

구조정보를 이용한 SGML문서 편집기

서 연구, 조 성배
연세대학교 컴퓨터과학과

An SGML Document Editor Using Structural Information

Yeon-Gyu Seo and Sung-Bae Cho
Dept. of Computer Science, Yonsei University

요약

컴퓨터 환경이 인터넷을 위주로 한 웹 환경에서 사용되기 시작하면서 전자상거래를 위한 문서교환에 대한 요구가 급증하고 있는 실정이다. SGML이 ISO 8879로 표준화되면서 SGML을 이용한 문서교환이 전자상거래의 표준으로 자리잡고 있다. 문서의 적용범위가 다양해짐에 따라서 SGML문서의 논리구조를 정의하는 문서형 정의(Document Type Definition, DTD)가 다양해지고 있다. SGML문서의 편집을 위해서는 SGML에 대한 많은 지식이 요구되며 다양한 DTD가 존재하므로 일반 사용자가 편집하기는 쉽지 않다. 따라서 SGML에 대한 일반적인 지식이 없어도 편집이 가능한 SGML전용 문서편집기가 필요한데 이 논문에서는 DTD를 파싱하여 얻어낸 구조정보를 저장하고 이 정보를 이용하여 사용자가 SGML에 대한 일반적인 지식 없이도 편집이 가능한 SGML편집기를 구현하였다.

1. 서론

웹을 이용한 문서교환이 가능해짐에 따라 전자상거래가 활성화되고 있다. 이기간의 문서교환을 위해서는 표준화된 문서형식이 필요한데 ISO 8879라는 표준안이 나오면서 SGML[2]을 이용한 문서교환이 활성화되고 있다. SGML 문서는 선언부, 문서형 정의부(DTD), 그리고 문서의 인스턴스(Document Instance, DI)로 구성되는데 DTD는 문서의 구조 및 작성규칙에 대해 정의하고 있으며 DI는 DTD에서 파생 가능한 하나의 인스턴스가 된다[1][2][3].

SGML문서의 편집을 위해 기존의 편집기를 사용할 수 있지만 다양한 문서형식의 요구에 따라 DTD가 다양해지고 그에 따른 규칙도 다양해지므로 사용자가 SGML문서를 편집하기 위해서는 복잡하고 다양한 DTD규칙을 알아야 하는 어려움이 따르게 된다. 따라서 DTD규칙을 모르더라도 일반 사용자가 쉽게 편집이 가능한 SGML문서 편집기가 필요하다고 볼 수 있다.

본 논문에서는 DTD로부터 구조정보를 얻어내고 이를 사용자에게 디스플레이함으로써 사용자는 문서의 전체구조를 보면서 편집할 수 있도록 하며 편집 시에 가능한 입력에 대한 정보를 제공하며 문서의 구조정보를 이용하여 간단히 편집할 수 있도록 하는 구조지향적 SGML문서 편집기를 구현하였다.

구체적으로는 SGML문서를 파싱하여 생성된 SGML의 구조정보인 Grove[6]로부터 DTD에 대한 정보를 얻어내고 이 정보를 이용하여 문서의 편집이 가능하도록 하였다. Grove로부터 얻어낸 DTD정보를 이용하여 DI를 생성해내기 위해서 고려할 수 있는 여러 방법 중에서 여기서는 DTD로부터 DI로의 매핑 방법을 사용하였다.

2절에서는 일반적인 SGML문서의 처리과정에 대해 언급하고 3절에서는 SGML문서 편집기의 전체적인 내용에 대해 설명한다. 4절에서는 DTD로부터 구조정보의 처리에 대해 설명하고 5절에서는 SGML문서의 편집에 대해 논의한다. 그리고 6절에서는 SGML문서 편집기의 인터페이스와 구현에 대해 설명하고 마지막으로 결론을 맺도록 한다.

2. SGML문서의 처리과정

일반문서는 구조정보를 포함하고 있지 않지만 SGML문서는 구조정보를 포함하게 되는데 SGML문서의 처리과정은 이러한 구조정보를 이용하는 과정이라고 할 수 있다. SGML문서는 일반문서와는 달리 프리젠테이션 정보가 포함되어 있지 않음으로 문서의 프리젠테이션을 위해서는 특별한 처리과정이 필요한데 DSSSL[4]이 이러한 문서의 프리젠테이션 정보에 대한 처리를 한다.

문서의 형 정의를 위해 DTD를 만들고 이것을 이용해 필

요한 여러 가지 문서를 만든다.

SGML문서는 여기서 만들어진 DTD에서 정의하고 있는 규칙에 따라 작성되어야 하므로 DTD로부터 문서에 대한 정보를 얻어내기 위한 파서가 필요하다. 파서는 문서에 대한 정보를 생성해내어 문서의 전체정보를 포함하는 Grove에 저장한다. 여기서 생성된 Grove를 이용하여 원하는 문서를 만들게 되는데 이 때 편집된 문서는 문서의 실제 정보와 함께 DI의 구조정보도 포함한다. 작성된 문서는 출력을 위한 포맷터가 필요하게 되는데 여기서 포맷터에 의해 출력형식이 지정되면 문서의 출력결과를 볼 수 있다 [3][4][8]. 그림1은 SGML문서의 처리과정에 대해 보여준다.

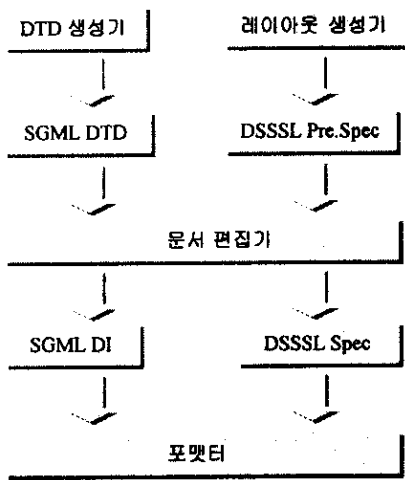


그림 1. SGML문서의 처리과정

3. SGML문서 편집기

SGML문서 편집기는 문서를 편집할 때 DTD를 파싱하여 얻을 수 있는 구조정보를 이용한다. SGML문서의 편집기는 일반적인 워드프로세서의 기능 뿐 아니라 현재의 입력이 DTD의 문서 작성규칙과 맞는지를 체크할 수 있어야 하므로 워드프로세서와 문법체크 기능을 수행하는 편집기라는 특징을 갖는다.

SGML문서의 편집을 위해서는 Grove에서 DTD정보를 얻어내고 DTD정보로부터 문서의 전체구조를 트리형태로 구성한다. 그리고 구조정보를 이용하여 문서의 작성순서를 제시함으로써 문서의 편집이 가능하도록 한다.

SGML문서의 편집과정은 Grove로부터 DTD정보를 얻어내고 얻어진 DTD정보에서 문서의 구조정보를 얻어 트리를 구성한 다음 구조정보를 이용하여 편집을 한다. 편집할 때 DI의 구조정보를 유지하기 위한 여러 방법이 있지만 본 논문에서는 DTD에서 DI로의 구조정보의 매핑방법을 사용하였다. 그리고 작성된 DI는 구조정보와 함께 문서형

태로 저장한다.

4. 구조정보의 처리

이 절에서는 문서의 구조정보를 이용하기 위한 노드의 구성과 문서의 구조정보 추출에 대해서 언급하고자 한다.

DTD에서 엘리먼트는 그 안에 내용모델(Content Model)을 가질 수 있는데 문서를 나타내는 엘리먼트는 그 안에 하위 엘리먼트들을 갖고 그 엘리먼트들은 또 다른 엘리먼트로 구성됨으로서 문서의 계층구조에 대한 정보를 유지하고 있다.

문서의 계층적 구조정보를 추출하기 위해 사용되는 방법에도 고려할 수 있는 여러 방법이 있지만 여기서는 Grove가 문서의 전체적인 정보를 유지하고 문서의 구조정보는 Grove를 액세스하여 얻어내는 방법을 사용한다. 그림2는 Grove로부터 DTD에 대한 정보와 SGML문서의 계층적인 정보를 추출하는 과정을 보이고 있다.

이 그림에서 문서이름은 문서를 나타내는 시작포인트가 되고 그 안에 DI에서 사용될 수 있는 여러 엘리먼트들이 리스트 형태로 유지된다. 일반적으로 SGML문서는 처음의 엘리먼트를 액세스하면 문서의 전체적인 구조정보를 얻을 수 있는데 시작 엘리먼트는 하위 엘리먼트를 내용모델로 가지고 있기 때문이다. 하위 엘리먼트의 정보를 엘리먼트 리스트에서 검색하고 다시 이와 같은 과정을 반복하면 문서의 전체적인 구조정보를 얻을 수 있다.

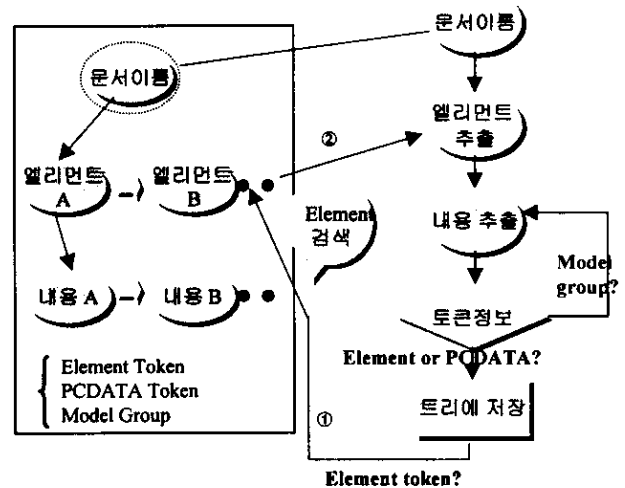


그림 2. 구조정보의 추출

그림2로부터 얻어낸 정보를 2진트리 형태로 구성하는데 하나의 엘리먼트에서 내용모델로 있는 엘리먼트들은 자식노드로 하며 같은 내용모델을 구성하고 있는 엘리먼트들은 형제노드를 구성하게 된다. 이러한 2진트리를 구성하기 위

해 사용되는 하나의 노드는 내용모델에 대한 정보를 액세스할 수 있는 모델그룹과 가질 수 있는 속성에 대한 포인터를 가지고 있다. 그림3은 노드의 구조와 노드를 구성하는 트리에 대해 보여주고 있다.

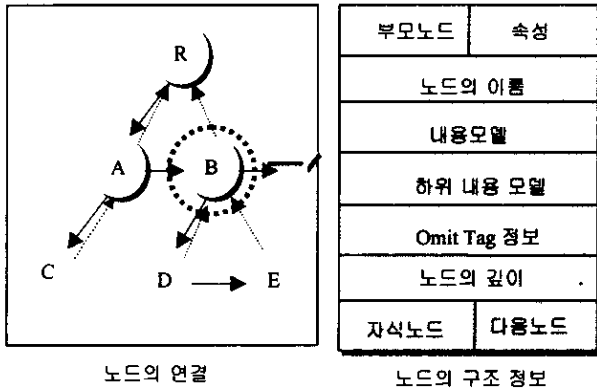


그림 3. 노드의 구조와 계층정보의 구성

그림에서 내용모델은 그림2와 같이 파싱이 가능한 'PCDATA'와 또 하나의 내용모델을 포함하고 있는 모델 그룹, 그리고 엘리먼트 등의 세 가지 토큰의 형태로 구성되어 있다. 만약 내용 모델 안에 모델그룹이 존재할 경우 문서처리의 복잡함을 제거하고자 내용모델에서 정보를 유지하지 않고 하위내용모델을 별도로 두어 액세스할 수 있도록 하였다. 여기서 노드의 깊이는 트리의 계층정보를 유지하는데 사용된다.

5. SGML문서의 편집

5.1 DTD에서 DI로의 구조정보 연결

문서를 편집할 때 사용자는 구조정보를 보고 편집하고자 하는 부분을 선택하면 된다. 사용자가 특정부분을 선택하면 선택된 부분에 대한 정보를 저장된 2진 트리로부터 검색하고 관련된 정보를 얻어온다. 이 때 얻어진 정보로 DI 트리에 삽입가능한지의 여부를 체크하게 된다.

삽입가능 여부는 현재 DI 트리의 구성과 문서의 작성순서를 비교 조사함으로써 가능하다. DI 트리에서 가장 마지막 편집위치를 검색하고 그 위치 다음에 올 수 있는 정보를 내용모델에서 유지되는 문서의 작성 순서정보로 검색한 다음 현재 삽입될 부분이 이 순서에 적합한지를 검사하면 된다. 만약 이때 입력 가능한 선택이라면 DI 트리에서의 삽입위치를 검색하고 해당 위치에 삽입함으로써 DI 트리의 구조정보를 유지하게 된다.

이러한 방법은 문서의 작성순서를 계속 검색해야 하는 불편함이 있지만 DI의 구조정보를 쉽게 구성할 수 있다는 장점이 있다. DI가 작성되고 난 후 수정할 때는 DTD의 구

조정보를 액세스할 필요가 없으므로 DI의 구조정보만을 이용하여 수정이 가능하다.

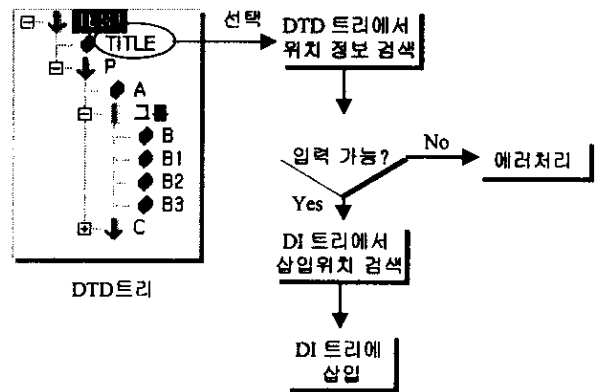


그림 4. DTD에서 DI로의 구조정보 연결

5.2 편집시의 에러처리

현재 프로그램에서 사용자가 편집하고자 하는 부분을 선택하고 입력하는 경우 입력된 내용이 선택된 태그 내에 삽입될 수 있는지를 Grove로부터 얻은 DTD정보를 이용하여 편집가능 여부로 체크한다.

이 때 발생할 수 있는 에러에는 크게 삽입순서의 에러와 잘못된 문자의 입력에러 등이 있을 수 있다. 삽입순서의 에러는 5.1절에서 설명한 것처럼 엘리먼트의 내용모델에 대한 정보를 검색하여 현재 DI의 구조에서 삽입가능한지를 비교 체크하면 가능하다. 그리고 잘못된 문자의 입력은 삽입 가능한 태그 안에서 입력할 수 없는 문자를 입력한 경우나 기입되어야 할 속성이 기입되지 않은 등 여러 경우가 있을 수 있는데 이에 대해서는 선택된 부분의 구조정보로부터 에러를 체크할 수 있다.

6. 편집기의 인터페이스 및 구현

SGML문서 편집기의 인터페이스는 3개의 창으로 이루어져 있다. DI와 DTD의 구조정보를 보여주는 창과 문서를 편집할 수 있는 편집 창, 그리고 태그를 선택했을 때 태그가 엘리먼트인 경우 삽입될 수 있는 속성정보를 보여주는 속성 정보 창으로 구성하였다.

문서의 전체구조를 트리 형태로 표시하여 보여주는데 문서를 편집할 때는 편집하고자 하는 트리 구조를 선택함으로써 가능하다. 창에 보이는 트리에는 문서의 작성흐름을 알기 쉽도록 이미지를 삽입하였다. 처음에 보이는 이미지는 작성 순서와 관련된 이미지를 보여주고선택이 되면 삽입 가능한 회수를 보여준다. 편집하고자 하는 부분을

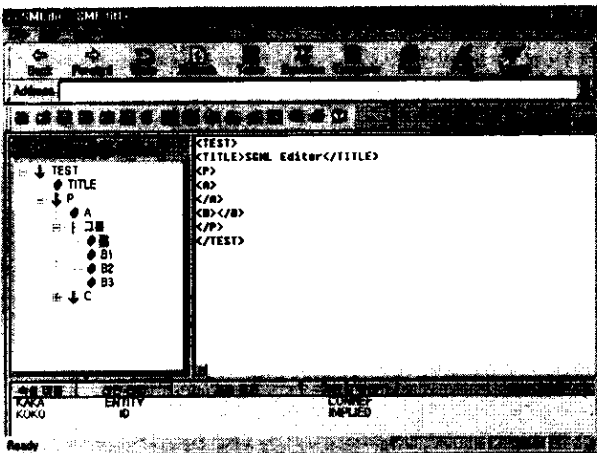


그림 5. 문서의 구조정보를 이용한 편집 예

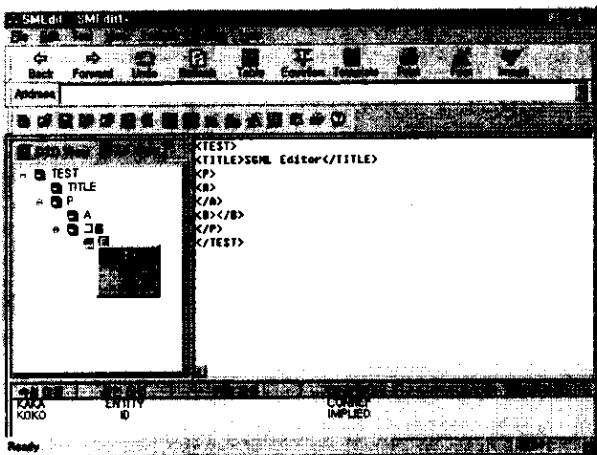


그림 6. DI 구조정보 및 DI 구조정보를 이용한 편집

선택하면 아래 속성 정보 창에서 삽입 가능한 속성을 제시하게 된다. 편집부분이 선택과 동시에 DI 트리로 구조정보가 연결되며 동시에 삽입가능여부를 판단하여 삽입 가능하다면 DI 트리에서 삽입위치를 검색하여 삽입한다.

사용자는 문서의 전체구조를 보면서 편집하게 되는데 현재 작성된 문서의 구조를 알고 싶다면 DI View를 선택함으로써 DI의 현재구조를 볼 수 있다. DI의 문서의 편집은 DI의 구조정보를 이용하여 편집할 수 있다.

그림5는 구조정보를 이용한 문서 편집의 예를 보여주고 있는데 사용자가 구조정보에서 삽입할 부분을 선택하면 관련된 태그가 자동 삽입되고 사용자는 실제 내용을 기입하면 된다. 이 때 현재 편집되고 있는 DI의 내용을 보려면 DI View를 선택하면 된다. 그림6은 편집중인 DI의 구조정보를 알아보기 위해 DI View를 선택한 화면을 보여주고 있는데 DI의 편집은 그림에서처럼 컨텍스트 메뉴를 이용한 편집이 가능하다. 현재 이 프로그램은 웹 기반이 아니라

단독으로 수행되고 있다. 인터페이스를 보면 웹을 위한 틀바 창이 보이는데 이 틀바는 SGML문서 편집기를 통해 작성된 문서를 웹 상에서 사용할 수 있도록 하기 위한 것이지만 현재는 구현되어 있지 않고 앞으로 계속 구현할 계획이다.

7. 결론

표준 ISO 8879와 함께 SGML이 전자상거래의 표준으로 자리잡으면서 다양한 문서가 요구됨에 따라 일반사용자도 쉽게 편집할 수 있는 SGML전용 문서편집기가 필요하다. 그러나 실제 사용 가능한 SGML문서 편집기는 많지 않은 것이 현재 실정이다.

본 논문에서 문서의 구조정보를 DI 구조정보로 연결하는 방법을 사용하여 SGML 문서 편집기를 구현해 보았다. 아직 부족한 점이 많으며 앞으로 보강해 나갈 계획이다. 특히, 현재 국내에서는 웹 기반 SGML편집기가 구현된 적이 없는데 앞으로는 그러한 SGML문서 편집기에 대한 연구도 수행되어야 할 것이다.

참고문헌

- [1] C. F. GlodFarb, *Parseme.1st SGML for Software Developers*, Prentice-Hall, 1998.
- [2] E. Herwijnen, *Practical SGML*, 2nd Ed, Kluwer Academic Pub., 1994.
- [3] ISO 8879, Information processing -- Text and office systems -- *Standard Generalized Markup Language(SGML)*, ISO, 1986.
- [4] ISO/IEC 10179, Information technology - Text and Office system -- *Document Style Semantics and Specification Language(DSSSL)*, ISO, 1996.
- [5] M. Bryan, *SGML*, Addison-Wesley Pub., 1988.
- [6] R. A. Milowski, *Grove Guide 1.1*
- [7] 박준서, 신경희, 문현주, 유재우, "구조 지향적 SGML 문서 편집기", 정보과학회, pp 56-58, 1998.
- [8] 윤승구, 이경호, 최윤철, "DSSSL기반 SGML 문서편집기의 설계 및 구현," 정보과학회, pp 467-469, 1998.