

DSSSL 에 기반한 SGML 문서의 변환에 관한 연구

백승욱^{*}, 장은영, 변창원, 고승규, 이경호, 최운철
연세대학교 컴퓨터과학과

A Study on the Transformation of SGML Documents based on DSSSL

SeungUk Baek, EunYeung Chang, ChangWon Byun, SeungKyu Ko, KyongHo Lee, YoonChul Choy
Department of Computer Science, Yonsei University

요 약

SGML은 이 기종간에 호환이 가능하고 논리적인 구조 정보를 포함한다는 장점 때문에 문서처리의 다양한 분야에서 표준 포맷으로 자리를 잡았다. 그러나 SGML은 문서의 스타일이나 변환 등의 처리 정보를 기술할 수 있는 방법을 제공하지 않는다. 이에 ISO는 SGML 문서의 다양한 처리 정보를 기술할 수 있는 표준으로 DSSSL을 제안하였다. 따라서 본 논문에서는 SGML 문서의 내용 및 구조를 수정할 수 있는 DSSSL 기반의 문서 변환기를 개발하였다. 이를 위하여 SGML 문서와 DSSSL 문서간의 표준적인 인터페이스를 지원하는 Grove 생성기를 개발하였으며 DSSSL의 변환 언어를 지원하는 처리 모듈을 개발하였다.

1. 서론

SGML(Standard Generalized Markup Language)[1] 문서는 이 기종간에 호환이 가능하고 논리적인 구조 정보를 가지고 있다는 장점 때문에 CALS(Commerce At Light Speed) 등의 문서 표준으로 자리잡았다. 이에 문서 처리의 다양한 분야에서 SGML 문서를 처리 및 가공하고자 하는 요구가 급증하고 있다.

특히 SGML 문서의 다양한 처리 중에서 문서간의 효율적인 변환은 매우 중요하다. 즉, 문서의 용도와 사용자의 요구가 변환에 따라 문서의 내용을 재 가공할 필요가 있다. 또한 기존의 문서를 상이한 구조의 문서로 변환할 필요가 있다[2].

이에 ISO(International Organization for Standardization)는 SGML 문서의 스타일 처리는 물론이고 변환 정보를 기술할 수 있는 표준으로 DSSSL(Document Style Semantics and Specification Language)[3]을 제안하였다. 따라서 본 논문에서는 DSSSL에 기반한 SGML 문서 변환기를 개발하였다. 특히 SGML 문서의 파싱을 위해서 기존에 본 연구팀이 개발한 파서[4]를 이용하였다.

^{*}본 연구는 정보통신부 대학기초연구지원사업의 연구비 지원에 의한 것임.

일반적으로 SGML 문서의 변환 과정은 처리 모델에 따라 이벤트 구동(event-driven)과 트리 탐색(tree-traversal) 방식의 두 가지로 구분된다[5]. 이벤트 구동 방식에서는 문서의 파싱 결과로 발생하는 이벤트를 순차적으로 처리하며, 트리 탐색 방식은 문서의 파싱 정보로부터 트리 구조를 구성한 후 이를 이용하여 변환 과정을 수행한다. 본 논문에서 개발한 문서변환기는 트리 탐색 방식의 처리 모델에 기반하고 있다.

본 논문의 구성은 다음과 같다. 2장에서는 DSSSL의 변환 언어에 대한 간략한 소개를 기술하고, 3장에서는 문서 변환기의 설계 및 구현에 대하여 기술하고 4장에서는 결론을 기술한다.

2. DSSSL의 변환 언어

문서의 처리 과정은 일반적으로 포맷팅과 변환의 두 가지로 구분된다. 이에 DSSSL은 SGML 문서의 다양한 처리 정보를 기술하기 위하여 변환 언어(transformation language), 스타일 언어(style language), 그리고 질의 언어(standard document query language)를 제공한다. 특히 변환 언어와 스타일 언어는 수식(expression) 언어에 기반하기 때문에 서로 유사하다.

이중에서 변환 언어는 SGML 문서의 내용 및 구조에 대

한 변환 정보를 기술할 수 있다. 예를 들어 변환 언어를 이용하면 임의의 DTD(Document Type Definition)[6]에 따라 작성된 SGML 문서를 다른 DTD의 문서로 변환할 수 있다.

DSSSL은 최근에 수정된 HyTime(Hypermedia/Time-based Structuring Language)[7]의 문서 모델인 그래프 형태의 Grove(Graph Representation Of property Value)에 기반하고 있다. 이에 SGML 문서에 DSSSL의 변환 정보를 적용하기 위해서는 먼저 문서의 파싱 정보를 이용하여 Grove를 생성한다.

변환 언어로 기술된 문서는 정의부(definition)과 결합 규칙(association)의 집합으로 구성된다. 실제로 변환 과정은 소스 Grove에 결합 규칙을 적용하여 결과 Grove를 생성하는 과정이다.

각각의 결합 규칙은 세 부분으로 구성된다. 첫번째 부분은 질의 수식(query expression)으로 소스 Grove에서 변환의 대상이 되는 노드 리스트(node list)를 선택한다. 두 번째 부분은 변환 수식(transform expression)으로 질의 수식의 결과값인 특정한 노드에 대하여 Grove 형태의 변환 결과를 생성하여 결과 Grove 상에 위치시킨다. 마지막 부분은 우선순위 수식(priority expression)으로 정수 값을 반환하며 결합 규칙의 우선 순위를 나타낸다. 특히 우선순위 수식은 생략이 가능하다. DSSSL의 변환 언어를 이용한 SGML 문서의 변환 과정은 그림 1과 같다.

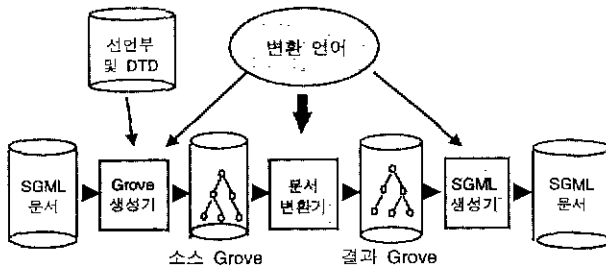


그림 1 DSSSL의 문서 변환 과정

3. 설계 및 구현

본 논문에서 개발한 SGML 문서 변환기의 주요 구성요소 에 대한 자세한 설명은 다음과 같다.

3.1 Grove 생성 모듈

진술한 바와 같이 SGML 문서에 DSSSL을 적용하기 위해서는 SGML 문서로부터 Grove를 생성하여야 한다. 응용 시스템이 SGML 문서를 처리한다는 것은 결국 Grove를 처

리한다는 것이다. 따라서 본 논문에서는 SGML 문서를 파싱 한 후, 그 결과로 제공되는 정보를 이용하여 Grove를 생성 할 수 있는 Grove 생성기를 개발하였다.

Grove는 Grove 계획에 의하여 생성된 노드의 집합이다. Grove 상의 모든 노드는 Grove 계획에 속하는 특정 클래스의 인스턴스(instance)이고, 해당 클래스에 대하여 지정 가능한 특성(property) 및 그 값으로 이루어진 특성 할당(property assignment)의 집합을 갖는다. 이에 Grove 계획을 통하여 응용 프로그램의 특성에 맞추도록 클래스와 특성의 집합을 모듈 단위로 정의할 수 있다.

특히 본 연구에서 개발한 Grove 생성기가 지원하는 Grove 계획은 intrbase(intrinsic base), baseabs(base abstract), prlgabs0(prolog abstract level 0), instabs(instance abstract), prlgabs1(prolog abstract level 1) 등의 모듈에 속하는 모든 클래스 및 특성을 포함한다

SGML 파서의 출력 결과를 이용하여 Grove를 생성하는 과정은 다음과 같다. Grove는 크게 프롤로그와 실제 문서부에 대한 두 부분으로 이루어진다. 따라서 Grove 생성기는 먼저 파서가 문서의 프롤로그 부분을 처리한 후 제공하는 정보를 이용하여 Grove의 프롤로그 부분을 생성한다. 마찬가지로 파서가 실제 문서부를 파싱한 후 제공하는 정보를 이용하여 실제 문서부에 해당하는 부분을 생성하여 최종적으로 Grove를 완성한다.

3.2 수식 언어 해석 모듈

수식 언어 처리 모듈은 LALR(1) 파싱 기법을 적용하며 어휘 분석기의 출력인 토큰을 사용하여 적합한 상위 생성 규칙(production rule)을 찾아 나간다. 이 과정에서 빌트인(built-in) 프로시저를 사용하여 기본 사칙 연산을 비롯하여 리스트 및 문자열 등의 연산을 수행한다. 본 논문에서 구현한 수식 언어 처리 모듈은 DSSSL에 정의된 수식 언어의 모든 기능을 지원한다.

3.3 변환 언어 해석 모듈

변환 언어 처리 모듈은 변환 언어로 기술된 결합 규칙에 따라 소스 Grove로부터 결과 Grove를 생성한다. 변환 언어는 특성상 순차적 파싱 기법의 적용이 불가능하다. 따라서 그림 2와 같이 먼저 변환 정보를 1회 스캐닝하여 정의부와 결합 규칙에 대한 리스트를 생성한다. 변환 언어에 포함된 수식은 파싱테이블을 통한 상향식 접근 방식을 취하고 변환

언어 자체는 하향식 접근 방식을 적용한다.

변환 언어 처리 모듈은 결과 Grove를 생성하기 위하여 소스 Grove를 재귀적으로 탐색하며 Grove 상의 각 노드에 해당하는 결합 규칙을 결합 규칙 리스트로부터 찾는다.

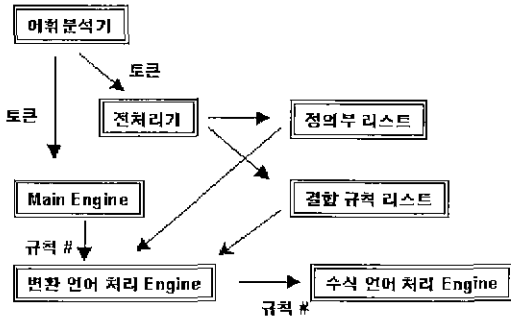


그림 2 변환 언어 해석 모듈

3.4 SGML 문서 생성기

SGML 문서 생성기는 결과 Grove로부터 변환의 최종 결과인 SGML 문서를 재구성한다. 특히 결과 Grove의 유효성 여부를 검증하는 방법은 다음과 같다. 즉, 생성된 문서로부터 세로이 Grove를 구성한 후, 결과 Grove의 각각의 노드로부터 검증 매핑(verification mapping)[3]이 존재하는지를 검사한다. 이때 검증 매핑의 모든 요건을 만족하면 결과 Grove는 유효하다.

4. 결론

DSSSL은 SGML 문서의 스타일 정보뿐만 아니라 변환 등의 다양한 처리정보를 기술할 수 있다. 따라서 본 논문에서는 SGML 문서의 변환 정보를 DSSSL의 변환 언어를 이용하여 기술하고 이를 처리할 수 있는 SGML 문서 변환기를 개발하였다. 이를 위하여 SGML과 DSSSL의 표준적인 처리를 지원하는 Grove 생성기를 개발하였으며 DSSSL의 변환 언어를 지원하는 DSSSL 처리기를 개발하였다. 개발된 SGML 문서변환기는 문서의 내용은 물론이고 구조의 변환을 지원한다.

참고문헌

[1] ISO 8879, *Information processing -- Text and office systems -- Standard Generalized Markup Language (SGML)*, ISO, 1986
 [2] Murata Makoto, "DTD Transformation by Patterns and Contextual Conditions," In Proceedings of SGML/XML '97, pp.

153-169, 1997

[3] ISO/IEC 10179, *Information technology - Text and office system - Document Style Semantics and Specification Language(DSSSL)*, ISO/IEC, 1996
 [4] 고영곤, 최윤철, "한글 SGML 문서처리를 위한 파서의 개발에 관한 연구", 한국통신학회 춘계학술논문발표회, pp. 535-543, 1997
 [5] Tony Graham, "Free SGML Transformation Tools," In Proceedings of SGML'96, pp. 127-130, Boston, MA, November 18-21, 1996
 [6] Charles F Goldfarb and Yuri Rubinsky, *The SGML Handbook*, Clarendon Press, Oxford, 1990
 [7] ISO/IEC 10744, *Information Processing - Hypermedia/Time-based Structuring Language (HyTime) 2nd edition*, ISO/IEC, 1997

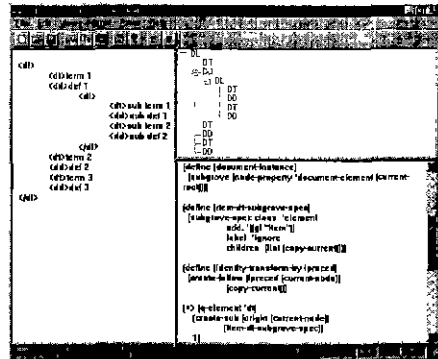


그림 3 SGML 문서와 DSSSL 변환정보



그림 4 변환된 SGML 문서