

가상문서에서 정형데이터 지원을 위한 스키마처리기와 질의링크 생성기

김철수⁰ 강지훈 강민구*
충남대학교 컴퓨터학과, 제노시스㈜*
(chulsul⁰, jhkang)@cs.cnu.ac.kr ngkang@xenosis.com*

Schema Processor and Query Link Generator for supporting Structured Data in Virtual Documents

Chul-Su Kim⁰, Ji-Hoon Kang, Min-Gu Kang*
Dept. of Computer Science, Chungnam National University
Xenosis Inc.*

요 약

XML을 기반으로 하는 가상문서는 다양한 데이터의 공유를 가능하게 하여 새로운 지식을 생성할 수 있도록 한다. 가상 문서를 지원하는 디지털 도서관 시스템에서 질의 링크는 인터넷 상의 정형데이터 공유를 가능하게 한다. 본 연구에서는 질의링크를 포함하는 XML 기반 가상문서를 효과적으로 생성하기 위해 질의링크 생성기와 스키마 처리기를 설계하고 구현하였으며 이를 지원하기 위해 디지털 도서관 시스템과 가상문서 저작시스템에서 서비스 관리기, 메타 검색기, 데이터베이스 관리기와 저작도구를 확장하였다.

1. 서론

수많은 데이터가 존재하는 인터넷 상에서 자신이 필요한 내용을 가져와 새로운 문서를 만들어야 하는 경우가 종종 있다. 분산 디지털 도서관 시스템에서 이런 지식 생성 기능을 위해 가상문서 개념이 제안되었다[1,2]. 가상문서는 인터넷 상에 존재하는 문서를 재사용하여 분산 환경에서 연관성이 있는 내용을 링크만을 이용해 표현한 새로운 문서형태이다. MIRAGE-III 디지털 도서관 시스템에서 XML기반의 가상문서는 비 정형데이터(텍스트, 이미지, 멀티미디어 데이터)와 준정형데이터(HTML, XML)를 링크로 연결하여 정보를 제공한다[3].

인터넷 상에서 이미 스크립트나 CGI같은 것을 이용해 정형데이터(데이터베이스)를 제공한다. 이에 가상문서에서도 제공하는 정보의 범위를 비정형데이터와 준정형데이터에서 정형데이터까지 확장하기 위한 방법이 제안되었다[4,5]. 가상문서에서 정형데이터를 지원하기 위해 SQL 질의를 포함하는 질의링크를 정의하고 디지털 도서관 시스템에서 가상문서를 재현할 때 질의링크를 처리하기 위해 데이터베이스 관리기를 이용한다.

본 논문에서는 가상문서를 지원하는 디지털 도서관 시스템에서 정형 데이터에 대한 질의 링크를 포함하는 XML 기반 가상문서를 효과적으로 생성하기 위해 질의링크 생성기와 스키마 처리기를 설계하고 구현하였고 이들을 이용한 질의링크의 생성을 지원하기 위해서 서비스 관리기, 메타 검색기, 데이터베이스 관리기와 저작도구를 확장하였다.

본 논문의 구성은 다음과 같다. 2절에서는 가상문서에서 질의 링크의 개념과 이를 처리하기 위한 디지털 도서관의 데이터베이스 관리기에 대해 다루고, 3절에서는 데이터베이스를 관리하기 위한 메타데이터와 질의링크 생성기와 스키마 처리기의 설계 및 구현에 대해 설명한다. 4절에서는 관련연구를 살펴보고 5절에서는 결론 및 향후 연구 방향에 대하여 살펴본다.

2. 정형 데이터를 지원하는 시스템

2.1 질의링크

가상문서는 실제 내용은 포함하지 않고 대상에 대한 링크로만 구성된 문서로서 내포링크, 참조링크, 양방향링크, 다중링크, 중첩링크, 질의링크의 다양한 링크를 지원한다. 질의링크는 URL 같은 주소형태로 인터넷상의 물리적인 문서를 가리키는 링크의 개념을 확장하여 데이터베이스상의 정형데이터를 가리키도록 데이터베이스 질의를 링크의 목적으로 이용하는 링크이다[4,5].

정형 데이터를 위한 질의 링크는 #query라는 한정어를 두어 서버 주소, SQL질의, 데이터베이스 이름을 이용해 다음과 같이 표현한다. 현재 질의링크는 RDB를 위한 SQL 질의만을 지원한다.

```
<ELink href="http://dl.cnu.ac.kr/#query(select * from Table DBName)" ...
```

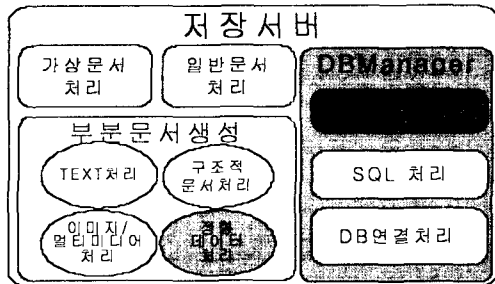
2.2 데이터베이스 관리기

우리의 디지털 도서관 시스템에서 데이터베이스 관리기는 저장서버의 한 부분으로 가상문서의 처리과정 중에서 정형 데이터를 처리하는 부분을 담당한다. 저장 서버는 일반 문서와

이 연구는 BK21 충남대학교 정보통신인력양성사업단의 지원을 받았다.

가상 문서를 저장하며, 문서 처리기에서 부분 문서 요청이 있을 경우 그것에 해당하는 문서를 만들어 URL을 돌려주고 문서 처리기에서 정형 데이터 처리 요청이 있을 때 저장 서버는 데이터베이스 관리자에게 처리 요청을 한다. 데이터베이스 관리기는 질의링크를 지원하기 위해 중요한 역할을 하는 콤포넌트이다. 데이터베이스 관리기는 가상문서상의 질의링크의 브라우징을 위한 작업을 수행한다. 질의링크에 포함된 SQL을 데이터베이스에 보내서 그 결과를 XML 문서로 만드는 SQL 처리기와 데이터베이스 연결에 관련된 작업을 하는 DB연결 처리기가 있다[4].

본 논문에서 데이터베이스 관리기에 스키마 처리기를 추가하여 데이터베이스의 스키마에 관련된 작업을 담당하게 하였다. [그림 1]은 디지털 도서관 시스템에서 정형 데이터를 지원하기 위해 데이터베이스 관리기를 가지는 저장서버의 구조를 보여준다.

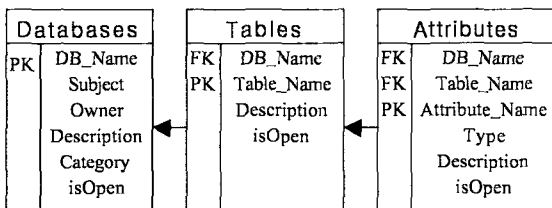


[그림 1] 정형데이터 처리를 위한 저장 서버의 구조

3. 스키마 처리기와 질의링크 생성기의 설계 및 구현

3.1 데이터베이스 관리를 위한 메타데이터

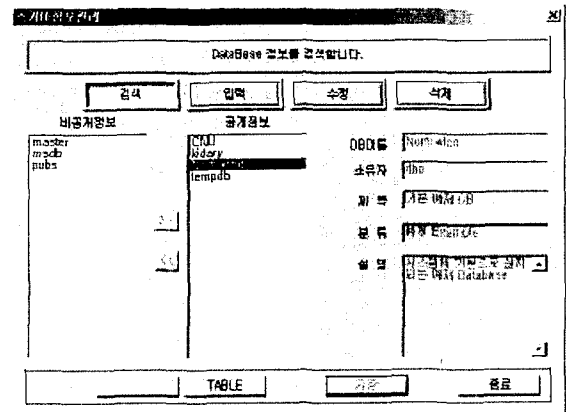
데이터베이스는 웹과 같이 아무나 접근하여 사용할 수 있는 시스템이 아니므로 데이터베이스에 접근하기 위해 허가된 사용자 정보가 필요하고, 질의를 포함하는 질의링크를 생성하기 위해서는 원하는 데이터가 어느 데이터베이스에 포함되어 있는지를 알아야 하고, 그리고 데이터베이스에 대한 스키마 정보가 필요하다. 본 논문에서는 이런 정보를 제공하기 위해서 [그림 2]의 형태로 스키마를 구성하여 정보를 저장하고, 요청을 받았을 때 저장된 정보로부터 적절한 스키마 정보를 제공하여 그 정보로 새로운 질의링크를 생성할 수 있도록 한다.



[그림 2] 메타데이터를 위한 스키마 다이어그램

3개의 테이블로 구성되며 각각은 데이터베이스, 테이블, 앵트리뷰트에 대한 정보를 저장한다. Databases 테이블의 주키를 Tables 테이블에서 외래키로 사용하고 Tables 테이블의 주키와 외래키를 Attributes 테이블에서 외래키로 사용하여 특정 테이블이 어느 데이터베이스에 속해있는지 특정 앵트리뷰트가 어느 테이블과 어느 데이터베이스에 속해있는지 알 수

있다. 그리고 Description, Category, Type에 스키마에 대한 메타데이터를 저장하고 isOpen을 통해 특정 데이터베이스, 테이블, 앵트리뷰트가 공개되었는지 알 수 있다. 관리자는 스키마 정보를 구성하기 위해 관리자 톨을 이용하며, 해당 데이터베이스의 공개 여부 및 메타 데이터의 관리를 해주어야 한다. [그림 3]은 스키마 정보를 검색하는 모습을 보여준다.



[그림 3] 스키마 정보를 관리하는 톨

3.2 스키마 처리기

스키마 처리기는 디지털 도서관 시스템의 데이터베이스 관리기의 일부분으로 데이터베이스 관리를 위한 메타데이터를 이용하여 데이터베이스 관리기의 요청에 의해 크게 세가지 기능을 수행한다. 먼저 질의링크 생성과정에서 원하는 데이터베이스를 찾기 위해 조건을 가지고 데이터베이스 검색 요청을 받으면 해당 데이터베이스를 찾아서 서버 주소를 포함한 데이터베이스 리스트를 만들어 결과를 돌려준다. 다음은 질의링크 생성과정에서 특정 데이터베이스의 스키마 정보를 요청 받을 때 해당 데이터베이스의 공개된 스키마 정보와 메타데이터를 XML 형태로 만들어서 준다. 마지막으로 질의링크 생성과정에서 생성된 질의의 수행 결과를 확인 할 때, SQL 질의를 분석하여 질의가 공개된 데이터베이스, 테이블, 앵트리뷰트에 대해서 수행되는가를 검증한다. 이 작업은 가상문서를 브라우징하기 위해 재연할 때 질의링크를 처리하는 과정에서도 수행된다.

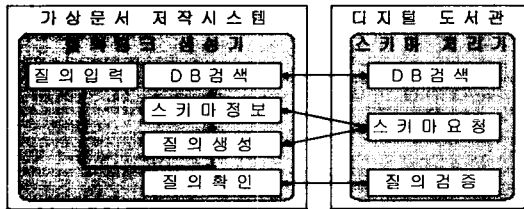
3.3 질의링크 생성기

질의링크 생성기는 가상문서 저작시스템에서 가상문서를 만들 때 질의 링크를 생성하기 위해서 사용자 인터페이스를 통해 사용자와 상호작용을 한다. 질의링크 생성기는 크게 네 단계를 거쳐서 가상문서에 삽입할 질의링크의 생성 과정을 완료한다. 우선 사용자가 어떤 데이터베이스를 원하는지 검색한다. 검색의 결과로 서버의 주소와 함께 데이터베이스 이름의 목록을 받는다. 두번째 단계에서는 앞에서 받은 데이터베이스 목록 중에서 사용자가 데이터베이스를 선택했을 때 해당 데이터베이스의 스키마와 메타데이터를 사용자에게 트리형태로 보여준다. 다음 단계에서는 데이터베이스를 선택하여 받은 데이터베이스의 스키마 정보를 토대로 사용자가 편리하게 SQL 질의를 들 수 있도록 도와준다. 마지막 단계에서는 앞 단계에서 만든 SQL 질의를 디지털도서관에 보내서 그 질의를 처리하고 그 결과를 사용자에게 직접 보여주어 사용자가 의도한 질의인가 확인하고 질의링크를 가상문서에 삽입하기 위한 준비를 한다. 질의링크를 생성할 때 데이터베이스와 스키마 정보를 알

고있으면 데이터베이스 검색, 스키마정보 요청, 질의생성 과정을 거치지 않고 사용자가 직접 질의를 입력하여 질의 결과를 확인할 수 있다.

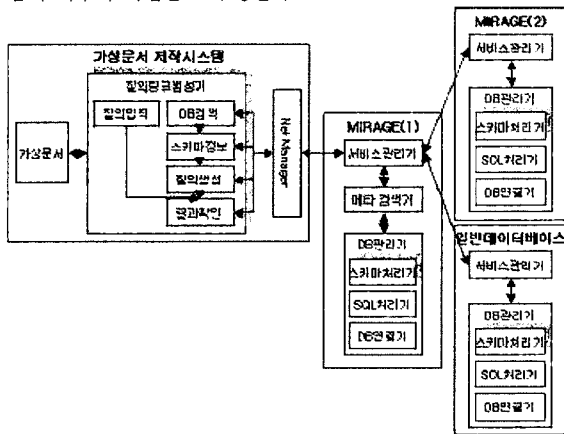
일반 데이터베이스 시스템에 데이터베이스 관리기와 디지털 도서관에서 오는 메시지를 인식할 수 있는 서비스 관리기로 구성된 조그만 시스템을 붙여서 디지털도서관에 정형데이터를 제공할 수 있다

가상문서 저작시스템의 질의링크 생성기와 디지털 도서관의 스키마 처리기는 [그림 4]에서와 같이 상호작용을 한다. 질의링크 생성기에의 DB검색 요청은 스키마 처리기의 DB검색 부분에서 처리되고 스키마 정보와 질의 생성에서의 작업 처리 요청들은 스키마 처리기의 스키마요청에서 처리가 되며 결과확인 단계의 요청은 질의검증 부분에서 처리가 된다.



[그림 4] 질의링크 생성기와 스키마 처리기의 상호작용

[그림 5]는 가상문서 저작도구(Authoring Tool)내의 질의링크 생성기를 이용해 질의 링크를 생성시 데이터베이스를 검색, 질의 생성, 결과를 확인 하는 과정을 보여준다. 우선 사용자 인터페이스인 AT는 도서관 시스템에 접속하여 해당하는 검색어를 주면 서비스관리기가 메타 검색기에게 검색 요청을 한다. 메타 검색기는 로컬, 다른 디지털 도서관 그리고 데이터베이스 관리기를 장착한 일반 데이터베이스를 검색하고 검색 결과를 통합한 데이터베이스 목록을 돌려준다. 질의링크 생성기는 접속한 도서관 시스템을 통해 다른 시스템들과 통신을 하면서 검색 이후의 작업들도 수행한다.



[그림 5] 데이터베이스 검색과정

4. 관련 연구

데이터베이스를 웹상에 공개하는 시스템인 Queensland 대학에서 만든 WEBFINDIT6은 데이터베이스를 주체별로 그룹화하고 메타 데이터를 데이터베이스에 저장하여 사용자가

순차적으로 원하는 데이터베이스를 찾아가는 방법을 제공하고 있다. 우리 시스템에서는 메타 검색기가 데이터베이스를 찾고, 그 데이터베이스의 스키마 정보를 이용해 데이터베이스에 대한 질의와, 질의를 포함하는 문서를 쉽게 만들 수 있는 방법을 제공한다.

Lie와 Pu는 DIOM(Distributed Interoperable Object Model)을 기반으로 분산환경에서 공유 가능한 데이터베이스 시스템인 Diorama를 개발하였다[7]. 이 시스템은 인터넷상의 데이터베이스 시스템들을 하나의 데이터베이스 시스템처럼 이용할 수 있게 하고 사용자들에게 동일한 인터페이스를 제공한다. 우리 시스템에서 Diorama 시스템을 MIRAGE-III가 이용할 수 있는 하나의 DBMS로 간주하여 Diorama 시스템의 데이터베이스를 이용하는 질의링크를 만들고 처리할 수 있다.

5. 결론

권한이 있는 사용자 정보와 스키마 정보를 알아야 질의링크를 생성할 수 있기 때문에 질의링크를 생성하기 위해 디지털 도서관 시스템과 가상문서 저작 시스템에 스키마 처리기와 질의링크 생성기를 추가하고 서비스 관리기, 메타 검색기, 데이터베이스 관리기를 확장하여 효율적으로 질의링크를 생성할 수 있게 되었다.

향후에는 질의링크 생성을 위한 데이터베이스 검색이 단순히 키워드를 포함하는 메타데이터를 찾는 형태이다. 향후 데이터베이스 검색요청 시 보다 나은 결과를 얻을 수 있는 새로운 검색 모델이나 방법론을 적용할 예정이다.

6. 참고 문헌

- [1] S. H. Myaeng, M.-H. Lee, J.-H. Kang, E.-I. Cho, Y.-B. Lee, D.-S. Lim, J.-M. Lim, H.-J. Oh, and J.-S. Yang, "Virtual Documents : a New Architecture for Knowledge Management in Digital Libraries," Proc. of Asian Digital Libraries Conf., Taipei, Taiwan, Nov. 1999.
- [2] 강지훈, 맹성현, 이만호, " 분산 디지털 도서관 시스템에서 XML을 이용한 가상문서의 표현처리," 제 2 회 디지털 도서관 컨퍼런스 논문집, pp.49-60, 1999 년 11 월.
- [3] S. H. Myaeng, M.-H. Lee, J.-H. Kang, E.-I. Cho, Y.-B. Lee, D.-S. Lim, J.-M. Lim, H.-J. Oh, and J.-S. Yang, "A Digital Library System for Easy Creation/Manipulation of New Documents over Existing Resources", Conf. of Proc of RIAO 2000, Paris, France, pp.196-207, April 2000.
- [4] 강민구, 김철수, 강지훈, "가상문서에서 질의 링크 처리를 위한 설계와 구현," 한국정보과학회 2001 년 학술발표논문집, 제 28 권 2 호, pp. 169-171, 2001 년 10 월
- [5] J.H.Kang, M.G.Kang, C.S.Kim, "Supporting Databases in Virtual Documents", Conference Papers of The 4th International Conference of Asian Digital Libraries 2001(ICADL'2001), Bangalore, INDIA, pp. 125-134, December 2001.
- [6] Bouguettaya, A., Benatallah, B., Ouzzani, M., & Hendra, L., "WEBFINDIT: An Architecture and System for Querying Web Databases," IEEE Internet Computing, Vol. 3, No. 4, pp.30-41, July-August. 1999
- [7] Liu, L. & Pu, C(1995), "The Distributed Interoperable Object Model and Application to Large-scale Interoperable Database Systems," Proc. Of Int'l Conf. On Asian Digital Libraires(ICADL'99), Taipei, Taiwan, November