

SDI 서비스를 위한 동적 RSS 채널 생성기 개발

A Study on dynamic RSS channel generator development to support Selective Dissemination of information Service

현미환, 이상기*, 유수현**, 예용희***
한국과학기술정보연구원 * ** ***

Hyun mi-hwan, Lee sang-gi*, Yoo su-hyeon**,
Yae yong-hee***
Korea Institute of Science and Technology
Information * ** ***

요약

선택적 정보 제공인 SDI(Selective Dissemination of Information) 서비스는 이용자가 탐색 전략을 입력하여 서비스를 신청하면 데이터베이스가 갱신될 때마다 자동으로 이용자에게 공지되는데, 본 연구에서는 이러한 SDI 서비스를 XML 포맷인 RSS로 제공하기 위해 동적 RSS 채널 생성기를 개발하였다. RSS는 웹사이트에 갱신되는 정보를 신속히 전달-교환하게 해주는 도구로서 많은 정보를 갱신과 동시에 받아볼 수 있게 해준다. 이러한 RSS를 활용함으로써 이용자는 미리 신청한 맞춤형서비스 뿐만 아니라 키워드 검색에 대한 검색결과 간략화면을 XML포맷의 RSS로 받아봄으로써 웹사이트를 방문하지 않고도 실시간으로 재검색 결과를 확인할 수 있다.

Abstract

SDI(Selective Dissemination of Information) is on demand service to inform to the user applies a service after inputting a search strategy, when new information is obtained. This Paper propose the ways to develop the dynamic RSS channel generator for supporting SDI service. 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. 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

과학기술 분야의 지식 정보는 다양한 정보 소스가 수시로 발생하며, 필요한 정보로의 즉각적인 접근은 정보 수요자들에게 매우 중요한 기능으로 인식되고 있다. 이런 필요성에 따라 오래전부터 문헌정보학 분야에서는 선택적 정보배포 서비스라 불리는 SDI(Selective Dissemination of Information)서비스를 수행하여 왔다. 국내의 대표적인 학술정보서비스인 KISTI의 NDSL(National Digital Service Library)은 NDSL 알리미서비스를 통해 연구개발 종사자들에게 SDI 서비스

를 제공해 왔는데, 직접 홈페이지나 정보검색 시스템에 접근할 필요 없이 요구한 정보를 주기적으로 수신하여 관심 기술의 최신 발전 동향을 신속하고 정확하게 파악할 수 있도록 하는 유용한 서비스이기 때문이다. 하지만 최근 메일링서비스는 스팸 메일 등과 같은 고객의 의도와 관계없이 제공되는 푸시 메일의 한계점으로 이용자가 감소하고 있어, 이러한 메일링서비스를 대체할 수 있는 서비스로 이용자가 정보를 요청할 때 제공하는 사용자 측면의 풀(pull) 방식 정보 배급 (syndication) 기술인 RSS는 핵심적인 기술이 되었다. RSS는 정보 제공자가 아닌 정보 이용자의 요구와 필요에 따라 정보의

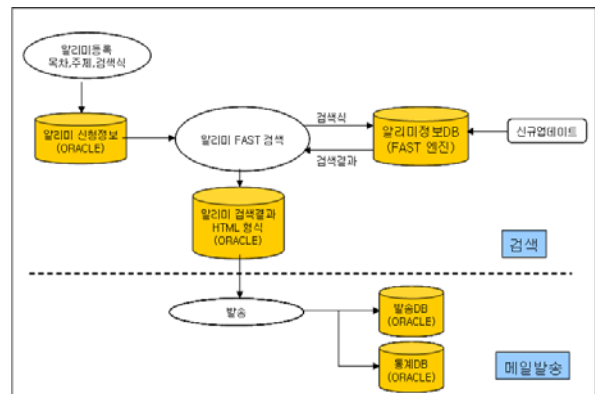
선택권이 부여되고, 수동적이었던 이용자는 정보의 생산자가 되면서 정보의 개별적인 선택과 활용, 그리고 공유를 용이하게 해준다.

KISTI의 알리미서비스는 이용자가 웹사이트에서 저널 목차정보나 검색어를 입력하여 서비스를 신청하면 신규 입수된 논문이나 특허 정보를 이메일로 보내주는데, 본 연구에서는 KISTI의 알리미서비스를 위해 사용되는 이용자의 프로필 정보를 통해 신규 정보가 입수될 때마다 동적인 RSS 피드를 생성하는 채널생성기를 개발하였다. RSS를 통해 제공되는 SDI 서비스는 실시간으로 입수된 정보를 볼 수 있고, IE 6.0 버전 이상에서는 별도의 리더기 없이 웹 브라우저를 통해 바로 구독이 가능하기 때문에 이용자가 편리하게 이용할 수 있다.

II. 관련 연구

1. SDI 서비스 과정

SDI 서비스 과정은 크게 자료입수, 정보검색, 정보제공 단계로 나누어 볼 수 있다. 자료입수는 사용자 정보 입수와 서비스 대상 데이터 입수로 구분되는데 사용자 정보 입수는 이용자의 요구정보를 프로파일로 생성하여 저장하는 과정이다. 정보검색 단계에서는 각 이용자가 설정한 검색 형식으로 자료검색을 수행하는데, 여기서는 정보검색을 최소화하기 위하여 이용자가 신청한 목차정보와 검색식 중 중복을 제거하고 유일한 주제(Unique item)에 대한 검색 결과를 신착정보 알리미 발송 DB 테이블에 HTML 형태로 저장하여 동일한 주제를 신청한 이용자들이 DB 정보를 공유할 수 있다. 정보제공 단계에서는 사용자 프로파일의 정보에 따라 알리미 발송용 DB 테이블에 저장된 HTML 형태의 검색결과를 메일로 구성하여 발송하는 단계이다. 다음 그림은 KISTI에서 제공하고 있는 SDI 서비스인 NDSL 알리미 서비스가 제공되는 과정을 보여주는 그림이다.



▶▶ 그림 1. SDI 서비스 과정

2. NDSL 알리미서비스

KISTI의 알리미서비스로 제공되는 자료는 학술지 및 프로시딩의 목차정보, 특허 및 실용신안의 초록정보, 최신 과학기술동향 및 첨단지식정보 등이다. KISTI의 알리미서비스는 NDSL 사이트를 통해 과학기술 관련 주제 및 잡지명을 목록에서 선택하거나 키워드를 이용하여 주제를 등록하면, 해당되는 새로운 정보가 KISTI DB에 입수될 때 이용자의 E-mail로 매주 화요일과 목요일 주2회 자동 발송된다.

서비스 메뉴 구성 화면은 다음 그림과 같이 과학기술 관련 학술잡지 3만 여종의 최신호에 대한 목차(TOC) 정보를 신청할 수 있는 목차정보 신청, 논문의 DDC(특히 IPC) 분류로 관련 주제 분야의 최신 유망 아이템과 기술에 대한 정보를 제공하는 주제별 정보신청, 과학기술 일반/정책, 물리, 화학, 우주과학 등 19개 과학기술분야에 대한 최신 동향과 첨단지식정보를 제공하는 동향분석 정보 신청, AND, OR 등의 연산자를 사용하여 검색식을 직접 작성하여 등록하면 해당하는 키워드에 대한 최신 논문, 특허 정보를 제공하는 정보검색 전문가를 위한 검색식 등록신청 기능을 제공한다.

개인정보

번호	등록일	자료종류	삭제
1	2009.04.24	디지털도서관 = Digital library	[삭제]
2	2009.04.24	한국도서관 정보학회지 = Journal of Korean Library and Information Science Society	[삭제]
3	2009.04.24	정보관리연구 = Journal of Information Management	[삭제]
4	2009.04.24	Current research in library & information science	[삭제]
5	2009.04.24	Library and information research news	[삭제]

주제어휘

번호	등록일	자료종류	주제어휘	키워드	삭제
1	2008.09.22	특허정보	0100-00100 0100-00102 0100-00104 0100-00300 0100-00304 0100-00302 0100-00500 0100-00700 0100-00702	특허	[삭제]
2	2008.04.25	논문	Computer Interfaces and Communications	user interface	[삭제]
3	2008.04.25	논문	Computer Programming, Software Engineering	computer	[삭제]
4	2008.04.24	논문	Programming for Specific Types of Computers	digital library	[삭제]
5	2008.04.24	논문	Computer Programs	digital library	[삭제]

검색어

번호	등록일	자료종류	검색어	삭제
1	2009.11.06	논문	TI (무분류) (37건)	[삭제]
2	2009.11.06	논문	TI (computational) and TI (education)	[삭제]
3	2009.09.19	논문	TI (정보) (47건)	[삭제]
4	2009.07.25	논문	TI (시스템)	[삭제]
5	2009.07.25	논문	TI (차량) (무분류)	[삭제]
6	2009.07.25	논문	시계열 (OH, OR, (의정문))	[삭제]
7	2009.07.25	논문	TI (무분류) and ASI (국가별)	[삭제]
8	2009.06.08	특허정보	CY (oxid) AND (나노)	[삭제]
9	2009.04.24	논문	(digital AvI library) or (전자도서관) or (디지털도서관)	[삭제]
10	2009.04.24	특허정보	(특허종 or 발명문) and (연세사) or (연세사)	[삭제]

▶▶ 그림 2. SDI 이용자 프로파일 등록 인터페이스

이용자가 등록한 정보는 검색엔진이 분석 가능한 쿼리를 변환되어 이용자별 프로파일 DB로 저장되고, 이것은 논문이나 특허 등의 정보가 신규로 업데이트 될 때 검색엔진을 통해 검색결과를 HTML 형태로 변환하여 알리미 발송용 개인별 DB에 저장되고, 스케줄링 된 시간에 메일이 발송된다.



▶▶ 그림 3. 메일을 이용한 NDSL 알리미서비스

III. 동적 채널생성기 개발

본 연구에서 제안하는 동적인 RSS 채널 생성기는

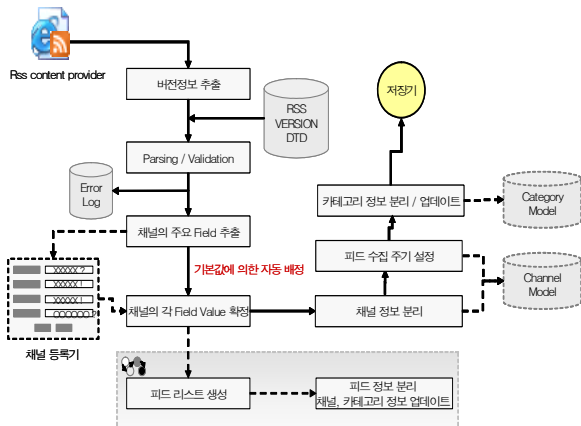
두 가지 형태의 SDI 서비스가 가능한데, 첫 번째는 RSS 피드 형태로 제공되는 콘텐츠에 대한 SDI 서비스이고, 두 번째는 신규 등록된 논문이나 특허에 대한 키워드 검색 결과를 RSS 피드 형태로 생성해 주는 동적인 RSS 채널 서비스이다. RSS 피드 형태로 제공되는 콘텐츠를 동적인 RSS 채널로 생성하여 서비스하기 위해 본 연구에서는 국내외 과학기술 RSS 피드 수집시스템을 제안한다. RSS 수집시스템은 웹상에 존재하는 RSS 채널 주소를 찾아서 수집해 오는데, 수집된 피드를 SDI 서비스로 제공하기 위해서는 피드에 대한 키워드와 주제 분야를 추출하는 과정이 필요하다. 키워드 검색결과를 동적인 RSS 채널로 생성해 주는 과정은 NDSL을 통해 제공되고 있는 알리미서비스에 대한 이용자 프로파일 정보를 동적인 RSS 채널로 생성하여 이용자는 RSS 채널 주소로 접근하여 프로파일에 대한 최근 등록된 논문이나 특허 정보를 RSS 피드 형태로 구독이 가능하게 된다.

1. RSS 맞춤형서비스를 위한 동적채널 생성기

제안된 RSS 수집시스템은 <그림 4>에서 보듯이 외부의 RSS 제공 사이트로부터 콘텐츠를 읽어 버전 정보를 추출하고 XML 파일을 파싱 하여 버전, 제목, 링크 정보 등의 기본적인 채널 정보를 시스템에서 자동으로 배정한다. 여기서 채널의 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 업데이트 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 DB에 직접 저장하지 않고, Memory상의 모델에 저장하여 다음 단계인 저장기로 전달한다.

이러한 수집시스템을 통해 수집되어 저장된 국내외 과학기술 RSS 채널과 피드에 대해 이용자가 검색 연산자가 포함된 검색 키워드를 등록하면, 이용자별 프로파일로 저장되고 동적 채널생성기는 프로파일 ID에 대한 동적인 RSS 채널을 생성하게 된다. 이를 통해 이용자는 자신만의 RSS 채널을 부여받게 되고, IE6.0 이상의 웹 브라우저나 RSS 리더기를 통해 할당 받은 RSS 채널

의 피드 정보를 조회함으로써 수집된 RSS 피드에 대한 맞춤형 정보를 실시간으로 받아 수 있는 SDI 서비스가 가능하다. 이러한 RSS 맞춤형서비스를 등록할 때 사용자는 동적인 채널에 대한 채널명을 부여할 수 있고, 채널의 분류에 대한 선택이 가능하다. 동적 채널 생성기는 FAST 검색엔진을 통해 RSS 피드를 검색하고, 이러한 결과를 실시간으로 가져 오며 RSS 채널을 동적으로 부여한다.



▶▶ 그림 4. RSS 수집 프로세스

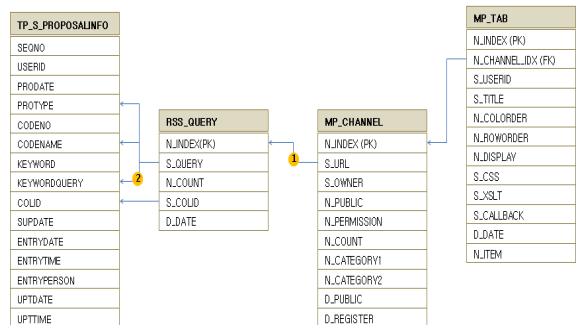
이용자가 맞춤 RSS서비스를 신청하면 제목, 카테고리, 키워드, 사용자아이디를 파라미터로 받아 검색 쿼리를 만들고, 쿼리에 대한 Key를 생성하여 채널 생성 Rule에 의해 Key가 조합된 동적인 채널을 생성하게 된다. 이렇게 신규로 생성된 채널은 동적 채널을 관리하는 DB 테이블에 등록되어 사용자가 RSS 리더기를 통해 채널을 구독할 때 실시간으로 검색하여 피드를 생성한다. 본 연구에서는 RSS 맞춤형서비스의 키워드 등록시 검색 연산자를 지원하고, 대량의 피드 데이터에 대한 빠른 검색을 가능하게 하기 위해 FAST 검색엔진을 연동하였다. FAST 검색 엔진 연동을 위해 표와 같이 RSS 채널과 피드 테이블에 대한 색인을 설계하고, 기존 채널과 피드에 대한 일괄 색인 작업을 진행하였다. 이후에는 색인 데몬 프로그램을 통해 주기적인(일 1회) 색인이 가능하다.

표 1. RSS 피드 검색을 위한 검색엔진 색인 설계

필드명	설명	FAST 적재필드	FAST 색인필드	정렬 항목
N_INDEX	고유번호	○	○	
S_TITLE	피드제목	○	○	○
S_DESC	피드요약	○	○	
S_LINK	피드링크	○	○	
S_PUBDATE	피드생성일	○		
S_IMAGE	이미지URL	○		
S_FEED	내용(XML)	○		
N_HIT	총 조회수	○		
N_CHANNEL_INDEX	채널INDEX	○	○	
D_DATE	피드수집일	○	○	○

2. 알리미서비스를 위한 동적채널 생성기

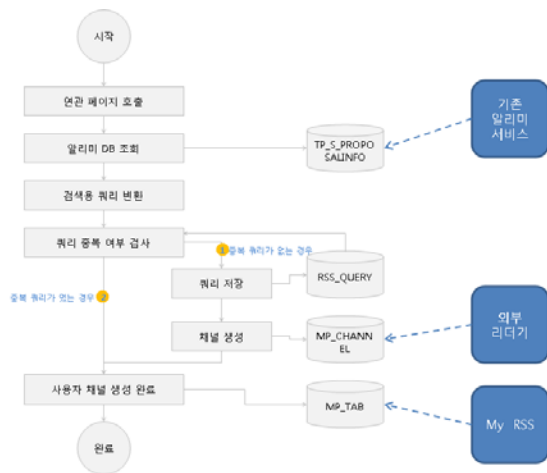
신규 등록된 논문이나 특허에 대한 키워드 검색 결과를 RSS 피드 형태로 생성해 주는 동적인 RSS 채널 서비스는 NDSL을 통해 메일로 제공되고 있는 알리미서비스를 위한 사용자 프로파일 정보를 활용하였다. NDSL 이용자는 알리미서비스를 등록할 때 저널의 목차정보, 특허의 IPC 주제분류, 혹은 검색연산자가 포함된 검색식을 입력하여 프로파일을 생성하게 되는데, 이 때 저장된 프로파일에 대한 Key 값을 이용하여 RSS 채널을 생성하게 된다. 다음 <그림 5>는 RSS 기반의 알리미서비스 구현을 위해 사용된 사용자 프로파일과 동적인 RSS 채널 생성을 위한 DB 테이블 스키마 명세다.



▶▶ 그림 5. 사용자 프로파일 기반 동적 채널 ER-Diagram

RSS 기반의 알리미서비스를 위해서는 사용자의 알리미 프로파일 DB 테이블을 이용하여 사용자가 등

록한 검색식과 목차정보, 주제정보를 가지고 동적인 RSS 채널을 생성하는 과정이 필요하다. 다음 그림은 이용자가 알리미서비스를 신청하여 프로파일을 등록할 때 동적인 RSS 채널이 생성되는 과정을 보여준다. 기존의 메일링서비스용 NDSL 알리미서비스에서 사용하는 사용자 프로파일 정보를 담고 있는 테이블을 사용하였고, 이것을 FAST 검색엔진용 쿼리로 변환하여 저장한다. 이용자는 자신이 등록한 프로파일에 대해 RSS 채널 주소를 부여받게 되는데, 이 주소를 통해 이용자가 피드를 읽는 시점에서 저장된 FAST 검색엔진용 쿼리를 통해 검색을 수행하여 결과를 가져오게 된다.



▶▶ 그림 6. RSS 채널 생성 프로세스

위 그림에서 보듯이 알리미 프로파일 정보를 가지고 RSS 채널 주소를 동적으로 할당할 때 채널 주소의 중복을 방지하기 위해서 FAST 검색 쿼리의 중복 체크를 하게 된다. 이때, 중복 쿼리가 없는 경우(1)에는 동적 URL 지정 방식을 통해 RSS_QUERY에 등록된 해당 레코드의 N_INDEX를 이용하여, "http://rss.ndsl.kr/feed.jsp?sURL=[N_INDEX]" 와 같은 형식의 채널 주소가 S_URL에 저장된다. 중복된 쿼리가 이미 존재하는 경우(2)에는 검색용 쿼리 정의 방식에 따라 알리미서비스를 위한 프로파일 소스 테이블(TP_S_PROPOSALINFO)의 PROTYPE, COLID 의 값에 따라 논문, 특허정보, 동향정보 및 저널로 구분된 RSS 생성을 위한 FAST 검색용 쿼리문을 생성한다.

IV. 결론

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 동적인 RSS 채널생성기를 개발하여 이용자가 키워드를 통해 선택한 콘텐츠를 자신만의 RSS 채널로 부여받아 RSS 피드 형태로 구독할 수 있는 시스템을 제안하였다. 또한 RSS 수집시스템을 통해 수집된 RSS 채널과 피드를 사용자가 키워드 등을 통해 자신만의 채널로 생성할 수 있는 동적인 RSS 채널서비스와 메일링서비스로 제공되는 신착정보 알리미서비스를 동적인 RSS 채널로 생성하여 서비스 할 수 있는 방안을 제안하였다.

제안된 시스템을 통해 KISTI에서는 NDSL을 통해 제공하는 신착 자료 메일링서비스를 RSS 기반의 알리미 서비스로도 제공하고 있으며, 웹기반의 RSS 리더기를 통해 신착정보를 웹브라우저에서 바로 확인해 볼 수 있는 서비스를 제공하고 있다. 현재까지는 뉴스 등 특정 분야의 정보를 RSS 서비스로 제공하는 것이 많이 활성화되어 있으나, 과학기술 분야의 지식정보를 RSS 형태로 받아보는 사례는 많지 않으므로 RSS를 통한 서비스를 준비하는 정보서비스 기관에는 좋은 사례가 될 것이다.

■ 참고 문헌 ■

- [1] 신수미, 이태석 "SDI(Selective Dissemination of Information) 서비스", 한국정보과학회 학술발표논문집, 제31권, 제1호, pp.172-174, 2004.
- [2] Ben Hammersley, "Content Syndication with RSS", O'Reilly & Associates, 2003

- [3] 권이남, 김재수, 신동구, 전성진, 정택영, “RSS 기반 과학기술정보 배급표준 시스템(STI-RSS)의 개발과 활용방안에 대한 연구”, 정보관리연구, 제36권, 제4호, pp.175-198, 2005.
- [4] 김선태, 김재훈, 한희준, 현미환, 이태석, 예용희, 여일연, 윤희준, “Pubmed 메타데이터 수집기 및 동적 RSS 채널 생성기 개발 연구”, 한국콘텐츠학회 학술 발표 논문집, pp.832-835, 2008.