

객체지향 데이터베이스를 이용한 구조화된 문헌의 저장 및 관리에 관한 연구

A Study on Storing and Managing Structured Documents with an Object-Oriented Database

이재진, 문성빈
연세대학교 문헌정보학과

Jae-Jin Lee, Sung-Been Moon
Dept. of Library and Information Science, Yonsei University

문헌의 구조화를 통한 전문의 이용은 검색에 있어 많은 효율을 가져올 것으로 기대된다. 여기에서 구조화된 문헌을 저장하고 관리하기 위한 데이터베이스는 그 특성을 반영하여 구축되어야 한다. 본 연구에서는 이를 위한 데이터베이스로서 객체지향 데이터베이스를 선정하고 객체지향 모델링 기법을 이용하여 구조화된 문헌을 위한 데이터베이스를 설계하고 구현하였다.

1. 서론

전문(full text)을 데이터베이스에 저장·관리하여 이를 이용하기 위한 최근의 연구 경향은 전문으로 제공될 문헌을 구조화하는 것이다. 문헌의 구조화는 문헌에서 나타나는 실제 내용과는 별도로 문헌의 구조 정보, 즉 문헌의 논리적 정보로서 장·절에 관한 정보를 이용할 수 있도록 하기 위해 이루어진다.

이와 같은 문헌에 포함된 구조 정보를 이용하기 위해서는 이를 저장하고 관리하여 나아가 검색할 수 있는 데이터베이스 시스템이 필요하

다.

이러한 배경에서 본 연구는 먼저 구조화된 문헌의 특성을 살피고, 이를 효율적으로 저장하고 관리할 수 있는 데이터베이스로서 객체지향 데이터베이스를 선정하였다.

이를 통해 실제 구조화된 문헌을 분석하여 이를 적절하게 객체지향 데이터베이스로 저장하고 관리할 수 있도록 객체지향 모델링 기법을 이용하여 데이터베이스를 설계하였다.

설계를 위해 제시한 방법을 토대로 실제 객체지향 데이터베이스·시스템 가운데 하나인 POET 3.0을 이용하여 데이터베이스를 구현하

여 평가하였다.

2. 구조화된 문헌을 저장·관리하기 위한 데이터베이스

2.1 구조화된 문헌의 특징

문헌을 구조의 측면에서 분석할 때 크게 논리적 구조와 물리적 구조로 나눌 수 있는데 흔히 '구조화된 문헌'이라 칭하는 경우 논리적 구조로서 실제 문헌의 내용과 관련되어 문헌을 일련의 장, 절, 문단 등으로 기술한 것을 의미한다. 문헌의 논리적 구조를 기술하기 위한 메타 언어로는 SGML이 국제 표준으로 제정되어 있다.

구조화된 문헌은 계층적인 구조를 갖고, 문헌을 구성하는 각각의 요소가 검색시 접근점이 될 수 있고, 구성요소간의 의미적 관계 파악이 용이하고 다양한 데이터 유형을 다룰 수 있다는 특징을 갖는다.

이러한 구조화된 문헌을 다루기 위한 데이터베이스는 그 특징을 최대한 살려줄 수 있는 것이어야 한다.

2.2 구조화된 문헌을 위한 데이터 모델

데이터베이스 시스템을 위한 데이터 모델 가운데 하나인 관계형 모델은 현재 상용 데이터베이스 시스템에서 널리 사용되고 있다. 그러나 관계형 모델은 문헌의 계층적 구조를 표현하기 어렵고, 다양한 길이를 갖는 가변장 필드를 다루기 힘들다. 또한 데이터의 의미 표현이 자유롭지 못하고 지나친 정규화로 인해 질의 표현이 복잡해 질 수 있다는 단점을 가지고 있다.

반면, 최근 많은 연구가 이루어지고 있는 객체지향 데이터베이스를 위한 객체지향 모델의 경우, 객체와 클래스의 계층을 계층의 개념으로 지원하여 문헌의 계층 구조를 자연스럽게

나타낼 수 있고, 객체나 클래스가 가질 수 있는 데이터형으로 사용자가 새롭게 정의한 사용자 정의형도 사용할 수 있다. 또한 객체나 클래스간의 다양한 의미 관계도 표현이 가능하다.

따라서 구조화된 문헌을 저장하고 관리하기 위한 데이터베이스로는 객체지향 데이터베이스가 적절하다.

3. 구조화된 문헌을 위한 객체지향 데이터베이스의 설계

3.1 구조화된 문헌의 분석

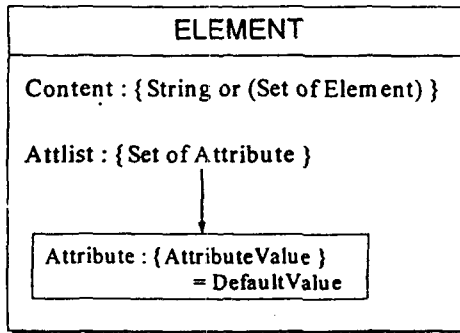
데이터베이스의 설계를 위해 우선 구조화된 문헌을 분석하여 문헌의 구조 정보의 개념을 얻었다. 구조화된 문헌으로는 SGML로 표현된 문헌으로서 ISO 12083 가운데 논문 기사 DTD를 선정하였다. DTD는 SGML의 가장 핵심적인 부분으로서 문헌에 나타나는 논리적 구조를 설명하는 문법이다. 따라서 이러한 DTD에 나타나는 내용을 통해 문헌에 나타나는 논리적 구조를 파악할 수 있을 뿐만 아니라 객체지향 데이터베이스의 설계를 위한 세부 내용을 결정할 수 있다.

3.2 구조화된 문헌의 객체지향 모델링

구조화된 문헌의 분석을 통해 객체지향 데이터베이스의 가장 기본이 되는 객체 및 클래스로서 문헌요소(element)를 추출하였다.

문헌요소는 DTD나 실제 문헌부에서 가장 중요한 요소로서, 실제 데이터베이스 검색시 기본 접근 단위가 된다. 이렇게 클래스로 이끌어낸 문헌요소의 메타클래스를 위해 문헌요소 선언 내용과 속성 정의 리스트를 그 성질로 모델링하고, 이렇게 모델링한 메타클래스를 각 개별 클래스에 모델링한다.

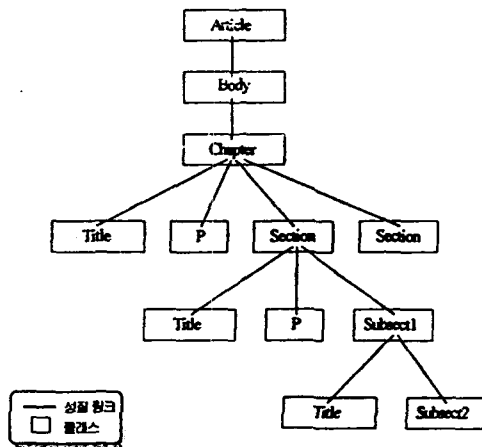
<그림 1>은 문헌요소 메타클래스의 모델링을 객체지향 모델링 방법 가운데 하나인 OMT 방법을 응용하여 나타낸 것이다.



<그림 1> 문헌요소 메타클래스의 모델링

개별 클래스의 모델링 과정에 의해 각 클래스는 클래스간의 의미 관계를 표현할 수 있게 되는데 특히 클래스의 성질이 연결된, 클래스 성질 계층을 형성할 수 있다.

<그림 2>는 클래스 성질 계층의 일부 예를 보인 것이다.



<그림 2> 클래스 성질 계층

모델링한 클래스는 실제 객체지향 데이터베이스로 사상되고 저장하게 되는데, 이 때 데이터 정의어가 사용되고 본 연구에서는 ODL (Object Definition Language)을 응용하였다.

모델링한 클래스의 하나의 예로서 클래스 Article을 ODL을 응용하여 나타내 보면 다음과 같다.

```

Class Article
{
  struct Content
  ( Front front;
    Body body;
    Appmat appmat;
    Back back;
  );

  struct Attlist
  ( // String HyTime = HyDoc;
    String id;
    // String SDAFORM = "article";
  );
}
  
```

4. 데이터베이스의 구현

본 연구에서는 이러한 과정의 실제 시스템에서의 적용 가능성을 살펴보기 위해서 데이터베이스로 구현하였다.

데이터베이스 구현을 위한 객체지향 데이터베이스 시스템으로는 POET 3.0을 채택하였고, 객체지향 프로그래밍 언어로는 POET 3.0과 함께 사용되는 C++(Microsoft Visual C++ 1.5)을 이용하였다. 생성된 데이터베이스를 이용한 검색은 POET 3.0에서 사용하는 OQL(Object Query Language)을 이용하였다.

구체적으로 문헌요소를 클래스로 생성하고 생성된 클래스를 통해 전체 데이터베이스 구조를 형성한다. 클래스가 생성되면 자연스럽게 개별 객체도 생성된다. 이 과정이 완료되면 데이터베이스의 틀이 짜여진 것이고 여기에 실제 문헌 인스턴스를 데이터로 저장하면 데이터베이스의 구축이 이루어지는 것이다. 구축된 데이

터베이스를 이용하여 객체에 의한 데이터 검색이 가능하게 된다.

5. 결론

본 연구에서는 구조화된 문헌을 객체지향 데이터베이스로 저장·관리하기 위한 방법을 제시하고 이를 데이터베이스로 구현하였다.

본 연구를 통해 제시한 과정의 특징은 구조화된 문헌의 분석을 통해 문헌요소를 접근 단위로 추출하여 이를 클래스로 구성하였다는 것과 내용 모델과 함께 속성 정의도 그 일부로 포함시켰다는 점, 그리고 이를 통해, 문헌의 내용 검색과 함께 논리 정보도 검색이 가능하다는 점이다.

그러나 모델링 과정에서 나타난 제한점은 데이터 소장 후에도 나타나는 엔티티 요소의 처리가 불가능하다는 것과 속성 정의시에 나타나는 다양한 데이터 유형을 구별하지 않았다는 점, 그리고 데이터 입력시에 하나의 문헌요소의 내용 모델이 다른 문헌요소와 혼합된 경우, 이들의 순서나 위치가 고려되지 않았다는 점 등을 들 수 있다.

이러한 설계 과정을 통해 검색시 사용될 적합한 질의어를 개발하고 데이터의 유지 및 보수를 위한 기능 개발은 본 연구의 앞으로의 연구 방향이 될 것이다.

참고문헌

- 김 원, 객체지향 데이터베이스, 서울 : 하이테크 정보, 1995.
- 이규철, "전자도서관에서의 구조적 정보 관리," 과학기술정보 워크샵 논문집 Nov. 1996, pp.53~80.
- 한예노, 강현석, "SGML 문서의 관리를 위한

객체지향 데이터베이스 설계," 한국정보과학회 가을 학술발표논문집 vol.23, no.2, 1996, pp.143~146.

Aberer, K., et al., "The Prospects of Publishing Using Advanced Database Concepts," *Conference on Electronic Publishing* vol.6, no.4, Dec. 1993. Available via <ftp://ftp.darmstadt.gmd.de/pub/dimsys/reports/P-93-23.ps.Z>

Christophides, V., et al., "From Structured Documents to Novel Query Facilities," *Proceedings of ACM SIGMOD 1994*, pp.313~324. Available via <http://www-db.stanford.edu/~abitebou/pub/sigmod94.ps>

Macleod, I. A., "Storage and Retrieval of Structured Documents," *Information Processing & Management* vol.26, no.2, 1990, pp.197~208.

Sacks-Davis, R., et al., "Database Systems for Structured Documents," *IEICE Transactions on Information & Systems* vol.E78-D, no.11, Nov. 1995, pp.1335~1342.

Tompa, F. W., "What is (tagged) text?" *Proceedings of the 5th Annual Conference of the UW Center for the New OED* 1989, pp.81~93. Available via <http://daisy.uwaterloo.ca/~fwtompa/papers/oed89.ps>

Zhang, J., "Application of OODB and SGML Techniques in Text Database : An Electronic Dictionary System," *SIGMOD RECORD* vol.24, no.1, Mar. 1995, pp.3~8. Available via http://bunny.cs.uiuc.edu/sigmod/sigmod_record/9503/Zhang.ps