioMed, Extenza, OUP(Oxford University Press), American Journal of Neuroradiology, AIP(American Institute of Physics), BMJ(British Medical Journal), Medscape 등이 있다[4]. 이처럼, RSS 이용이 확산되면서 국제적인 표준 제정의 필요성이 증대되어 'Atom'이라는 프로젝트로 구체화되었고, 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[6].

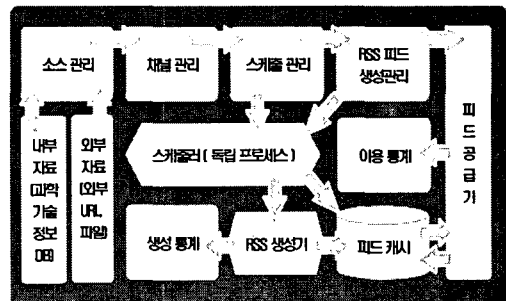
3. RSS기반 과학기술정보 배급 표준시스템(STI-RSS) 설계

RSS기반 과학기술정보 배급 표준시스템(STI-RSS)이라고 함은 정보이용자들에게 다양한 채널에서 제공되는 정보를 이용자가 일일이 웹사이트를 방문하지 않고 채널주소만 등록해서 언제든지 새로운 정보를 선택적으로 하나의 S/W로 다양한 채널과 합

계 받아볼 수 있는 풀(Pull)방식의 정보제공시스템이며, 과학기술정보의 다양한 형태와 서비스 환경 변화를 유연하게 받아들여 이 표준시스템을 통해 언제든지 RSS로 배급 및 통합 관리가 가능한 시스템을 말한다. 즉, 시스템 관리자가 언제든지 RSS 피드의 배급시간이나 배급량, RSS버전, 해당 채널, 해당 소스, 채널 관리자 등을 구성하고 관리할 수 있는 유연한 표준화된 시스템이어야 한다.

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 배급시간과 배급량을 설정 관리할 수 있어야 하며, 다섯째, 피드의 생성과 이용에 관한 통합적인 통계정보를 제공하고 관리할 수 있어야 한다. 여섯째, 채널별 관리뿐만 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 극대화하였다. 이러한 주요 설계 방향을 근거로 시스템을 설계하였으며, [그림 2]는 이러한 요구조건을 반영한 RSS기반 과학기술정보 배급 표준시스템(STI-RSS)의 구성도이다.



▶▶ 그림 2. STI-RSS 시스템 구성도

과학기술정보의 다양한 정보 소스를 수용하기 위한 '소스관리'는 과학기술정보의 입력 자료를 지정하는 것으로 과학기술정보 데이터베이스, 파일, 외부RSS 채널주소 등 다양한 형태의 내부 및 외부 정보 소스를 수용하도록 설계하였다. 계층적이며 세분화된 과학기술 분류를 수용하기 위해서는 무엇보다 기술 분류별 채널의 자유로운 구성과 채널 통합이 가능해야 하는데, '채널관리'는 다양한 분류별로 유연하게 구성하기 위하여 개별 채널 생성뿐만 아니라 개별 채널들을 그룹 채널로 묶어서 그룹단위로 한꺼번에 설정하고 관리할 수 있도록 설계함으로써, 다양한 과학기술의 주제별, 기술 분야별 그룹핑이 가능한 맞춤형 채널을 구성할 수 있게 된다.

보다 정확한 피드의 생성과 배급을 위해 '스케줄관리'를 통해 채널별, 그룹채널별, 피드 단위별로 피드 생성 스케줄을 만들고 관리할 수 있도록 하고, 개별 채널이나 그룹채널의 스케줄 외에 보다 정교한 생성 스케줄의 보안을 위해 피드단위 스케줄을 추가 배정하여 설계하였다. '스케줄러'는 채널별 그룹별 설정된 시간에 따라 피드 생성 신호를 'RSS생성기'에 보내고, '생성관리'의 피드캐시 운영조건에 따라 캐시의 피드 삭제를 위한 신호를 보낸다. 'RSS생성기'는 스케줄러의 생성 신호에 따라 각종 조건에 따른 피드를 생성하여 피드캐시에 저장하고, '피드 캐시'는 생성된 스케줄에 따라 미리 생성된 피드와 서비스 유효기간 내의 피드를 보관하는 임시저장소 역할을 한다. 이것은 과학기술정보의 생성주기를 고려해 '피드 캐시'를 통해 가끔 방문하는 이용자에게도 최신 정보 뿐만 아니라 캐시에 저장된 이전 정보 제공도 가능하게 한다.

생성된 RSS 피드가 제대로 배급되었는지와 캐시가 너무 많을 경우를 고려하여 '생성관리'에서 생성 스케줄에 따른 피드의 생성과 배급 및 피드캐시의 운영 현황을 모니터링할 수 있도록 설계하였다. '통계관리'는 단위 채널별, 그룹 채널별, 또는 전체 통계를 통합하여 제공하며, 피드의 생성과 이용에 관한 통계정보를 제공하도록 설계하였다. '권한 관리'는 그룹과

개별 채널에 대하여 메뉴별 권한을 각각 부여할 수 있도록 하고, 그룹채널 및 단위 채널, 또는 통합 관리할 수 있도록 관리자별 그룹과 권한을 지정하여, 소속 그룹에 따라 권한을 부여하도록 설계하였다.

4. STI-RSS 구현

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

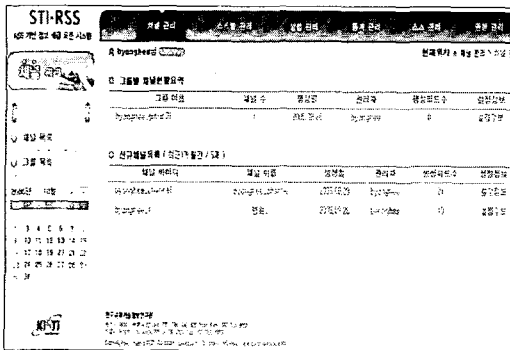
RSS기반 과학기술정보 배급 표준시스템(STI-RSS)의 구현은 리눅스 환경에서 JSP와 Java 언어를 사용하여 구현하였으며, DB는 오라클과 My SQL로 구현하였다. 구현된 시스템의 기능구조도는 [표 1]과 같고, 초기화면은 [그림 3]과 같다.

피드가 생성되어 제공되는 프로세스 흐름은 RSS 피드를 생성하고자하는 소스를 먼저 선택하고, 선택된 소스에서 과학기술 분류별 그룹채널 또는 단위채널을 만든 다음 피드 생성을 위한 스케줄을 생성한다. 피드생성 스케줄에 따라 생성기를 통해 생성된 피드는 피드캐시에 저장되고 이용자가 접속하여 요청시 제공하게 되며, 이용 내역을 통계관리에서 확인하는 순서로 진행된다.

[표 1] STI-RSS 기능구조도

주기능	서브기능	설명
소스 관리	소스 목록	등록된 소스 목록 관리
	신규 소스 등록	소스(DB, 파일, RSS주소)등록
채널 관리	소스 목록	채널 생성 및 관리
	신규 소스 등록	그룹 채널 생성 및 관리
스케줄 관리	그룹 스케줄	그룹채널 스케줄 생성 및 관리
	채널 스케줄	단위채널 스케줄 생성 및 관리

	피드 스케줄	피드단위 스케줄 생성 및 관리
	시스템 관리	시스템 타이머 운영 관리
생성 관리	피드 생성 현황	피드 생성 현황 모니터링
	피드 공급 현황	피드 공급 현황 모니터링
	캐시 운영 관리	채널 캐시현황 확인 및 관리
통계 관리	통계 보기	그룹별, 채널별, 스케줄별 피드 생성 및 공급 통계 확인
권한 관리	관리자 목록	그룹 및 채널 관리자 등록 및 관리
	권한 목록	등록된 관리자의 메뉴별 권한 관리



▶▶ 그림 3. STI-RSS 시스템 초기화면

4.1 소스관리

소스관리는 RSS 피드 생성을 위한 첫 단계로서 과학기술정보DB(오라클, My SQL), 파일, RSS 채널 주소의 신규 소스 등록과 함께 이미 등록된 소스 목록을 관리한다. 해당 소스에 대한 접속정보의 유효성 검사를 통해 오류를 최소화하고, 소스설정의 편의성을 높이기 위해 기술적인 설정값을 미리 XML파일로 만들어놓고 지정하면 소스등록을 위한 전체 설정에 반영되도록 하였다. 피드의 중복 생성을 방지하기 위해 별도의 고유 식별값을 소스에 배정하여 동일 채널에 같은 피드가 생성되는 것을 방지한다.

4.2 채널관리

채널관리는 단위채널의 그룹핑이 가능한 그룹채널과 개별채널, 피드단위채널의 목록을 생성하고 관리한다. 그룹채널과 단위채널간의 종속관계를 부여함으로써 그룹에 대한 각종 설정정보가 해당 그룹에 속한 채널을 신규 생성할 때 기본적으로 사용되도록 하여

채널별로 재설정해야 하는 번거로움 없이 일관성을 유지하도록 한다.

세계적인 표준화의 흐름에 맞춰 다양한 RSS버전으로 생성할 수 있도록, 채널 생성시 채널별로 채널 관리자, RSS버전(0.91, 0.92, 1.0, 2.0, ATOM 0.2, 1.0) 선택이 가능하다. 특정 조건만을 만족하는 피드를 생성할 수 있도록 피드 생성조건 설정을 통해 동일한 소스더라도 해당 조건만을 만족하는 정보를 선별하여 피드로 생성할 수 있고, 유효성 검사를 위해 '미리보기' 기능을 제공한다. 과학기술정보 RSS 피드가 한꺼번에 많은 량이 발생할 수 있으므로 피드의 생성과 배급량 조절을 위한, 피드 생성시간, 피드 생성량 및 캐시운영 방법(캐시시간, 캐시건수)을 설정하고 관리한다.

4.3 스케줄관리

스케줄관리는 그룹채널 스케줄, 개별채널 스케줄, 피드단위스케줄의 신규 생성 및 관리와 함께 스케줄러의 동작시간과 간격을 관리한다. 그룹스케줄을 작성하면, 해당 그룹에 속한 모든 채널의 스케줄도 동시에 생성된다. 피드단위 스케줄은 특정 조건에 해당하는 개별 피드를 추가하거나 제외하는 기능을 제공한다. 스케줄러의 수행을 위한 타이머를 서비스 데몬으로 독립 구성하여, 타이머의 주기적인 신호에만 피드생성 프로세스를 수행하도록 구현함으로써 지속적인 스케줄 체크의 부하를 줄일 수 있다.

4.4 생성관리

생성관리는 생성된 피드의 배급을 위한 캐시량을 조정하도록 피드의 생성현황 및 공급현황, 캐시사용현황을 확인하고 운영관리하며, 지정한 기간별로 그래프를 제공한다. 매일 발생하는 정보도 있지만 비정기적으로 발생하는 과학기술정보의 양과 주기를 고려하여 피드 캐시시간과 량을 조정하도록 채널별 모니터링 정보를 제공한다.

4.5 통계관리

통계관리는 그룹채널별, 개별채널별, 스케줄별 피드 생성과 공급 통계뿐만 아니라, 전체 통합 피드의 생성과 공급, 에러 통계를 확인한다. 통계는 각 기간 별로 구분하여 그래프로 표현하고 엑셀파일로 다운로드 할 수 있다.

통계그래프 생성 기능은 플래시(Flash) 기술을 활용하였으며, 시스템 연동시 XML파일로 입력하는 방식과 자바스크립트로 직접 콜(Call)하는 방식을 제공한다.

4.6 권한관리

권한관리는 관리자목록에서 신규관리자를 등록하고 권한목록에서 등록된 관리자가 그룹채널 및 단위 채널별 각 서브 메뉴별로 읽기, 쓰기, 삭제, 수정 등의 다양한 권한을 부여하고 관리한다. 관리자 그룹별(그룹관리자, 채널관리자)로 묶어서 한 번에 권한을 부여하고, 개별 관리자 별로도 권한을 부여할 수 있도록 하기 위해서 관리자별 메뉴별 권한 테이블을 매트릭스 형태로 정의하여 구현하였다.

정보배급시스템은 과학기술분야 외에 다양한 분야의 지식정보에도 활용할 수 있도록 적용 방안 연구가 필요하다.

■ 참고 문헌 ■

- [1] 권이남, 김명일, 김재수, 신기정 "효율적인 맞춤형 정보 서비스를 위한 RSS기반 개인화 지원 연구", KOSTI Workshop 2004 - 과학기술정보, pp.69-82, 2004.
- [2] 구중익, "RSS 메타데이터기반 SDI시스템 구축방안 연구", 한국정보관리학회 학술대회, pp.107-115, 2005.
- [3] Ben Hammersley, "Content Syndication with RSS", O'Reilly & Associates, Inc, 2003.
- [4] Tony Hammond, Timo Hannay, Ben Lund, "The Role of RSS in Science Publishing", D-Lib Magazine December 2004, Vol.10, No.12.
- [5] http://www.chinainfo.gov.cn/data/200505/1_20050510_109960.html
- [6] <http://www.ietf.org/html.charters/atompub-charter.html>

5. 결론

본 연구는 과학기술분야 지식정보의 다양한 정보소스와 환경을 고려한 RSS기반의 표준화된 정보배급 시스템의 구현에 관한 연구로서 과학기술분야의 다양한 서비스별 중복 개발 가능성을 미연에 방지할 수 있고, 무엇보다 표준화된 포맷을 사용함으로써 얻게 되는 콘텐츠 재사용 및 시스템간 상호운용성, 통합관리 측면의 효과에서 매우 의미가 있다. 본 STI-RSS 시스템에서는 외부 소스가 데이터베이스인 경우, 오라클 및 MySQL에 대해서만 지원하지만 향후 JDBC 기반의 표준 인터페이스를 제공하는 모든 데이터베이스로 확장이 필요하고, 스케줄별 조건과 운영환경에 따른 시스템 성능평가도 필요하다. 본 표준화된