

가상문서에서 질의 링크 처리를 위한 설계 및 구현

강민구⁰ 김철수 강지훈
충남대학교 컴퓨터학과
(mgkang, chulsul, jhkang)@cs.cnu.ac.kr

A Design and Implementation for processing Query Links in Virtual Documents

Min-Gu Kang⁰, Chul-Su Kim & Ji-Hoon Kang
Dept. of Computer Science, Chungnam National University

요 약

XML을 기반으로 하는 가상문서는 인터넷 상의 정보 공유를 가능하게 하여 새로운 지식을 생성할 수 있도록 한다. 가상문서에서는 비정형(텍스트, 이미지, 멀티미디어 데이터)과 준정형(HTML, XML) 데이터를 링크로 연결하여 정보를 제공하고 있다. HTML과 같은 기존의 웹 문서에서는 스크립트나 CGI 같은 것을 통하여 정형 데이터(데이터베이스)의 정보를 제공하고 있으므로 가상문서에서도 자연스럽게 정형 데이터를 연결하여 사용이 가능하다. 본 논문에서는 디지털 도서관 시스템에서 정형 데이터를 지원하기 위해 질의 링크를 포함하는 가상문서를 효율적으로 처리할 수 있도록 데이터베이스를 설계 및 구현하며, 데이터베이스 스키마 정보를 관리하여 필요한 데이터베이스를 검색하여 질의 링크의 생성을 돕도록 하였다.

1. 서론

인터넷의 급격한 발달로 웹상에는 수없이 많은 데이터가 존재한다. 수 많은 데이터 중에는 텍스트 형태의 데이터뿐만 아니라 이미지, 오디오 비디오와 같은 다양한 종류의 미디어 데이터가 존재한다. 이러한 분산 환경에서 이미 존재하는 수 많은 미디어 데이터 중에서 자신이 필요한 일부 또는 전체를 가져와 새로운 문서로 만들어야 하는 경우가 종종 있다. 이러한 지식 생성 기능을 분산 디지털 도서관 시스템에서 효율적으로 충족 시키기 위한 새로운 개념의 문서인 가상문서를 정의하였다. 이러한 가상문서를 디지털도서관 시스템이 지원하기 위해서는 가상문서 양식이 필요하며, 표현력, 처리 효율성, 인터넷 접근성 등을 고려하여 XML DTD로 정의되었다 [MLK99, 강지훈99].

가상문서에서는 비정형(텍스트, 이미지, 멀티미디어 데이터)과 준정형(HTML, XML) 데이터를 링크로 연결하여 정보를 제공하고 있다. HTML과 같은 기존의 웹 문서에서는 스크립트나 CGI 같은 것을 통하여 정형 데이터(데이터베이스)의 정보를 제공하고 있으므로 가상문서에서도 자연스럽게 정형 데이터를 연결하여 사용할 수 있도록 한다. 그러나 데이터베이스는 웹과 같이 아무나 접근하여 사용할 수 있는 시스템이 아니므로 허가된 사용자가 필요하고 데이터베이스 스키마를 알고 있어야 사용 가능하다. 그러므로 가상 문서에서 링크를 사용하여 정형 데이터를 연결하고 처리할 수 있는 새로운 방법이 필요하다. 즉, 우리는 링크의 목적지를 데이터베이스 질의로 보고

가상문서 링크의 의미를 확장하였으며 이것을 질의 링크라고 한다. 또한 디지털 도서관 시스템은 가상문서 내의 질의 링크를 처리하기 위해 데이터베이스 관리기를 두는 방향을 제안하였다[KKK2001].

본 논문에서는 디지털 도서관 시스템에서 정형 데이터 형태의 콘텐츠를 제공하기 위해 질의 링크가 포함된 가상문서를 효과적으로 처리할 수 있도록 데이터베이스 관리기를 설계하고 구현 한다. 특별히 데이터베이스 스키마 정보를 관리하여 가상문서 저작자로 하여금 필요한 데이터베이스를 검색할 수 있도록 함으로써 질의 링크 생성을 도울 수 있도록 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 가상문서에 대한 개념과 가상문서 기반의 디지털 도서관 시스템 구조에 대해 다루고, 3장에서는 질의 링크를 지원하고 처리하기 위한 문서 처리기와 저장 서버 및 데이터베이스 관리기에 대해 설명한다 4장에서는 관련연구를 살펴보고 5장에서는 결론 및 향후 연구 방향에 대하여 살펴본다.

2. 가상문서

2.1 가상문서의 개념

가상문서는 인터넷 상에 존재하는 문서 전체 또는 그 일부를 재사용하여, 분산 환경에서 연관성이 있는 내용을 링크만을 이용해 표현한 새로운 문서이다. 즉, 새로 만들어진 가상문서에는 실제 내용은 존재하진 않고 기존에 존재하던 콘텐츠의

* 이 연구는 BK21대전,충남 정보통신인력양성사업단의 지원을 받았음

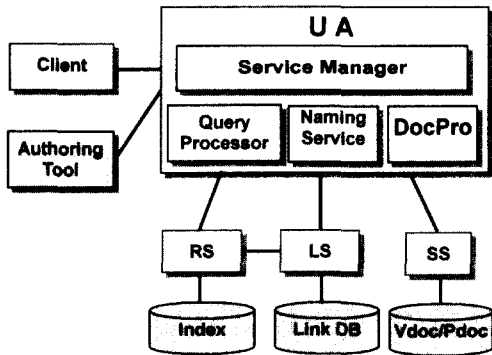
링크만을 갖게 된다. 가상문서는 허브(hub)와 스타일시트(styleshet)로 구성된다. 허브는 문서의 구조와 내용을 나타내고, 스타일시트는 가상문서 전체 및 구성요소를 위한 스타일 정보를 가진다[장지훈99, MLK+ 2000].

가상문서는 허브와 스타일시트로 구성된다. 허브는, 내포링크 순서(a sequence of embedding links), 참조링크 집합(a set of referential links), 그리고 메타데이터(metadata)의 세 개로 구성된다. 즉, VdocHub는 가상문서에서 가장 상위 요소(Element)로 ELinkSeq, RLinkSet, Metadata의 순서를 갖는 세 개의의 요소를 가지고 있다.

내포링크는 가상문서가 실제화 될 때 그 구성 요소가 되는 문서들에 대한 링크를 말한다. 가상문서 상태에서는 단지 링크이지만, 클라이언트가 가상문서를 브라우징하게 되면 그 가상문서가 가지고 있는 모든 내포링크의 목적지에 존재하는 문서들이 불러와서 하나로 취합이 되어 온전한 문서 형태로 클라이언트에게 보여야 한다. 참조링크는 링크의 다양한 방향을 지원한다. 기존의 HTML 문서에서 사용하던 단방향 링크는 물론, 양방향링크(bi-directional link), 목적지가 여러 곳인 다중링크(multi-directional)를 지원한다.

2.2 가상문서를 지원하는 디지털 도서관 시스템

우리의 디지털 도서관 시스템[Myaeng99]에서 문서처리 모듈(Document Processing Module)인 DocPro(문서 처리기)는 UA(User Agent)의 한 부분으로 가상문서의 처리를 담당한다. 문서처리기는 가상문서의 고유 ID(Identifier) 생성과, 링크서비스에 필요한 링크정보를 반환하며, 문서의 검색을 위한 가상문서 ID와 문서에 정의된 메타데이터, 내포 문서들과 그 문서에 대한 ID들을 반환해준다. 그리고 추가기능인 문서변환의 기능을 가지고 있다. [그림 1]은 디지털 도서관시스템에서 가상문서 처리를 위한 시스템의 구조를 나타내고 있다.



[그림 1] 가상문서 처리를 위한 시스템의 구조

3. 정형 데이터 지원을 위한 설계

3.1 질의 링크

가상문서에서 링크의 목적지는 인터넷 상의 물리적인 문서를 지시하며, URL형태의 주소로 표시된다. 데이터베이스가 인터넷 상에 존재함으로써 가상문서에서 링크의 목적지로 사용하려 한다. 즉, 링크의 목적지로 데이터베이스 질의를 사용하며, 그것을 질의 링크라 한다[KKK2001].

3.2 가상문서에서 질의 링크 표현

미디어 데이터의 종류에 따라서 각각의 표기법이 필요하다. 텍스트 데이터의 표기 방법은 #text라는 한정어를 두어 Start, Offset으로 부분문서를 표현한다. Start는 부분문서가 시작되는 byte이고, Offset은 포함될 데이터의 범위를 지정한다.

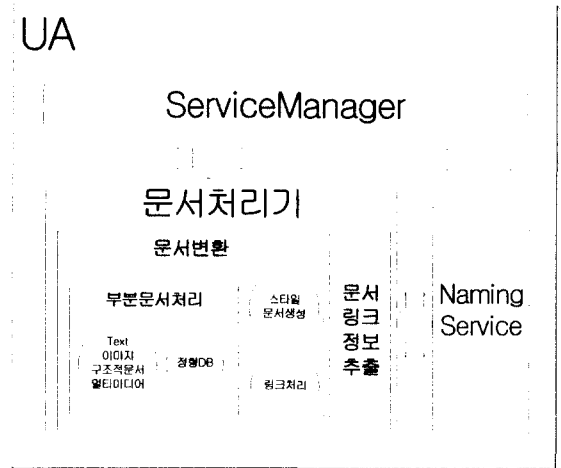
```
<ELink href="dl.cnu.ac.kr/pdoc/bach.txt#text(Start, Offset)" ...
```

정형 데이터를 위한 질의 링크는 #query라는 한정어를 두어 SQL과 데이터베이스 이름을 이용해 부분문서를 표현한다. 현재 SQL은 RDB만을 위한 표준 쿼리를 지원한다.

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

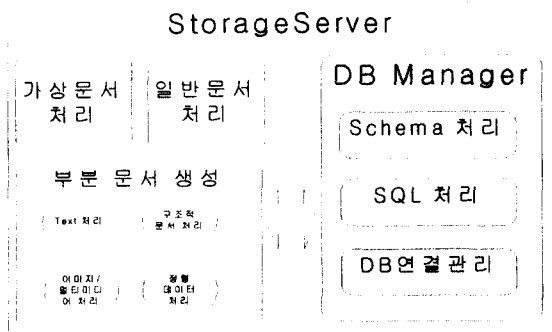
3.3 질의 링크 처리를 위한 설계

디지털 도서관 시스템에서 정형 데이터 처리를 위해 기존의 문서처리기와 저장 서버를 확장하였다. 문서처리기의 중요한 역할은 링크 요소로만 되어있는 가상문서를 사용자가 가상문서를 작성한 의도대로 볼 수 있도록 한다. 문서처리기에서는 질의 링크를 처리하는 부분을 인식하고 저장서버에 요청할 수 있는 부분이 필요하다. 또한 정형데이터를 위한 스타일 추가하였으며, 링크 처리 부분은 수정하였다. [그림 2]는 정형 데이터를 지원하는 질의 링크를 처리하기 위한 문서처리기 구조를 나타낸다.



[그림 2] 질의 링크를 처리를 위한 문서처리기의 구조

저장 서버는 일반 문서와 가상 문서를 저장하며, 부분 문서 요청이 있을 경우 즉 텍스트의 일부분이나 그림의 일부분 또는 동영상의 일부분 요청이 있을 경우 그것에 해당하는 문서를 만들어 문서 처리기에 URL을 되돌려준다. 문서처리기에서 정형 데이터 처리 요청이 있을 때 저장 서버는 정형 데이터 처리를 위해서 SQL을 처리하고 DB 연결을 위한 모듈이 필요하다. 검색이나 가상 문서 작성 할 때에 편의를 제공하기 위해 메타 정보(데이터베이스 스키마)를 저장 관리하는 부분이 필요하다. 그래서 정형 데이터를 잘 관리할 수 있도록 데이터베이스 관리를 추가하였다. [그림 3]는 저장서버에서 정형 데이터를 위한 처리 구조를 나타낸다.

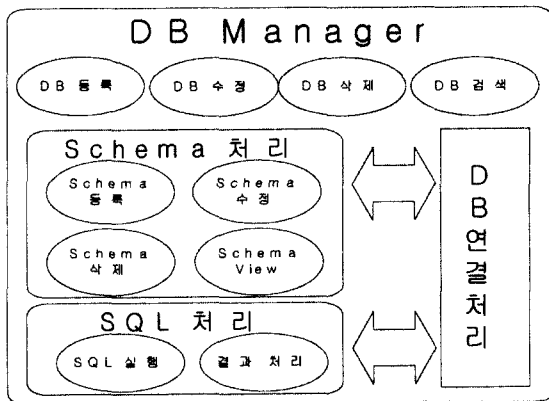


[그림 3] 질의 링크 처리를 위한 스토리지 서버 구조

3.4 데이터베이스 관리기

데이터베이스 관리기는 질의 링크를 지원하기 위해 중요한 역할을 하는 컴포넌트이다. 데이터베이스 관리기는 가상문서상의 질의 링크를 브라우징을 위해 처리하고 검색을 위한 메타 정보를 관리하는 역할을 한다.

데이터베이스 관리기는 데이터베이스를 등록하고 관리하는 역할을 하는 부분, 이 데이터베이스에 존재하는 테이블 스키마 정보와 간략한 메타 정보를 등록하고 관리하는 스키마 처리기, 질의 링크에 있는 SQL을 데이터베이스에 보내서 그 결과를 받아 XML 문서로 만드는 SQL 처리 부분과 데이터베이스 연결에 관련된 작업을 하는 DB 연결 처리 부분이 있다. [그림 4]는 데이터베이스 관리기의 구조를 나타낸다.



[그림 4] 데이터베이스 관리기의 구조

가상문서를 브라우징 할 경우 질의 링크의 처리과정을 살펴 보면, 문서 처리기에서 질의 링크를 발견하고 그것을 스토리지 서버에 요청(getQueryResult())하면, 스토리지 서버는 가상문서에서 표현된 질의 링크에서 서버와 SQL을 분리하여 데이터베이스 매니저에게 요청(excuteQuery())하고 데이터베이스 관리기는 데이터베이스에 연결한 후 질의를 하여 그 결과를 받아서 XML 처리(generateXML0)를 한 후 스토리지 서버에 보내고 스토리지 서버는 XML 문서의 URL를 문서 처리기에 보내면 브라우징을 위한 스타일 처리를 해서 클라이언트에 보내지게 된다.

4. 관련 연구

다가문서 모델[PW96]과 FEDORA[PC98]에서의 디지털 객

체 개념은 디지털 도서관을 위한 문서 구조를 정의한다는 점에서 우리의 시스템과 공통점이 있다. 전자의 경우, 문서는 다양한 형태의 층으로 이루어지며, 각 층은 "behavior"라 부르는 프로그램의 지원을 받는다. 후자의 경우에는 디지털 객체는 사용자 측면에서는 내부 구조가 감추어져 있으며 "disseminator"라 부르는 서비스 인터페이스의 집합을 통하여 사용자가 디지털 객체에 개념적으로 쉽게 접근할 수 있도록 하고 있다. 기존의 두 접근 방식은 자료를 틀 안에 넣고 구조화하는데 주안점을 둔 데 비하여, 우리의 방식은 다양한 링크를 지원함으로써 자료를 연결하고 확장하는 데에 초점을 맞추고 있다.

데이터베이스를 웹상에 공개하는 WEBFINDIT[BBOH99]은 데이터베이스를 주제별로 그룹화하여 웹상에서 쉽게 접근할 수 있도록 하고 있으며, 메타 데이터를 데이터베이스에 저장하여 데이터베이스의 위치를 알 수 있는 방법을 제공하고 있다. 그러나 인터넷 상에서 있는 데이터베이스 정보를 새로운 문서를 만들 수 있는 기능을 제공하지는 않는다.

5. 결론

이미 존재하는 문서들을 링크를 이용하여 새로운 형태의 문서를 만들 수 있는 가상문서는 인터넷 상에서 자유롭게 지식을 공유하고, 이용할 수 있도록 해준다. 본 논문에서는 기존에 디지털 도서관의 문서처리기에서 지원한 텍스트, 이미지 데이터, 멀티미디어와 구조적 문서를 포함하여 정형 데이터를 지원 하도록 처리 방법을 제시하였다. 현재에는 관계형 데이터베이스만을 지원하나, 앞으로 다양한 모델의 데이터베이스를 지원할 필요가 있다.

6. 참고 문헌

[강지훈 99] 강지훈, 맹성현, 이만호, " 분산 디지털 도서관 시스템에서 XML을 이용한 가상문서의 표현처리," 제 2회 디지털 도서관 컨퍼런스, 서울, 1999년 11월.

[MLK99] S. H. Myaeng, M.-H. Lee, & J.-H. Kang, "Virtual Documents: a New Architecture for Knowledge Management in Digital Libraries," Proc. Int'l Conf. on Digital Libraries, Taipei, Taiwan, Nov. 1999.

[BBOH99] A. Bouguettaya, B. Benatallah, M. Ouzzani, & L. Hendra, "WEBFINDIT: An Architecture and System for Querying Web Databases," IEEE Internet Computing, Vol. 3, No. 4, pp.30-41, July-August. 1999.

[PC98] S. Payette, and C. Lagoze, "Flexible and Extensible Digital Object and Repository Architecture," Second European Conf. on Research and Advanced Technology for Digital Libraries, Heraklion, Crete, Greece, Sept, 1998. (<http://www.cs.cornell.edu/payette/papers/ecdl98/fedora.html>)

[PW96] T. Phelps and R. Wilensky, "Toward active, extensible, networked documents: multivalent architecture and applications", Proc. 1st ACM Int'l Conf. on Digital Libraries, Bethesda, Maryland, U.S.A., Mar 1996.

[MLK+2000] S. H. Myaeng, M.-H. Lee, J.-H. Kang, E.-I. Cho, Y.-B. Lee, D.-S. Lim, J.-M. Lim, H.-J. Oh, & J.-S. Yang, "A Digital Library System for Easy Creation/Manipulation of New Documents over Existing Resources", RIAO 2000, Paris, France, Apr 2000.

[KKK2001] J.-H. Kang, M.-G. Kang & C.-S. Kim, "Supporting Databases in Virtual Documents", submitted to ICADL 2001.