

SDI(Selective Dissemination of Information) 서비스

개선방안 연구

신수미^o 이태석

한국과학기술정보연구원 정보서비스실

{sumi^o, tsyi}@kisti.re.kr

A Study on Improvement plan in Selective Dissemination of Information Service

Sumi Shin^o Taeseok Yi

Dept. of Information Marketing, Korea Institute of Science and Technology Information

요 약

선택적 정보 배포의 의미를 갖는 SDI(Selective Dissemination of Information)서비스는 이용자가 탐색 전략을 입력하여 서비스를 신청하면 데이터베이스가 갱신될 때마다 또는 이용자가 지정한 간격으로 자동으로 질의되어 그 결과를 이용자에게 제공하는 서비스로 정의되는데, 이때 대용량의 데이터베이스로부터 복잡한 데이터 질의과정을 거친 데이터가 다수의 사용자들에게 보다 빠르게 전달할 수 있도록 하는 것이 SDI 서비스의 주요한 요구사항이 되고 있다. 본 연구는 다수의 대용량 서비스 데이터베이스를 대상으로 이용자의 요구 주체에 해당하는 정보를 빠르게 서비스하기 위하여 검색시간을 단축할 수 있는 방안으로 'SDI 서비스 전용 데이터베이스'와 '검색결과저장 데이터베이스'를 별도로 두는 시스템을 제시한다. 이를 통해 동일한 질의에 대한 데이터베이스 검색이 최소화되고 검색대상 정보의 범위가 축소되며, 아울러 서비스 대상 데이터베이스의 확장도 용이해진다.

1. 서 론

최근의 인터넷 기술은 웹사이트를 통해 원하는 자료를 물리적, 시간적 제약 없이 획득할 수 있게 해준다. 하지만 정보획득 절차에 있어서는 필요한 자료가 있을 때마다 자료제공이 가능한 서비스에 접속하여 필요한 정보를 검색해야 하는 불편함이 여전히 개선되어야 할 부분으로 남아있다. 특히 정보의 갱신속도와 수량이 기하급수적으로 증가하는 현재와 같은 정보의 홍수시대에서 필요한 정보로의 즉각적인 접근은 정보 수요자들에게 매우 중요한 기능으로 인식되어 왔으며 이런 필요성에 따라 오래 전부터 문헌정보학 분야에서는 선택적 정보배포 서비스라 불리는 SDI(Selective Dissemination of Information) 서비스를 수행하여 왔다. SDI 서비스는 특히 연구개발 종사자들이 많이 활용하였는데, 직접 홈페이지나 정보검색시스템에 접근할 필요 없이 요구한 정보를 주기적으로 수신하여 관심 기술의 최신 발전 동향을 신속하고 정확하게 파악할 수 있도록 하는 유용한 서비스로 인식되어 왔기 때문이다. SDI 서비스는 과거 전통적으로 관리자가 수작업으로 개개인의 프로파일에 따라 오프라인 인쇄물을 서비스하는 방식에서부터 출발하여 최근에는 서비스의 신청에서부터 검색기능, 이용자별 신청정보를 맞춤형 e-mail로 제공하는 기능까지 모든 과정이 자동화 기능으로 발전하여 왔다. 즉, 이용자가 웹사이트에서 필요한 정보주체를 입력하여 서비스를 신청하면 신청정보는 데이터베이스를 검색할 수 있는 프로파일로 변환되어 저장되고, 데이터베이스가 갱신될 때마다 또는 이용자가 지정한 간격으로 프로파일을 참고하여 해당 주체를 위한 질의가 대상 데이터베이스에 실행되며, 그 결과는 이용자

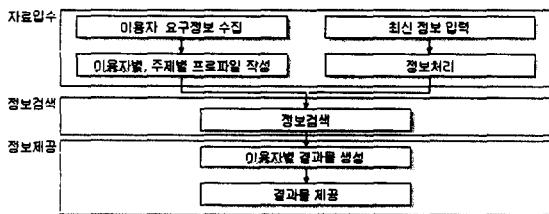
에게 e-mail 등 다양한 매체를 통해 제공되는 방식으로 서비스 일련의 과정이 모두 자동화 된 것이다. SDI 서비스가 발전하고 서비스 이용자와 대상 데이터도 함께 증가하면서, 많은 대상 데이터로부터 복잡한 질의가 신속히 시행되고 또한 검색된 다량의 데이터가 다수의 사용자들에게 빠르게 전달될 수 있도록 하기 위하여 정보검색이나 결과제공 성능의 지속적인 개선이 SDI 서비스 시스템의 주요한 요구사항이 되어 왔는데 특히 여러 이용자들이 신청한 다수의 질의를 대용량 데이터베이스에서 검색하는데 많은 시간이 소요되면서 이 시간을 획기적으로 단축할 수 있는 시스템 설계와 개발이 절대적으로 필요하게 되었다. 본 연구는 정보검색과 결과를 제작에서의 시스템 효율을 증대시키기 위하여 정보검색의 범위와 횟수를 축소할 수 있는 방안을 제시하고 그에 대한 시스템 설계안을 소개한다.

2. 구축시스템

2.1 SDI 서비스 과정

SDI 서비스 과정은 크게 자료입수, 정보검색, 정보제공단계 등으로 나누어 볼 수 있다. 자료입수는 이용자 정보입수와 서비스대상 데이터 입수로 구분되는데 이용자 정보입수는 이용자의 요구정보를 프로파일로 생성하여 저장하는 과정이다. 대상 데이터 입수는 SDI 서비스를 위한 데이터를 데이터베이스에 입수하고 입수된 데이터를 확인, 제한하는 절차로 이 과정에서는 특히 대상 데이터 입수현황에 대한 결과확인 필요하는데 이는 주기적으로 새로운 자료 입수현황을 검색하거나 데이터베이

스 갱신 관리 모듈이 자동으로 입수상태를 모니터링 하는 방식으로 가능하다. 하지만 이런 방법은 데이터베이스에 불필요한 검색을 시행하게 되므로 실제 이용자들에게 데이터를 서비스하기 위한 자원을 효율적으로 사용할 수 없게 한다. 본 시스템에서는 2.2절에서 소개한 바와 같이 자료입수를 위한 SDI 전용 데이터베이스를 별도로 운영하는데 이럴 경우 신규자료에 대한 정기적인 검색과 모니터링을 생략할 수 있게 된다. 정보검색 단계에서는 각 이용자가 설정한 검색형식으로 자료검색을 수행하는데 본 시스템에서는 정보검색을 최소화하기 위하여 이용자가 신청한 주제 중 중복된 주제를 제외한 유일한 주제(Unique item)를 검색하고 그 결과를 데이터베이스에 저장하여 해당 주제를 신청한 이용자들에게 중복적으로 사용할 수 있도록 하였다. 정보제공 단계에서는 이용자 프로파일의 정보에 따라 신청주제에 대한 결과를 검색결과 저장DB에서 검색하여 메일로 구성·발송하는 단계이다. 서비스 과정의 단계별 구분은 [그림1]과 같다.

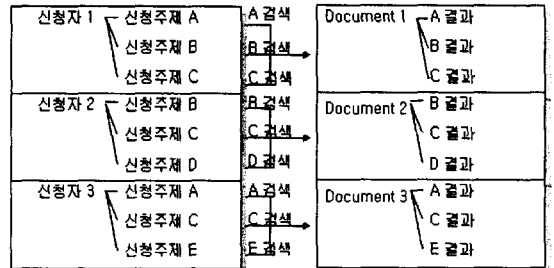


[그림 1] SDI 서비스 과정

이스를 비운 후 다시 새로운 서비스를 위한 데이터를 적재하도록 되어있다. 이때 서비스데이터와 저장데이터의 일관성 유지를 위하여 SDI서비스를 위한 검색 및 결과물 제작이 일어나는 동안에는 SDI 서비스 전용 데이터베이스의 Update를 제한하였다. 그 외 데이터의 일관성 유지를 위해 고려할 사항은 3.2절에서 언급하였다.

2.2.2 검색결과저장 DB 운영

SDI 서비스에서는 하나의 주제를 여러 이용자가 중복하여 신청할 수 있다. 특히 SDI서비스의 이용자가 증가하고 3.1절에 소개한 것과 같은 미리 정의된 특정주제에 대한 SDI 서비스가 시행될 경우 중복이 더욱 많아 질 수 있는데 이런 경우 일반적으로 시행하는 이용자별 신청주제별 정보검색을 시행할 경우 [그림2]와 같이 동일한 검색이 여러 번 반복되게 된다.



[그림2] 일반적인 검색과정

2.2 제안모델

2.2.1 SDI 서비스 전용 데이터베이스 운영

SDI 서비스는 데이터베이스가 갱신될 때 마다 신착정보를 검색하여 제공하게 되므로 서비스 대상 정보는 새로 갱신된 정보로 제한될 수 있다. 이때 실제 운영 중인 서비스 데이터베이스에서 SDI서비스를 위한 검색을 시행할 경우 기 저장되어 있던 데이터의 수량이 방대하여 불필요한 검색시간이 소요되는 문제가 있다. 또한 서비스 대상 데이터베이스가 여러 개인 경우 상이한 구조의 데이터베이스 검색을 위해 검색항목 매핑과 같은 복잡한 절차가 이용자 프로파일이나 주제별 프로파일 작성시 포함되어야 한다. 이에 본 시스템은 검색속도 향상과 서브스 대상 데이터베이스의 확장을 용이하게 하기 위하여 SDI 서비스를 위한 전용 데이터베이스를 구축하였다. 이 SDI 전용 데이터베이스에는 SDI 서비스만을 위한 데이터가 적재되며 데이터 구성항목은 실제 저장위치에 대한 정보와 질의에 필요한 주요필드가 된다. 이렇게 SDI 서비스 전용 데이터베이스를 구축하면 실제 서비스 데이터베이스의 구조에 관계없이 표준화된 항목만으로 구성된 메타데이터형 정보의 저장만으로 SDI서비스가 가능하게 된다. SDI 서비스 전용 데이터베이스는 SDI 서비스 시점에서 대상이 되는 신착정보만 저장되어 있도록 유지하기 위하여 SDI 서비스가 시행되어 결과물이 이용자들에게 완벽히 제공되고 나면 기존 데이터를 백업하고 데이터베

이의 개선을 위해 동일주제에 대한 검색은 1회만 시행하여 그 결과를 중복 활용할 수 있도록 검색결과를 저장하여 관리하는 검색결과DB를 별도로 운영하였다. 검색결과DB를 별도로 운영함으로써 SDI 서비스에 신청된 유일한 주제(Unique item)수와 만큼의 검색만이 필요하게 된다.

이때 정보검색의 효율성을 증가시키기 위하여 이용자가 요구정보를 정의하고 이용자별 프로파일 생성되는 시점에서 필요한 검색을 간략히 관리하는 신청주제별 프로파일도 함께 생성되도록 하였다. 정보검색은 2.2.1절의 SDI 서비스 전용 데이터베이스를 대상으로 한다. 검색결과 저장 데이터베이스는 이용자별 SDI서비스 결과물을 제작하기 위해 사용되며 이 데이터베이스의 데이터 구성항목은 유일한 주제(Unique item)별 구분자와 실제 서비스 데이터베이스의 Article 제어번호가 된다. 검색결과 저장 데이터베이스는 오직 주제(item)에 따른 검색에만 시행되므로 이를 인덱스로 하여 효율을 더욱 향상시킬 수 있다.

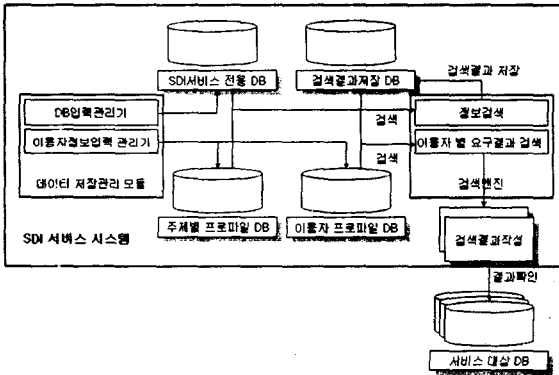
2.2.3 결과물 제작

한 이용자가 여러 아이템의 정보를 신청했다라도 결과물은 하나의 메일로 구성되어야 하므로 이용자 프로파일에 등록된 신청정보를 확인하여 검색결과 저장 데이터베

이스에서 해당 주제(item)의 결과를 모두 검색한 후 그 검색결과를 하나의 개체로 통합하는 과정이 수행된다. 검색결과 저장 DB에 주제별로 인덱스가 구성되어 있으므로 이용자별 신청 아이템에 대한 검색결과를 보다 빨리 획득할 수 있으며 아울러 질의는 신청자의 구분자만으로 검색 제한항목을 작성한 간단한 형태로 실행된다. 검색은 SDI 서비스의 이용자(Unique user)수 만큼 시행된다.

2.3 시스템 구성

본 연구에서 제안하는 SDI 서비스 시스템 구성은 [그림3]과 같이 이루진다. 기능별로 시스템으로의 데이터 저장관리 부분과 검색관리 부분으로 구분할 수 있는데 입력관리 부분은 SDI서비스 데이터 저장, 프로파일 저장, 유일한 주제별 검색결과 저장 등을 수행하고, 검색부분은 SDI 전용 데이터베이스로부터 주제별 검색 시행, 검색결과저장 데이터베이스로부터 이용자별 요구결과 검색 시행 등의 절차를 나누어 수행한다. SDI 시스템은 데이터의 저장과 검색기능의 성능이 전체 시스템의 성능에 크게 영향을 줄 수 있다.



[그림3] 시스템 구성도

3. 시스템 구축에 대한 기타 이슈

3.1 참고목록을 이용한 신청기능

본 시스템은 특정 주제(item)가 여러 이용자들에게 중복적으로 신청되는 사례가 많을수록 보다 효율적인 서비스를 할 수 있다. 이러한 특성을 고려하여 본 시스템에는 무작위적인 아이템의 신청이 아닌 참고 목록을 활용한 아이템 신청기능을 추가한다. 참고목록을 활용할 경우 이용자들이 제한된 범위 내에서 주제물 신청하도록 하여 정보검색의 횟수를 축소할 수 있는데 이러한 참고 목록을 이용한 신청서비스로 일본과학기술진흥원의 표준최신정보제공 서비스(Standard SDI Service)가 대표적이다.

3.2 SDI 전용DB의 데이터 신뢰성 확보

SDI 전용DB의 데이터가 실제 서비스 DB에 등록이 되어 있지 않는 데이터가 생기게 되면 SDI 서비스를 통해 전달받은 데이터를 실제 웹사이트 등을 통해 확인할 방법을 상실하게 된다. 이런 문제의 예방은 다수의 이용자를 대상으로 하는 서비스에서 매우 중요한 문제이다. 이런 문제의 발생을 방지하기 위하여 SDI 서비스 전용 데이터베이스에는 실제 서비스 데이터베이스에 Insert가 완료된 것이 확인된 데이터만 적재할 수 있도록 하였으며 이를 위해 서비스 DB에 Insert가 완료된 자료만을 표준 데이터 파일로 만들어 SDI 전용 DB 적재를 위한 자료로 활용하는 방법을 채택하였다. 즉, 서비스 DB와 SDI 전용 DB의 데이터를 실시간으로 일치시키는 것이 아니라 서비스 DB에 완전히 반영된 자료만 SDI전용 DB에 Insert하여 운영하는 방법을 택하였다.

4. 결 론

SDI 서비스는 이용자가 필요로 하는 정보를 신속하고 정확하게 보급하는데 의의가 있다. 신속한 서비스를 위해서는 정보의 검색과 검색결과물 제작시간의 단축이 필수적으로 요구되는데 본 연구에서는 이를 위해 검색범위와 검색횟수를 최소화 할 수 있는 방안으로 SDI 서비스 전용 데이터베이스 운영과 검색결과 저장 DB 운영을 제안하였다. 또한 본 시스템에서 제안하는 방법을 통해 SDI 서비스 대상 DB를 효과적으로 확장할 수 있으며 본래의 서비스에 부담을 주지 않고 SDI 서비스를 위한 검색이 시행될 수 있는 방안을 살펴보았다. 본 연구를 통해 SDI 서비스의 더욱 확산될 수 있기를 기대한다.

[참고문헌]

- [1] Yan, Tak W. and Hector Garcia-Molina. 1994. "Index structures for information filtering under the vector space model". In Proc. International Conference on Data Engineering, 337-47.
- [2] Yan, Tak W. and Hector Garcia-Molina. 1994. "Index structures for selective dissemination of information under the boolean model". ACM Transactions on Database Systems.
- [3] M. Altinel and M.J. Franklin. Efficient Filtering of XML Documents for Selective Dissemination of Information. In Proceedings of VLDB, pages 53--64. Cairo, Egypt, 2000.
- [4] Wyle, Mitchell F. 1996. "Effective Dissemination of WAN". 1996. [1996.10.18]
- [5] Bing, Michelle. 1997. "Push: a technology with staying power?". Database, 20(4) : 27-30.