

TheoCMS®

XML/SGML Enabled CMS (Content Management System)

Technical Publication Team/K4M
February, 1999
Copyright © 1999, K4M

정보 구조화를 위해 1986 년 SGML (Standard Generalized Markup Language: ISO8879)이 국제 표준으로 제정되고 산업계의 수용도를 높이기 위해 1998 년 W3C 의 추천안으로 XML(eXtensible Markup Language)이 소개됨에 따라 XML/ SGML 에 기반한 구조화 문서정보 인프라 구축의 저변이 각 산업계로 급속히 확대되어 가고 있다. 이제까지의 정보의 단순 생성 측면보다는 정보의 효율적인 활용이라는 측면에서 정보 재사용과 유통이 촉진되도록 구조화된 정보의 유연한 활용이 강조되고 있다.

TheoCMS(Content Management System) 제품은 정보 구조화의 특성을 반영한 새로운 정보 인프라로서 주요 데이터베이스 시스템과 연동되는 Repository 기능을 갖는 XML/SGML Enabled CMS 이다. 이는 기존의 인트라넷/엑스트라넷/그룹웨어의 발전된 형태의 지식관리를 위한 정보 인프라 구축의 필수적인 모듈로서 인터넷 환경에서의 다양한 콘텐츠의 효율적 관리와 정보 유통을 촉진한다. 본 고에서는 XML/SGML 의 응용 영역에서의 지식관리를 위한 요소로서의 TheoCMS 서버의 기능을 기존의 문서정보 인프라와 비교하여 소개한다.

Contents	
PART I	XML/SGML 과 문서정보 인프라
PART II	XML/SGML Enabled CMS 와 CALS/EC
PART III	지식지향 정보인프라를 위한 CMS 솔루션 - TheoCMS

PART I: XML/SGML 과 문서정보 인프라

XML/SGML 문서 출판 모델
XML/SGML 정보 표현의 장점
XML/SGML 기반 문서정보 인프라의 장점

이제까지 널리 알려진 인터넷 웹(Web)은 SGML의 응용으로서 HTML(HyperText Markup Language)이라는 언어를 기반으로 하여 전 세계의 모든 정보를 거대한 정보망의 형태로 묶어 넣으로써 정보의 전달/공유에 대한 기존의 모든 관념을 일시에 무너뜨렸다. 즉 사이버 공간으로 이름 지어진 가상의 공간에서 정보를 교류하고 전달하는 새로운 정보유통 구조를 만들어 내게 된 것이다. 그러나, 누구나 손쉽게 전 세계적으로 분산된 다양한 정보를 손쉽게 접하게 됨으로서, 정보를 습득/이해하고 나아가서는 개인적으로 또는 조직적으로 관리해야 할 정보량이 일순간에 폭발적으로 증가하게 된 것이다. 이러한 정보의 폭발적인 증가는 이미지 또는 HTML에 기반한 정보처리의 한계를 인식하게 하고 필연적으로 정보 콘텐츠의 효율적 관리에 대한 필요성이 대두되게 되었다.

기존의 문서정보 관련 제품들은 이미지 형태 또는 데이터가 특정 응용 제품에 종속되는 상태로 콘텐츠를 관리하게 됨으로써 정보의 재활용 시 정보의 손실이 많았다. 문서정보의 콘텐츠를 관리하고 재활용하려는 입장에서 보면 특정 제품에 종속되는 데이터는 정보의 혼란도가 높아 구조화를 위한 문서 포맷인 XML/SGML에 기반한 콘텐츠 관리보다 재활용성이 매우 낮다. XML/SGML 국제 표준에 기반한 문서정보 구조화는 데이터를 조직화/정규화/지식화 하는 효과를 가지게 되어 문서 정보의 일관성과 재활용성을 높여 준다. 이는 이제까지의 문서 정보처리 방식의 비효율성을 초월하여 정보 획득 기회의 증대 및 정보 재사용을 통한 정보 유통 비용의 감소를 가능하게 하면서 문서정보 인프라로서의 새로운 패러다임을 제공한다. 즉 기존 정보저장 방법의 비일관성, 자료의 중복 및 불일치, 정보접근의 한계 등의 문제로 궁극적으로 발생하는 사용자 정보처리 부하를 줄일 수 있는 가능성을 제공하는 것이다. 특히, 정보구조화를 위한 XML이 각종 인터넷 응용 제품들의 표준 스펙으로 자리를 잡아가고 있기 때문에 미래 지향적 지식화를 지향하는 문서정보 인프라를 구축하기 위해서는 문서의 구조정보를 어떻게 잘 활용하고 서버간 콘텐츠를

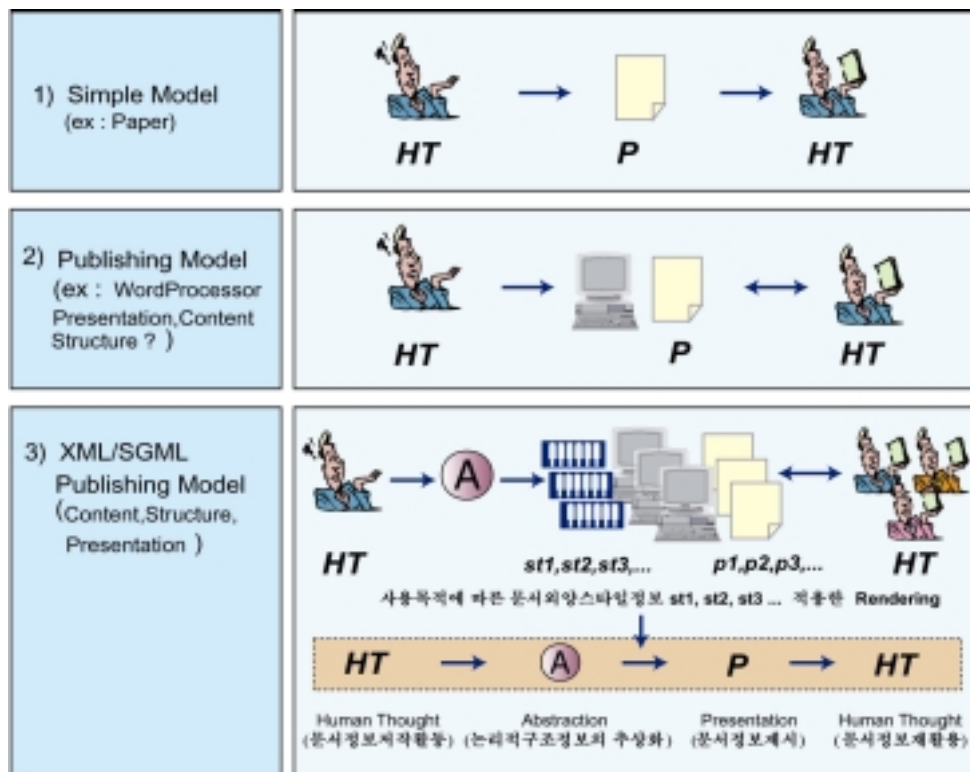
어떻게 잘 통합하고 효율적으로 유통하는가에 대한 문제 해결이 경쟁력 강화를 위한 필수적 고려요소로 대두되고 있다.

구조화된 문서는 주로 텍스트나 그래픽 등의 엔티티 또는 엘리먼트들의 집합으로서 각 엔티티 간의 계층적 및 수평적 관계성에 대한 정의 요소들을 포함한다. 이러한 엔티티들의 단위들은 기존의 문서관리에서는 설계되지 않고 사용되지만 구조화된 문서는 사전에 인간이 사용하기 유용한 단위로 설계되고 정의되어 진다는 점이 기존의 문서 표현 방법과는 다르다. 이는 설계되어진 소프트웨어가 설계되지 않은 소프트웨어들 보다 소프트웨어 활용 측면에서 유용한 것과 마찬가지로 설계된 문서는 생산적이고 활용성이 높은 문서 관리를 가능하게 한다. 이는 문서의 생성, 저장, 검색, 관리, 제시, 교환 등의 여러 단계에서 기존의 문서처리 방법과는 매우 다른 활용성을 지니며 비용 절감 효과와 직결된다. 따라서, 선진 각국의 CALS/EC, 전자도서관, 정부문서관리등의 추진 전략에서는 구조화된 문서 즉 XML/SGML 표준을 채택하고 새로운 천년의 지식경쟁 사회의 새로운 문서 정보 인프라로 준비하고 있다. 즉 정보 통합 및 재사용의 극대화를 통한 정보생산 비용 절감, 정보관리의 효율화를 통한 정보관리 비용 감소, 원하는 정보를 빠르게 검색 및 전달 하여 정보 접근성을 극대화함으로써 각 국의 국가 경쟁력 강화를 위한 XML/SGML 기반 문서 정보 인프라 구축에 적극적으로 투자하고 있다.

XML/SGML 문서 출판 모델

XML/SGML 에 의한 정보 구조화는 즉 한번 생성되고 전달되면 그 생명을 다하는 1 회성의 정보유통이 아닌 정보 생산/유통/재사용의 모든 과정에 있어 좀 더 효율적인 문서 출판 모델에 근거하여 정보의 지식화를 추구한다. 기존의 문서 단위 콘텐츠는 문서의 내용/구조/스타일 정보가 하나의 파일로 특정 제품 종속 데이터로 생성되어 콘텐츠 데이터간 변환에 의해서만 정보 재활용이 가능 하다. 이에 반하여 XML/SGML 에서는 문서의 논리적 구조 정보의 추상화 작업이 선행된 후 사용 목적에 따른 문서의양 스타일 정보를 적용한 Rendering 을 하기 때문에 정보를 제시하기 전 정보 Fragments 단위로의 다양한 정보접근과 동적인 정보 취합이 가능한 Fragments 단위의 콘텐츠 관리를 통하여 보다 유연한 정보 활용성을 보장한다.

그림 1. XML/SGML 에 의한 문서구조화 출판 모델



위 그림 1 에서와 같이 XML/SGML 에 의한 문서 구조화 과정은 인간의 생각 (HT: Human Thought)이 문서로서 표현되고 종이 등의 매체로 출판되는 과정을 문서 데이터의 추상화 또는 지식화 과정과 최종 매체로의 출력을 위한 장식화 과정을 두 단계로 분리한다. 즉 다양한 스타일 정보를 적용하여 다양한 프리젠테이션을 가능하게 하거나 최소한의 비용으로 CD-ROM, Web, RTF, LATEX, PDF 등의 포맷으로 변환하여 다양한 목적으로 Re-tasking 또는 Re-purposing 되게 된다. 따라서, XML/SGML 을 통해 구조화된 정보는 외형적인 형태를 배제하여 기존의 Raw 데이터 보다 지식화 되어 존재하므로 원하는 정보만을 추출하고 전달하며 효율적인 재사용을 보장하게 된다. 태그가 고정된 HTML 과는 달리 XML 에서는 사용자 태그 확장이 가능하여 Self-Describing 한 정보로 사용자에게 다양한 시멘틱스를 제공하는 각종 응용 분야 즉 데이터베이스 정보교환, 검색, 제시 등의 분야에서 활용성이 높다. 또한 이제까지의 워드프로세서나 인터넷 정보의 대부분을 차지하는 HTML 문서는 링크 및 매우 간단한 계층 구조만을 표현하고 있는데 반해 XML/SGML 에 기반한 구조화 문서는 마크업으로 구분된 엘리먼트의 구조적 연관성과 포괄적인 링크 정보를 일관성 있게 잘 표현하고 있기 때문에 사용자 요구에 보다 유연하게 적응 가능한 문서 정

보 인프라를 구축할 수 있게 한다.

XML/SGML 정보 표현의 장점

XML/SGML 로 구조화된 정보 인프라의 장점을 기존의 문서정보 표현 방법들과 개괄적으로 비교하면 표 1 과 같다.

포맷 구분	XML/SGML	PDF/RTF/HWP	TIFF
문서 유형	<ul style="list-style-type: none"> 국제표준 다양한 DTD 수용 	<ul style="list-style-type: none"> 특정 벤더 종속 제한된 문서 수용 	<ul style="list-style-type: none"> 특정 응용 종속
정보의 지식화 정도	<ul style="list-style-type: none"> 국제 표준 기반 마크업에 의한 정보의 지식화 정도 높음 	<ul style="list-style-type: none"> Raw 데이터 수준 	<ul style="list-style-type: none"> 기본적으로 이미지 상태의 Dumb 데이터
정보접근 유연성	<ul style="list-style-type: none"> 사용자가 원하는 문서 엘리먼트 단위별 접근이 가능하여 접근 유연성 높음 	<ul style="list-style-type: none"> 문서 단위로서 접근 유연성 낮음 	<ul style="list-style-type: none"> 이미지 문서단위로서 접근 유연성 낮음
데이터의 가독성	<ul style="list-style-type: none"> ASCII 데이터로 사용자 가독성 높음 	<ul style="list-style-type: none"> 특정 벤더 종속 데이터로 사용자 가독성 낮음 	<ul style="list-style-type: none"> 매우 낮음
문서 저장	<ul style="list-style-type: none"> 내용 및 구조정보통합 저장의 일관성 	<ul style="list-style-type: none"> BLOB 형태 저장 	<ul style="list-style-type: none"> 이미지 기반 BLOB 형태 저장
문서 검색	<ul style="list-style-type: none"> 빠른 전문 텍스트 내용검색 고도의구조정보검색 	<ul style="list-style-type: none"> 텍스트 변환 Filter 필요 구조정보검색불가능 	<ul style="list-style-type: none"> 메타데이터 수준 의 부가정보에의한검색 구조정보검색불가능
문서정보 변환 및 조작	<ul style="list-style-type: none"> 저렴하게 다양한 Back-Ends 로 Re-tasking 이나 Re-purposing 촉진 	<ul style="list-style-type: none"> Re-tasking 이나 Re-purposing 에 한계 	<ul style="list-style-type: none"> 거의 불가능
스타일/Display	<ul style="list-style-type: none"> 다양한 스타일 조작 (DSSSL/XSL) 에 의한 Multiple Display 	<ul style="list-style-type: none"> One Display Only 	<ul style="list-style-type: none"> One Display Only
문서교환 Network Bandwidth	<ul style="list-style-type: none"> Low 	<ul style="list-style-type: none"> Relatively High 	<ul style="list-style-type: none"> High
정보 링크 /Navigation	<ul style="list-style-type: none"> XLL/HyTime 에 의한 일관된 링크 관리 	<ul style="list-style-type: none"> 특정벤더종속 및 일관성 부족 	<ul style="list-style-type: none"> 이미지 단위

PDF : Portable Documents Format, RTF : Rich Text Format, TIFF : Tag Image File Format

표 1. 문서정보 인프라를 위한 정보 표현 방법 비교

XML/SGML 로 구조화된 문서정보 표현 방법은 벤더 독립적인 아스키 데이터로 사용자가 XML/SGML 구문에 대해 알면 쉽게 문서정보 데이터를 이해 할 수 있고, 다양한 정보 접근 단위로 정보 조작이 가능하며, 정보 교환 시에도 Network Bandwidth 이 상대적으로 낮다. 또한 Self-describing 한 정보를 포함하여 정보의 지

식화 정도가 높아 정보 생성자가 예측하지 못한 활용성을 지닌다.

XML/SGML 기반 문서정보 인프라의 장점

표 2 는 새로운 XML/SGML 기반 문서 정보 인프라를 기존의 접근 인프라들과 요약하여 비교한 표이다.

구 분	XML/SGML 문서정보 인프라	기존 문서정보 인프라
정보 생성 방법	다사용자에 의한 엘리먼트별 동시 공동 저작 가능	문서 별 개별적 문서 생성
정보 저장 및 확 장성	XML/SGML 의 DTD 설계에 따른 정보 유연성을 살린 확장성 제공	고정된 스키마 이외의 정보 추가시 전체 시스템 재구성 필요
구조 정보검색 수 행 능력	지정된 모든 엘리먼트 항목별 구조 및 애트리뷰트 검색 지원	5-6 개 정도의 항목 분류를 통한 제한된 정보 검색 제공
정보 관리	XML/SGML 저장 및 관리를 위한 Repository 서버를 통한 엘리먼트별 라이브러리 (Check-in/ Check- out/Versioning)서비스기능	문서 단위의 메타데이터에 대한 수 작업을 통한 관리
통계정보 추출	다양한 자동 통계정보 추출	수작업을 통한 통계정보 추출
정보 수정	개별 정보 수정자가 직접 수정	시스템관리자를 통한 수정
타기관과의 정보 공동 활용	기관간 정보 메시징 기능에 의해 타기관에서 직접 목록을 등록/수정	단순 목록보기
분산 DB 간 연동	Messaging Oriented Middleware 에 의 해 XML/SGML 정보의 상호 연동 이 가능하며 응용 프로그램에서의 정보 재사용율이 높음	스키마가 고정되어 DB 간 정보의 상호 연동이 비효율적임
유지 보수 비용	정보 공유 및 재사용의 촉진으로 수작업 최소화하며 통합된 정보 콘 텐츠 작성 비용의 최소화	지속적인 수작업 및 정보 통합 비 용 증가

표 2. XML/SGML 기반 문서정보 인프라와 기존 인프라와의 비교

본 고에서의 TheoCMS 는 미래 지향적인 구조화된 정보 관리를 위한 서버 사이드
의 인프라 기능을 충족하는 XML/SGML 기반 문서정보의 생성, 저장, 검색, 관리,
전달을 지원하는 문서정보 인프라이다. 지식화된 조직을 위한 고유의 목적을 달성

하기 위해서는 습득되고 생산되는 정보가 홍수처럼 증가해 가고 있는 상황에 대처하여 정보의 고립을 방지하고 정보과다 상황에 적합한 새로운 정보 인프라를 구축해야만 한다. 그 첫 단계로 제시되는 것이 XML/SGML 을 이용한 정보의 구조화이다. 그러나 아직 대부분의 조직내 80-90%의 문서정보가 데이터베이스 등의 조직화된 형태로 관리되지 않고 있는 것이 현실이다. XML/SGML 기반 문서구조화 기술은 조직화되지 않은 문서정보의 혼란도(entropy)를 낮추어 주도록 문서 정보 데이터에 마크업을 추가하여 정보의 조직화/정규화/지능화를 추구하는 새로운 대안이다. 이는 결과적으로 문서정보의 적절한 통제와 저장/변경 그리고 원활한 정보유통이 가능한 문서 정보 인프라를 제공한다.

XML/SGML 에 의한 정보 구조화는 기존 문서 단위 보다 세분화된 정보의 단위(Fragments)로 즉 보다 인간이 정보를 다루고 활용하기 적합한 단위로 문서를 재정의하는 과정이다. 정보의 구조화는 정보를 분자와도 같이 가장 작은 단위로 분할하고 이를 구조적으로 통합하는 방식으로 각 단위 정보들이 서로 분리되어 통합정보가 달라지더라도 단위정보(엘리먼트)자체에 기술된 정보는 그대로 유지하도록 하는 것이 XML/SGML 에 의한 정보 구조화의 역할이다. 단순한 레고블럭이 모여 점차로 큰 단위의 모양을 이루고 이러한 모양이 합쳐 전체의 커다란 구조물을 이루는 것과 같은 것으로 전체 구조물을 전체를 하나의 단위로 정보 인프라를 구축하는 것과는 정보의 Fragments 의 유연한 재사용 측면에서 큰 차이를 보인다. 정보의 통제와 저장을 위해서는 정보가 되도록 큰 단위로 존재하는 것이 유리하지만 정보유통과 변경을 위해서는 정보는 세분화 될 수 있는 가능한 작은 단위로 존재하는 것이 유리하다. 정보가 너무 큰 단위로 존재하면 정보의 변경과 유통을 위해 한번에 움직이고 처리해야 할 정보의 양이 너무 많아지게 된다. 기존의 콘텐츠 관리 시스템은 정보의 단위가 정해진 단위로 미리 고정되어 사용자로 하여금 정해진 정보의 단위에 맞게 활용하는 방안을 요구하는 부담을 주게 된다. 예를 들어 웹 검색엔진의 경우 웹 페이지 단위로 검색이 이루어지거나, 도서검색시스템의 경우 도서단위로 정보를 관리한다. 그러나 이러한 단위의 제한은 정보의 원활한 재사용을 저해하며 한번 구축한 정보를 사용하는 응용 프로그램의 요구사항이 바뀌면 (원하는 정보단위가 바뀌면), 정보 시스템 전체를 다시 구축하고 정보도 이에 맞게 다시 구축하는 작업을 반복적으로 하는 경우가 자주 발생한다. XML/SGML 은 정보 내용에 마크업을 추가하여 구조화함으로써 정보자체가 원하는 작은 정보단위로 유연하게 나뉠 수 있고

이를 원하는 단위로 다시 통합할 수 있게 한다. 따라서 사용자가 원하는 단위로 모든 관리가 이루어 질 수 있어 응용에 대한 요구사항이 바뀌어도 유연하게 대처할 수 있다. 이는 정보의 지식화를 위해 기존과는 다른 다음과 같은 새로운 정보관리 및 활용 환경을 가능하게 한다.

▪ 공동저작(collaborative authoring)을 통한 정보생성의 효율화

정보의 생성은 제품을 생산하는 생산라인에 비유할 수 있다. 즉 다듬어 지지 않은 생각을 가진 다수의 생성자가 서로가 맡은 부분에 대한 생각을 정리하여 작성하는 과정이다. 기존의 정보관리 환경에서는 한 생성자가 한 정보의 생성을 모두 담당하거나 부분을 작업한 후 이를 다시 통합하는 과정을 여러 번 거치면서 정보를 생성한다. 이러한 과정은 산업혁명이전의 형태와 유사한 프로세스로서 많은 비효율을 내포하고 있다. 레고블럭과 같이 단위정보를 바탕으로 하여 논리적 분할하에 각자의 정보를 생성하고 이러한 정보를 통합함으로써 정보를 생성할 수 있도록 해 주는 것이 정보의 구조화이다. 레고블럭을 가지고 중세의 성을 만들고자 한다면 성벽, 성문, 정원 등을 분담하여 구축하고 이를 연결하여 전체 큰 구조물을 만들어 낼 수 있다. 이는 레고블럭이라는 기본 단위를 서로 묶어서 점차로 확장시켜갈 수 있기 때문이다. 이러한 기본단위의 확장과 통합이 바로 정보의 구조화이다. 거대한 성이 완성되면 성벽 수정등의 일부 수정이 가능한 것도 이러한 논리적 분할에 따른 구조화에 기인한다. 대량의 정보를 생성하는 환경에서 전체와 부분의 통합을 보장하여 주는 정보인프라가 구축된다면 정보생성에서의 비효율은 산업혁명과도 같은 정보혁명을 촉진시켜 정보생성 비용을 현저히 감소 시킬 수 있다. 또한, XML/SGML에 의한 문서 정보의 생성은 Raw 데이터에 마크업을 추가하여 지식화 함으로써 문서 사이클상의 여러 단계에서 다수 사용자에게 의한 지속적인 정보 활용 효과가 매우 높아 질 수 있는 기반을 마련해 준다.

▪ 정보통제와 유통의 효율화

정보통제는 앞서 기술한 바와 같이 정보단위가 커질수록 효율이 향상된다. 그러나 유통 시에 큰 단위의 정보는 정보시스템의 부하를 증가시키게 된다. 그러나 구조화된 정보라면 통제되는 단위와 유통되는 단위를 모두 가변적으로 유지할 수 있다. 즉 행정 단위를 시.도-읍.면-동.리와 같이 원하는 크기로 세분하여 사용하듯이

XML/SGML 로 구조화된 정보는 사용자가 원하는 단위로 정보를 세분화할 수 있고 이들의 포함관계를 다룰 수 있으므로, 통제와 유통에 있어서의 효율성을 모두 극대화 할 수 있다. 이러한 효율성은 구조화된 문서의 생성 과정이 내용 뿐아니라 문서의 활용성을 고려한 문서 구조 및 링크의 설계 과정을 통한 데이터의 지식화 과정을 포함하는데 기인한다.

▪ 정보 과부하에 대한 정보검색(Information Retrieval)의 효율화

일반적으로 정보검색은 대량의 정보 속에서 정보사용자가 원하는 정보를 빠르게 찾아낼 수 있도록 도와주는 중요한 정보인프라 중의 하나이다. 정보검색 시스템은 정보를 색인하여 색인구조로 저장하고 이를 바탕으로 정보의 내용에 대한 질의를 통해 정보를 찾도록 도와주는 시스템이다. 정보검색 시스템의 성능은 사용자가 원하는 정보를 얼마만큼 효율적으로 찾아내느냐에 달려있다. 이러한 정보검색 시스템의 성능은 두 가지의 구성요소를 갖고 있다. 정보를 얼마나 빠르게 찾아내느냐에 관한 정보탐색(searching) 요소와 더불어 정보검색 시스템 자체의 정보표현 능력이 어느 정도인가에 관한 정보표현(representation or indexing) 요소가 있다. 도서관시스템을 예로 들면, 아무리 대량의 도서를 빠르게 찾아내더라도 책제목으로만 찾아낼 수 있다면 정작 사용자는 자신이 원하는 정보를 빠르게 찾아 낼 수 없고, 수많은 검색 결과리스트를 헤매다가 진정 원하는 정보를 찾는 것을 포기하게 될 것이다. 만약 저자정보 혹은 출판사, 요약정보로도 검색이 가능하다면 사용자에게는 자신이 질의할 수 있는 영역자체가 확장되므로 좀 더 정확한 정보를 빠르게 찾아낼 수가 있다. XML/SGML 에 의한 정보의 구조화는 정보표현 영역을 넓혀 사용자가 자신이 원하는 정보를 정확하게 표현할 수 있게 하여 원하는 정보만을 정확하게 찾아낼 수 있는 방법을 제공한다.

▪ 효율적 정보 관리, 전달 및 정보의 지식화

문서 정보의 생성, 관리, 검색, 전달, 유통의 과정에서 구조화 정보를 바탕으로 한 정보시스템이 구축되어 연동될 때 앞서 기술한 모든 효용을 얻고 조직의 개인에게는 정보 과부하 상태에서 지식화된 정보를 얻고 활용할 수 있는 기반이 마련된다. 정보 구조화 및 지식화를 위해서는 기존의 정보시스템에 대한 한계성과 비효율성을 사용자 자신이 인식하고 이를 새로운 방향으로 모색하기 위한 인식이 전제되어야

한다. 자신의 상황에 필요한 정보구조화의 방향과 정확한 응용처에 대한 인식이 새로운 정보혁명에 진입하는 가장 중요한 요소가 된다. 정보의 구조화는 정보 재사용과 유통에 대한 고려를 정보의 생성 단계에서부터 필수적으로 동반하게 된다. DTD(Document Type Definition)으로 불리는 일종의 정보 템플릿(Information Template)에서부터 정보의 재사용과 유통, 향후 발전 방향에 대한 모든 고려가 통합되어야 한다. 이러한 고려 외에 정보를 생성하는 정보생성자 개개인 모두가 이러한 정보템플릿에 대한 이해를 바탕으로 정보를 생성해야 한다. 즉 한번 생성되고 버려지는 정보가 아닌 영속성을 가진 유기적 생명체적인 정보를 조직에 맞도록 생성하고 사용하는 고려가 전제되어야 한다. 이와 같이 인적인 요소들과 연계되는 체제에서의 XML/SGML 에 기반한 정보 구조화는 이러한 정보의 지식화를 위한 틀을 제공하게 되는 것이다.

▪ 효율적 응용프로그램간의 연계 (XML/SGML and Messaging Oriented Middleware)

이제까지 HTML 에 기반한 웹 응용 프로그램의 한계점은 각 응용 시스템간의 연결고리가 부족하다는 점이었다. HTML 문서는 H1, H2, DD 등의 태그만으로 데이터의 성격을 알 수 있는 방법이 제한되어 있어 호스트 시스템들 간의 응용 프로그램들을 연결하는데 제한이 있었다.

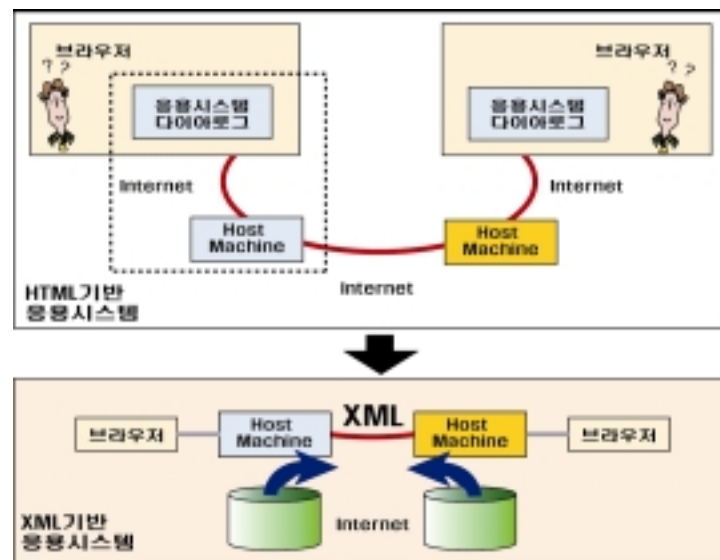


그림 2. XML 기반 Messaging Oriented Middleware

XML/SGML 에 의해 표현된 정보는 HTML 정보와 비교하여 Self-describing 한 정

보로서 해당 정보의 시멘틱스를 자체적으로 알수 있다. 예를들어 그림 3 에서 HTML 문서와 XML 문서는 정확히 같은 내용을 정보사용자에게 보여주고 있다. 그러나 정보를 이해하는 응용 프로그램 또는 사용자 입장에서 보면 XML/SGML 로 되어 있는 문서를 통해 ‘客口’라는 것이 책의 이름을 의미하고, 저자가 ‘황석영’ 이라는 정보의 이해를 훨씬 쉽게 할 수 있다. 따라서, XML/SGML 로 기술된 정보는 응용 시스템간의 연계성을 높일 수 있는 접근법을 제공한다. 특히 XML 은 웹 환경에서 SGML 의 간략화된 스펙으로서 Self-describing 한 특성으로 인해 각종 데이터베이스들 간의 정보 교환 또는 정보를 응용프로그램에서 활용 하기 위해 매우 유리하다.

<pre> <HTML> <BODY> <TABLE> <TR> <TD> 客口 </TD> <TD> 황석영 </TD> <TD> 창작과 비평사 </TD> <TD> 1974 </TD> </TR> </TABLE> </BODY> </HTML> </pre>	<pre> <?XML version="1.0"?> <단행본> <도서명> 客口 </도서명> <저자> 황석영 </저자> <출판사> 창작과 비평사 </출판사> <출판년도> 1974 </출판년도> </단행본> </pre>
HTML 문서	XML 문서

그림 3. HTML 과 XML 예제 문서

이와 같이 Messaging Oriented Middleware 는 응용 프로그램들 간에 연결고리를 담당하는 소프트웨어를 통칭한다. 데이터베이스 정보들간의 연계하거나 각 응용 프로그램 간에서 비교적 짧은 정보 내용들을 변환하여 로컬 프로그램 환경에 적합한 형태로 데이터를 생성한다. 일반적으로 Two-tier Client/Server 모델에서 Middle Tier 가 추가되어 Third Tier 형태로 작동된다. 이때 XML 형태로 Middle Tier 에서 정보를 변환하여 Client 응용 프로그램으로 정보가 전달된다. 이러한 XML 에 기반한 Messaging Oriented Middleware 는 특히 네트워크 환경에서 Web Traffic 양을 줄일 수 있으며 엔티티 관리 기능에 의하여 분산되어 있는 관련 있는 데이터를 일관된 프레임워크에 의해 정보를 교환 및 활용하는 효과를 창출할 수 있다. 이때 전달되는 XML 문서의 형태는 비교적 양이 적고 간단한 문서 구조를 가진 경우에 해당되며 장기적으로 관리되고 Mission Critical 한 긴 문서인 경우에는 XML/SGML Enabled CMS 와 연계되어 관리되어야 한다.

최근 전 세계적으로 관심이 높은 전자 상거래에서의 대부분의 응용들은 Repository 기능을 갖는 CMS, 서버간 정보 교환 및 변환을 위한 Middleware, 응용프로그램들로 구성되어 지며 XML 형태의 정보는 보다 지식화되고 효율적인 정보 교환 및 제시 측면에서 활용성이 높아 오라클, 선, IBM 등의 유명 IT 업체에서 경쟁적으로 XML 스펙지원을 발표하고 있다. 예를들어 그림 3 의 문서들 경우 도서명과 저자만을 따로 모아 리스트를 만들고자 한다면 HTML 문서는 추가 수작업이 필요한 반면 XML 문서에서는 Self-describing 한 정보를 이용하여 자동으로 수행할 수 있다.

이어지는 PARTII 에서는 시스템 측면에서의 구조화된 문서정보 관리를 위해 요구되는 기능들과 CALS/EC 등의 응용에 대하여 살펴보고, PARTIII 에서는 지식관리를 위한 정보관리 솔루션으로서의 TheoCMS 에 대하여 정리하고자 한다.

PART II : XML/SGML Enabled CMS 와 CALS/EC

XML/SGML Enabled CMS CALS/EC 에서의 CMS 적용 CALS/EC 에서의 CMS 적용 효과

XML/SGML 을 활용한 CALS/EC 에 적용되는 정보 시스템을 구축하기 위해서는 정보 구조화를 위한 인적요소 뿐만 아니라 이러한 인적요소를 뒷받침할 수 인프라 시스템들이 새로이 도입되어야 한다. 이러한 시스템은 XML/SGML 로 표현된 구조화된 정보의 특성을 반영하여 가장 적절하게 구현하고 기능을 제공하는 시스템으로서 이를 위한 가장 중요한 두 기능이 구조화 정보검색과 정보관리를 위한 정보 저장소(Repository) 기능이다. 이러한 정보저장소 기능을 갖는 XML/SGML Enabled CMS 를 위해 기본적으로 요구되는 전제는 다음과 같다.

첫째, 기존의 정보검색의 단순확장이 아닌 XML/SGML 정보 구조화의 패러다임을 그대로 유지시킬 수 있는 새로운 정보검색 시스템이어야 한다. 엘리먼트 단위로 세분된 정보를 분할하여 색인 검색할 수 있고 이러한 엘리먼트 간의 논리적 관계도 검색에서 사용할 수 있는 검색시스템 기능이 제공되어야 한다.

둘째, XML/SGML 이 갖는 정보의 논리적 분할과 통합에 대한 메타포어를 모두 제공하면서 구조화된 정보 자체를 정보손실 없이 저장하고 수정하며 관리할 수 있는 정보 저장소 기능이 제공되어야 한다. 구조화정보 관리 시스템은 엘리먼트라는 정보 블록으로 이루어진 정보를 분할/통합/수정/제시할 수 있는 기본 기능이 제공되어야 한다.

셋째, CMS 는 정보검색 기능과 정보저장소의 기능의 단순 연동이 아닌 상호 융합되어 문서 사이클상에서 정보의 지식화된 활용을 지원하기 위한 지식관리 차원의 기능들을 지원하여야 한다.

XML/SGML Enabled CMS

정보는 그 생성, 유통, 저장, 검색 및 활용에 있어 나름대로의 생명주기를 가진다. 또한 정보의 생명주기는 그 정보가 활용되는 맥락과 이용하는 정보기술에 따라

다양한 형태로 구성되게 된다. 개인용 컴퓨터, 워드프로세싱 프로그램, 탁상용 전자출판 등의 활용 및 인터넷 사용의 폭발적 증가에 따라 요즘의 정보 생명주기는 이전의 종이에 인쇄한 문서가 유통의 주된 형태였던 시기와는 전혀 다른 양상을 나타내고 있다. 1 회적인 정보의 생산 및 유통에 끝나던 과거와는 달리, 한번 생성된 정보는 문서로의 출력, 인터넷을 통한 검색 및 서비스 등 다양한 형태로 재활용되고, 변경되면서 그 생산성을 극도로 활용할 것이 요구되고 있으며, 이는 기업이나 정부 등 조직에 있어 경쟁력 확보를 위한 절실한 요구 사항이 되고 있는 상황이다. 여기에서 정보의 전 생명주기에 걸쳐 다양하고 탄력성있는 관리 기능을 제공해 줄 수 있는 정보 인프라의 필요성이 절실히 요구되고 있다.

또한 현재의 정보는 과거와는 달리 텍스트, 이미지, 비디오 등의 다양한 콘텐츠가 나름대로의 구조와 논리적 의미를 가지고 혼재되어 있고, 이들 콘텐츠는 또 다른 정보의 일부분으로서도 활용되어지는 등 복합적인 전자적 형태를 이루고 있는 것이 보통이다. 따라서 전체적인 하나의 단위로서의 정보가 아니라, 그 정보를 구성하고 있는 개개의 논리적 콘텐츠를 관리 단위로 하여 이를 전체적으로 관리할 필요성이 있다.

CMS 는 복합적인 정보 생명 주기에 걸쳐 조직의 복잡한 정보 요구에 대한 솔루션으로 제시되었다. CMS 를 통해 조직은 자신의 소중한 자산인 정보를 재활용 가능한 형태로 생성, 저장, 검색, 관리하면서 필요한 다양한 형태의 정보 요구에 유연하게 대처할 수 있게 될 뿐 아니라, 정보 관리에 필요한 비용을 최소화할 수 있게 된다. 예를 들어 백과사전을 관리하기 위해 CMS 를 적용할 경우를 생각해 보자. 백과사전에 있어서는 최신의 정확한 정보가 수록되어 있을 것이 경쟁력을 결정짓는 요소가 될 것이다. 백과사전은 수만가지 이상의 항목으로 구성되어 있어, 기존의 수작업이나 문서관리시스템으로 관리할 경우, 보통 하나의 판본을 출판하고 다음의 변경된 판본을 출판할 때까지 약 반년에서 1 년의 시간이 걸리므로, 새로운 정보가 쏟아져 나오는 요즘의 정보 환경에 발빠르게 대처하기가 어렵게 되며, 이는 경쟁력 저하의 요인으로 작용하게 된다. 또 이를 CD-ROM 이나 인터넷을 통해 서비스하고자 할 경우에는 이를 위한 새로운 데이터 입력이나 번거로운 변환 작업이 필요하게 된다. 또한 새로운 항목을 추가하거나, 이전의 항목을 변경하는 경우 다른 항목과의 링크 등 정합성 유지에 실제적인 문제점이 드러나게 되며, 이를 적절히 제어할 수

단은 기존의 관리시스템에서는 제공되지 않는다. CMS 를 적용할 경우, 백과사전에 담겨 있는 각 정보가 서로간의 연관성을 유지하면서 관리되게 되고, 다수의 저자에 의한 공동 저작이 가능하게 되므로, 새로운 항목의 추가나 이전 항목의 변경이 손쉽게 이루어져, 관리되는 정보를 항상 최신의 상태로 유지할 수 있게 된다. 또한 CMS 는 정보의 콘텐츠 만을 관리하므로, 필요할 경우 인쇄, CD-ROM, 인터넷 등을 통해 필요한 부분 또는 전체를 서비스할 수 있어 다양한 형태의 출판이 가능하게 되며, . 이를 통해 하나의 판본에서 다음 판본을 출판하는 데 걸리는 시간이 거의 걸리지 않게 되어, 새로운 정보의 정확한 공급이라는 백과사전 본래의 목적에 충실하게 되어 경쟁력 향상에 매우 큰 도움이 된다.

XML/SGML Enabled CMS 를 구성하는 주요 요소들은 다음과 같다.

▪ 데이터 Repository

관리할 정보를 저장하고 요청시 이를 제공하는 구성 요소이며, 정보의 저장과 제공을 신속하고 효율적으로 수행할 수 있어야 하며, 이와 더불어 권한별 접근 제어 기능도 제공해 주어야 한다. 통상 Repository 내부의 정보 구성은 하위 콘텐츠 구성 단위까지 분할되어 저장하고 관련된 엔티티에 대한 관리 기능을 갖는다. 이는 응용의 특성에 따라 관리되어야 할 분할의 단위가 다양한 형태로 구현될 수 있을 것이다. 또한 하부 저장 구조로는 파일시스템, 관계형 DBMS 또는 객체지향형 DBMS 등이 이용될 수 있다.

▪ 사용자 인터페이스

CMS 가 관리하고 있는 정보를 사용자가 손쉽게 브라우징하고 찾아 볼 수 있도록 하는 구성 요소이다. 이는 인터넷 브라우저나 아래의 편집 도구를 통해서 제공될 수도 있고, 별개의 전용 프로그램을 이용해 접속할 수도 있을 것이다. 또한 필요시 독자적인 인터페이스를 구현할 수 있는 라이브러리 등의 도구도 제공되어야 한다.

▪ 편집 도구

CMS 가 관리하게 될 정보의 생성 또는 관리하고 있는 정보의 편집을 위한 도구이

다. 관리할 정보의 포맷에 따라 다양한 프로그램이 사용될 수 있고, 필요에 따라 전용 프로그램을 구현하여 활용할 수 있다.

▪ 워크플로우

통상 CMS 가 관리할 정보는 다수의 사용자에게 의해 생성, 변경, 검색되게 된다. 이러한 다수 사용자 환경에서 각 콘텐츠를 구성하고 있는 요소들이 현재 어떠한 상태에 있는가를 적절히 추적하고 통제, 관리하는 것은 정보의 정합성 유지를 위해 필수적인 일이며, 이러한 기능을 제공하는 구성 요소가 워크플로우이다. 여기에는 콘텐츠 locking 에 기반한 Check-out/Check-in, 접근관리, 버전 관리 및 나아가 정보 라우팅 및 통보 서비스등이 포함되게 된다.

▪ 검색 엔진

CMS 가 관리하고 있는 정보를 사용자가 손쉽게 활용하기 위해서는 다수의 정보 중 자신이 원하는 정보를 정확하고 신속하게 검색할 필요성은 절대적이며, 이 경우 지원되어야 할 검색 단위는 전체로서의 정보가 아니라 정보를 이루고 있는 각 콘텐츠 단위로 검색할 필요성이 있다. 따라서 CMS 에서 제공되어야 할 검색 서비스는 이제까지의 전문 검색을 넘어, 각 콘텐츠간의 관계성 및 이들에 대한 관리 정보등을 포함하여 복합적인 형태의 검색 기능이 되어야 한다.

▪ 출력 도구

출력 도구는 데이터 Repository 에 저장되어 있는 정보를 특정한 사용 유형에 맞도록 포맷해 주는 도구이다. 즉 동일한 정보가 인쇄, CD-ROM, 인터넷 등을 통해 다양한 형태로 출판될 수 있고, 또 필요시 정보의 일부분만을 선택적으로 출판할 수도 있게 된다. 사용자의 요구에 맞는 출력이 가능하도록 다양한 형태의 출력 도구를 구비할 것이 요구된다.

이러한 CMS 는 대량의 정보를 서로의 연관성을 유지하면서 관리할 필요성이 있는 경우, 다수의 저자에 의해 동일한 정보 집단이 생성, 변경되면서 정보의 내용이 빈번하게 변화하는 경우, 동일한 정보를 다양한 형태로 서비스할 필요성이 있는 경우 등에 있어 최상의 해결책으로 간주되고 있다.

CALS/EC 에서의 CMS 적용

이제까지의 CALS/EC 를 추진 함에 있어 기존의 문서 관리시스템에서 문서 관리의 단위는 단일체로서의 문서단위 관리를하는 파라다임에 근거하여 발전되어 왔다. 즉 다양한 프로그램을 이용해 생성된 문서들을 하나의 객체로 간주하고, 적절한 필터를 이용해 이들의 내용을 추출해 색인, 검색하고 관리하는 것이 주된 목적이었다. 하지만 이러한 문서 단위의 관리는 다양한 콘텐츠를 이용해 복합적으로 구성되어 있는 요즘의 문서를 고려하면, 필요한 문서 일부만을 재활용할 요구가 있을 때 유연하게 대처할 수 없고, 문서에 담겨 있는 정보를 다른 목적에 사용하기 위해서는 비용과 시간이 많이 소요되는 등 현재 요구되고 있는 정보 수요에 효과적으로 대처할 수 없는 것이 문제점으로 지적되고 있다.

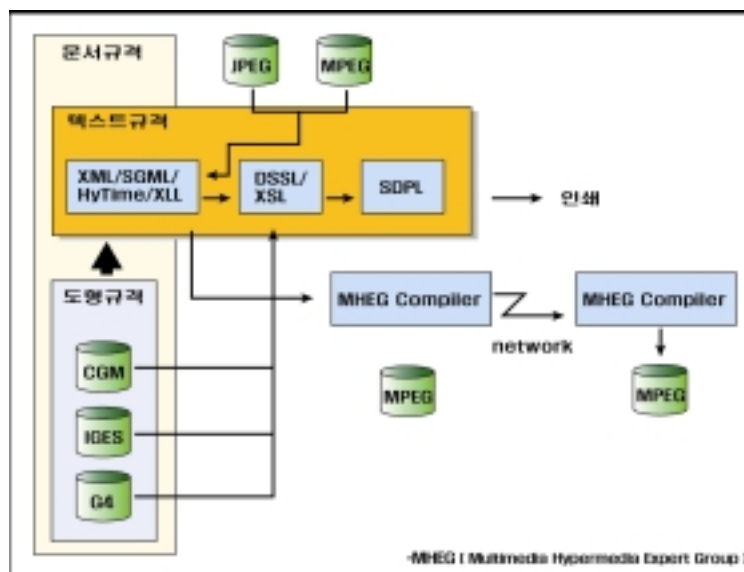


그림 4. 문서처리를 위한 CALS/EC 표준

이에 반해서 XML/SGML Enabled CMS 를 적용할 경우 동일한 정보를 이용해 다양한 유형의 목적에 용이하게 활용이 가능하고, 필요할 경우 정보의 일부만을 추출하여 서비스할 수 있는 등 기존의 문서 관리시스템에 비해 보다 다양한 유형의 서비스를 적은 비용으로 제공할 수 있다. XML/SGML Enabled CMS 는 CALS/EC 를 구현하기 위해 미래 지향적으로 요구되는 스펙을 포괄하는 기능을 구현 서버 모듈이다. 예를들어, Business to Customer 및 Business to Business 전자상거래를 실현 시키기 위해서는 XML/EDI Repository 모듈을 필요로 하는데 이를 지원하기 위한 인프라로서 CMS

기능을 필수적으로 요구한다. XML/SGML 스펙은 그림 4에 나타난 대로 이미지, 비디오, 오디오, 도면 등의 멀티미디어 스펙을 포괄하고 제품 데이터 STEP 표준과 연계하여 다양한 응용이 가능하여 각 표준이 통합된 CALS/EC를 구현하기 위해서 XML/SGML 관련 표준 스펙을 포괄하는 CMS가 전체 인프라의 구심점 역할을 수행한다. 따라서, CALS/EC를 추진하는 효과로서 매년 20-40% 각국의 GNP에 해당하는 문서정보 생성, 유지, 관리하는 비용을 절감하는데에 XML/SGML Enabled CMS가 기여하는 바가 매우 높음을 유추해 볼 수 있다. 이는 곧 CALS/EC의 주된 목표인 기업 및 국가의 환경 변화에 대응되는 신속한 광속 거래의 실현을 위하여 고객의 요구에 따른 신속한 정보의 제공과 국제표준에 기반하여 국제간의 상거래 활성화를 추진할 정보 인프라로서 XML/SGML Enabled CMS에 대한 요구가 높아져가고 있는 것이다. 다음 그림 5는 CALS/EC를 추진하는 기업의 개발, 생산, 영업, 서비스 부분을 통괄하여 XML/SGML 스펙이 응용되고 있는 영역을 나타내어 주고 있는데 CMS는 각 부문의 정보를 저장하고 관리하는 Repository 모듈로 또는 지식관리 정보 저장소로 데이터베이스와 연동되어 활용된다.

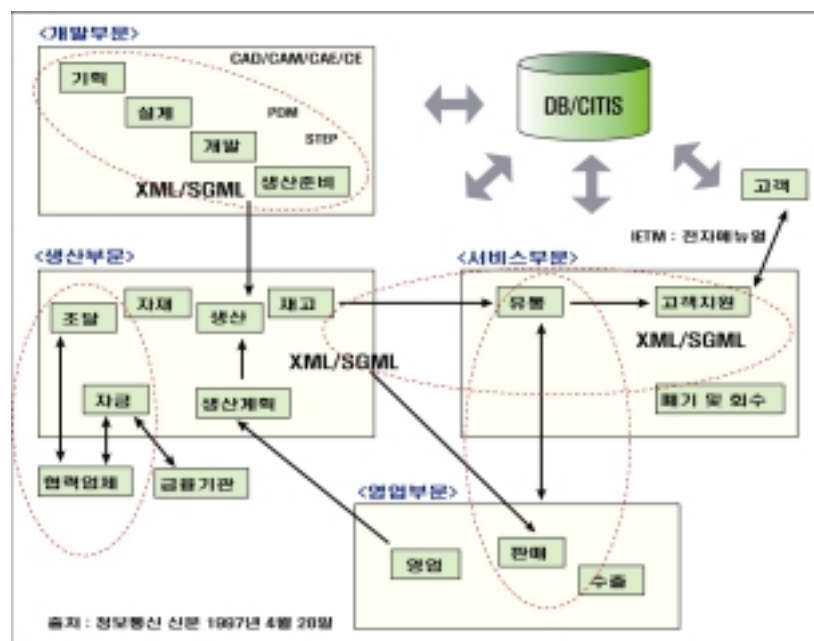


그림 5. XML/SGML 표준의 CALS/EC 적용 부문 예시

XML/SGML Enabled CMS의 부문별 적용 분야는 다음과 같이 나누어 지면서 발전 될 것으로 예측된다.

- CITIS 를 위한 CMS (Government to Business Solution)
- EC 를 위한 CMS (Business to Business, Business to Customer Solution)
- 기업 정보 포탈 (CIP:Corporate Information Portal) 서비스를 위한 CMS
- 전자도서관 정보의 지적 관리를 위한 CMS

CALS/EC 에서의 CMS 적용 효과

CALS/EC 를 추진함에 있어 다양한 분야에서 XML/SGML Enabled CMS 를 적용하여 얻을 수 있는 문서처리 관점에서의 구체적 효과를 정리하여 보면 다음과 같다.

▪ 정보 변경 비용 감소

CMS 에 저장되어 있는 정보중 변경이 필요한 부분만을 선택적으로 변경할 수 있고, 최신의 상태로 유지되는 정보는 다양한 유형으로 즉각적으로 서비스될 수 있다. 이는 문서 단위로 정보가 변경되어야 하는 기존 문서 관리시스템에 비하면 정보의 생성 및 변환에 드는 비용을 크게 줄일 수 있으며, 이러한 서비스에 드는 비용은 통상 50% 정도로 생각되고 있다.

▪ 항상 사용 가능한 콘텐츠

CMS 가 관리하고 있는 정보는 특정 유형의 활용이나, 특정 포맷으로 이루어진 것이 아니고, 다양한 활용이 가능한 중립적인 포맷으로 유지되고 있으므로 필요시 항상 최신의 정보를 서비스할 수 있는 상태를 유지할 수 있으며, 이는 정보 서비스에 있어 경쟁력 향상에 지대한 요인이 될 수 있다.

▪ 콘텐츠의 status 에 대한 효과적인 관리

현재 정보를 구성하고 있는 각 콘텐츠의 status 를 즉각적으로 파악할 수 있게 되어, 최종 출판물 생성, 마케팅 및 경영상 이러한 정보를 이용할 수 있게 됨으로써 다양한 이점을 확보할 수 있게 된다.

▪ 출판 비용 감소

CMS 를 이용하면, 최종 출판에 있어 최대의 유연성을 확보할 수 있는 형태로 정보가 유지, 관리되기 때문에 실제 출판에 드는 비용을 절감할 수 있고, 또한 출판 요구상에 변경이 있더라도 최소의 비용으로 이에 대응할 수 있게 된다.

▪ **맞춤 출판에 대한 적응성**

CMS 를 채용함으로써 이제까지는 비용 문제로 인해 거부되었던 맞춤 출판에 대한 기회를 최대한 확보할 수 있게 된다. 인터넷의 폭발과 더불어 동일한 정보를 더욱 다양한 형태로, 또 동일 정보내에서도 상황에 따라 필요한 부분만을 발췌하여 출판할 필요성이 더욱 더 증대하고 있으며, CMS 를 이용함으로써 이러한 상황에 더욱 효과적으로 대처할 수 있게 된다.

이러한 CMS 는 일반적인 CALS/EC 뿐 아니라 전자도서관, 의료, 금융 등의 각종 구체적 응용 시스템의 기본 인프라로서 다양하고 유연한 정보서비스 구현을 위해 비용, 시간적인 면에서 다양한 장점을 제공해 주고 있다.

PART III : 지식지향 정보인프라를 위한 CMS 솔루션- TheoCMS®

구조화정보 관리시스템의 구현 요건

TheoCMS : 시스템 구성도

TheoCMS : 기능 목록

TheoCMS : 사양

TheoCMS의 주요 고객은 기존의 정보 검색 및 문서 관리 분야의 전문가들이다. 최종 사용자 측면에서 보면 중장기적으로 볼 때 정보검색 기능과 문서관리 기능을 이용하는 지식 엔지니어라고 통칭될 수 있는 계층이다. 이제까지의 정보 검색 시스템들은 정확도와 재현율이라고 하는 성능 지표로서 평가되어 왔지만 점점 인터넷/인트라넷/엑스트라넷 환경에서의 정보검색은 급격히 늘어나는 정보량에 대비한 대용량 처리와 분산환경에서의 다양한 정보원을 처리하는 기능에 대한 성능이 점점 중요시되고 있다. 이런 측면에서 볼 때 정보 검색 및 문서정보 저장소 시스템은 향후 다양한 정보원으로부터 고부가가치화된 정보처리 기능을 추가한 지식관리 시스템으로의 확장을 고려한 경쟁력을 가지고 있어야 한다.

TheoCMS가 대상으로 하고 있는 것은 전체적인 하나의 단일체로서의 문서나 정보가 아니라, 그 정보를 구성하고 있는 개 개의 콘텐츠이다. 따라서 CMS를 구현하기 위한 기본적인 정보 표현 도구로 사용될 포맷은 다음과 같은 요건을 만족할 수 있어야 한다.

첫째, 정보의 논리적 구조를 표현할 수 있어야 한다. CMS가 관리할 정보는 해당 정보를 구성하고 있는 콘텐츠 및 이들간의 논리적 관련성, 또 정보와 정보간의 연관성 등이 기본이다. 사용할 기본 포맷은 이들을 효과적으로 표현할 수 있어야 하고, 또 이들 논리적 정보를 특정 출판 유형과는 별개로 추상적으로 표현할 수 있는 기능을 제공하여 관리되는 정보가 다양한 활용 유형에 효과적으로 적용될 수 있도록 해야 한다.

둘째, 충분한 정보 표현력을 가져야 한다. CMS가 대상으로 하는 정보는 문자, 이미지, 음성 등의 다양한 미디어로 구성되어 있는 복합적인 것이며, 기본 포맷에서는 이러한 다양한 물리적 미디어간의 관련성을 적절히 표현할 수 있는 수단을 제공하여야 한다.

셋째, 단일 플랫폼에 종속되지 않아야 한다. CMS 가 관리하는 정보는 다양한 활용에 적을 가능해야 하며, 이를 위해서는 특정 플랫폼이나 어플리케이션에 독립적으로 정보를 표현하고 관리할 수 있어야 한다. PDF(Portable Document Format), 워드 포맷 등은 해당 어플리케이션(PDF 뷰어, 워드)이 없으면 내용을 볼 수가 없을 뿐 아니라, 특정 회사의 특정 어플리케이션에 종속되어 있으므로 CMS 의 기본 포맷으로 적절한 것이 될 수 없다.

상기의 요구 조건을 만족시킬 수 있는 가장 현실적인 선택은 XML/SGML 이다. 이들은 특정 플랫폼이나 어플리케이션에 종속되지 않고 정보를 표현할 수 있는 포맷으로서, 충분한 신뢰성을 확보하고 있는 국제적 표준 기관에 의해 제정, 유지되고 있으므로 특정 소프트웨어 벤더나 프로그램에 종속될 우려가 없이 정보를 생성하고 관리할 수 있다. 또한 XML/SGML 에서의 문서의 기본적인 구성은 엔티티라는 개념을 기반으로 하고 있어 다양한 미디어들의 표현 및 이들간의 논리적 연관성을 적절히 표현할 수 있고, 요청되는 다양한 출판 유형에 유연하게 대처할 수 있는 능력을 제공한다. 또한 이들 포맷을 기반으로 구현에 필요한 도구들도 다양하게 구비되어 있으므로 실제적인 구현에 있어서도 상당한 이점을 확보할 수 있다. 따라서 CMS 를 구현하기 위한 기본 정보 포맷으로는 XML/SGML 이 가장 이상적이라 할 수 있다.

TheoCMS 는 (주)K4M 에서 자체기술로 개발한 정보관리시스템으로 구조화정보 관리시스템이 가져야 할 기능적 요구사항을 국제표준인 XML/SGML 에 기반해 구현하였으며, XML/SGML 기반 정보인프라 구성의 필수 요소를 제공하며, 현재 하부 저장 구조로 객체관계형 DBMS 인 ORACLE 과 객체지향형 DBMS 인 O2 를 이용하는 두 가지 버전으로 제공되며 사용자의 필요에 따라 선택적으로 이용가능하도록 구현되어 있다. 또한 구조 정보 검색을 위하여 본사의 구조 정보 검색 솔루션인 TheoSEARCH 와 밀결합하여 수행될 수 있도록 구현되었다. 본 장은 구조화정보 관리시스템을 구현하기 위한 방법론과 더불어 TheoCMS 의 기능적 특성을 고찰하고자 한다. 본 고에서의 TheoCMS 는 정보검색 시스템과 정보저장소 기능이 융합되어 있는 시스템으로 정보구조화를 통한 정보의 지식관리를 지원한다.

*Theo 는 ‘신의 선물’을 의미하는 그리스어이며, TheoCMS 는 ‘신의 선물’과도 같은 이상적인 정보인프라를 완성하고자 하는 (주)K4M 의 목표를 구체화한 K4M 의 정보관리 제품명이다.

구조화정보 관리시스템의 구현 요건

XML/SGML 을 기반으로 한 구조화정보 관리시스템의 구현에 있어 고려하여야 할 요건은 다음과 같이 생각해 볼 수 있다.

▪ 정보 모델링

XML/SGML 기반 정보 관리시스템에서 정보는 기본적으로 XML/SGML 문서로 표현된다. 관리시스템 외부에서 생성된 XML/SGML 문서는 관리시스템에 저장될 때, 그 논리적인 구성 요소별로 분할되어 저장되고, 필요시 다시 결합되어 원래의 문서 또는 문서의 일부분이 복원되게 된다. 이 과정에서 원래의 문서내에 있던 정보를 완벽하게 복원하기 위해서는 XML/SGML 문서의 여러 특성들을 저장시스템이 인식하고 이에 대해 지원할 수 있어야 하고, 또한 관련되는 멀티미디어 콘텐츠도 원본 문서와의 관련성하에서 보존되어야 한다. 이를위해 문서관리시스템에서는 XML/SGML 문서의 특성에 따라 정보의 모델링 및 이에 기반한 구현이 이루어져야 하고, 이와 더불어 사용의 용이성을 위해 기존 파일시스템과 유사한 메타포어를 제공하는 가상 디렉토리 구조가 구현되어 정보의 조직화가 용이하도록 지원되어야 한다.

▪ 라이브러리 서비스

문서의 전 생명주기에 있어서, 생성된 문서를 웹브라우저나 특정 프로그램을 통해 관리시스템에 편리하게 저장할 수 있는 API, 사용자 권한 설정에 의한 문서, 문서 일부분 및 멀티미디어 콘텐츠에 대한 접근 제어, 문서의 변경에 따른 버전 등의 형상관리 등, 다양한 라이브러리 서비스가 구현되고 이를 용이하게 이용할 수 있는 API 가 제공되어야 한다.

▪ 공동 저작

다양한 콘텐츠를 구성 요소로 하여 이루어지는 문서는 그 생성에 있어서도 단일 저자에 의한 생성보다는 다수 저자가 동시에 공동 저작을 통해 생성할 필요성이 있다. 이러한 공동 저작을 위해 어떤 저자가 편집중인 문서나 문서의 부분을 다른 저자가 편집할 수 없도록 하는 Check-out/Check-in 기능을 비롯한 공동 저작 환경에서

의 지원이 구현되어야 한다. 더불어 사용자에게 대한 접근제어도 사용자별 뿐만이 아니라 사용자들이 모인 그룹별로도 제어가 가능하여야 할 필요성이 있다.

▪ 동적 문서 관리

CMS 에서 관리하는 정보는 한번 생성되면 그 내용 그대로 지속되는 죽어 있는 정보가 아니라, 그 자신의 생명 주기에 따라 수시로 또는 주기적으로 변경되고 진화되어 가는 생명체와 같은 존재이다. 따라서 CMS 에서 관리하는 정보의 변경 및 삭제, 추가 등이 항상 용이하게 이루어질 수 있는 시스템 구현이 필요하고, 이에 따라 문서에 대한 색인 정보 및 관리 정보 등도 동적으로 변경될 수 있도록 구현되어야 한다.

▪ 구조 정보 검색

CMS 가 관리하고 있는 정보의 검색에 있어서도, 기존의 전문검색엔진에 의한 검색만으로는 부적합하고, 정보의 내용 및 각 콘텐츠간의 연계성을 최대한 이용하여 검색할 수 있도록 구현되어야 한다. 또 대량의 정보에 대한 효율적이고 빠른 검색을 위해서는 기존 전문검색엔진의 색인 구조에다 필드 개념에 의한 구조 정보 색인을 덧씌운 형태의 통합되지 않은 색인 방식은 부적절하며, 색인 구조의 하부에서부터 다양한 정보들이 통합되어 있는 방식의 색인 알고리즘이 구현되어야 한다.

▪ 문서 변환/출판

CMS 의 궁극적인 목적은 관리하고 있는 정보를 다양한 단위의 형식으로 서비스 하는 것이다. 이를 위해 웹(CGI), ODBC 및 JDBC 등의 표준적 데이터베이스 접근 방식, API 등 다양한 서비스 접근 통로가 구현되어야 하고, 제공되는 정보의 형태도 XML/SGML/HTML 등의 표준적 문서 포맷, RTF, LaTeX, HWP 등의 일반적인 문서 포맷 등 다양한 포맷의 정보로 제공될 수 있도록 구현되어야 한다.

TheoCMS : 시스템 구성도

TheoCMS 는 전술한 바와 같이 하부 저장 구조로 객체관계형 DBMS 인 ORACLE 을 사용한 버전과 객체지향형 DBMS 인 O2 를 사용한 버전이 있다. 각 버전은 하부 플랫폼에 최적화된 상태에서 수행될 수 있도록 구현되었으며, ORACLE 버전의 경우 Stored Procedure 형태로 구현하였고, O2 버전의 경우 ODMG C++ 인터페이스를 사용하여 구현하여, 사용자가 필요에 따라 최적화된 버전을 선택할 수 있도록 하였다. 또한 제공되는 API 의 경우 양버전 모두 최대한 동일하도록 구현하여 본 제품을 이용한 통합 어플리케이션의 개발을 최대한 신속하고 효율적으로 가능하도록 하였다.

다음 그림은 본 TheoCMS 의 구성도이다.

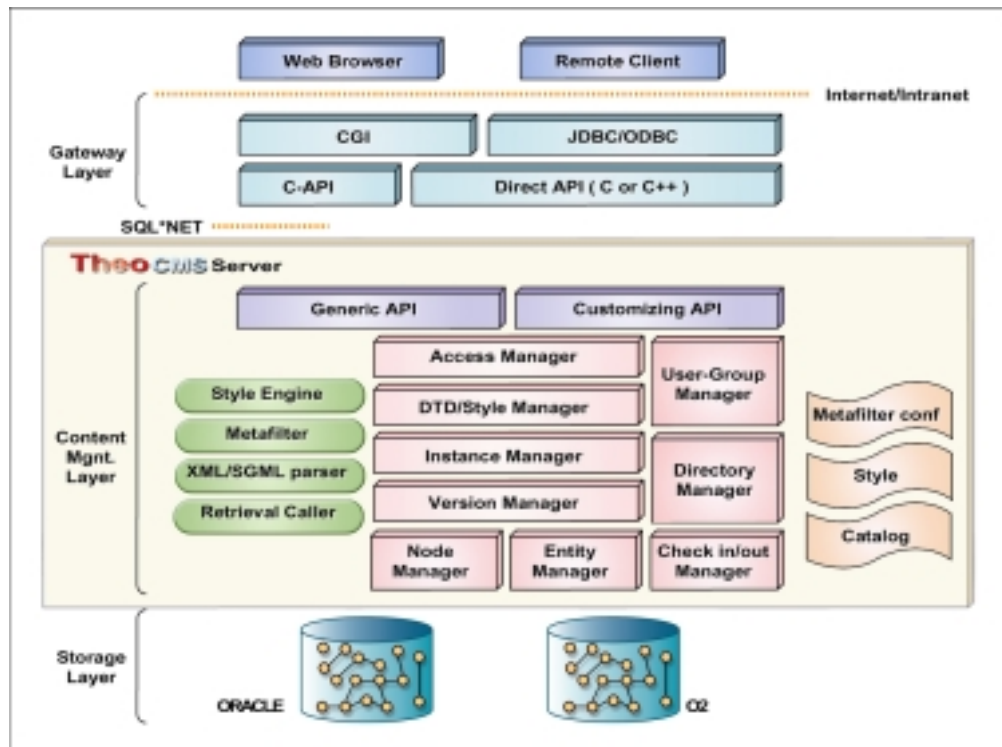


그림 6. TheoCMS 구성도

Storage Layer : 실제 XML/SGML 문서를 저장하는 레이어로서, 가장 높은 시장 점유율 및 저변을 확보하고 있는 ORACLE 를 사용하여 구현한 버전과, 점차 저변을 확대해 가고 있는 객체지향형 DBMS 인 O2 를 사용하여 구현한 버전이 있다. 특히 상위 레이어인 CM 레이어와는 명확히 구분된 인터페이스를

정의하여 구현함으로써, 사용자 요구에 따라 여타의 DBMS 또는 저장 수단으로 용이하게 확장할 수 있도록 구현하였다.

Content Management Layer : 콘텐츠 관리에 필요한 로직을 구현한 레이어로, 정보의 분할과 저장, 복원, 접근제어, 버전제어, Check-out/Check-in, 정보 변환 모듈 등으로 구성되어 있다. 또한 완벽한 한글 처리가 가능한 자체 개발 XML/SGML 파서 및 XSL/DSSSL 등의 스타일 처리가 가능한 스타일 엔진을 탑재하고 있으며, 기본적인 기능을 외부에 제공하는 Generic API 를 제공하고, 특정 사이트나 사용자의 필요에 따라 커스터마이징된 API 를 제공할 수 있도록 구현하였다.

Gateway Layer : CMS 에서 제공되는 API 를 이용하여 다양한 통로를 통해 CMS 가 관리하고 있는 정보를 서비스하는 레이어이다. 본 제품에서는 기본적으로 통합 어플리케이션의 개발을 위한 서버 API 와 클라이언트 API(C 와 C++), CM 레이어의 API 를 이용하여 웹을 통한 CGI, ODCB 및 JDBC 를 이용한 정보 제공 API 등을 제공하고 있고, ORACLE 버전의 경우 ORACLE*NET 을 통한 API 를 제공하여 더욱 편리한 개발이 가능하도록 하고 있다.

이와 더불어 본사의 구조화정보 검색시스템인 TheoSEARCH 와 밀결합이 가능하도록 API 를 제공하고 있으며, MetaFilter 개념을 지원하여 필요시 시스템의 수행 성능을 한층 높일 수 있도록 하고 있다. MetaFilter 란 정보의 분할 정도를 사용자가 정의할 수 있도록 하여 정보의 저장, 검색 및 복원에서 시스템의 수행 성능을 더욱 높일 수 있도록 하는 것이며, 사용자가 정의한 정보는 시스템의 외부에 존재하므로 편리한 시스템 configuration 이 가능하도록 하였다. 또한 기본 클라이언트로서 웹 브라우저에 의한 서버 브라우징, Java 로 구현된 서버 브라우저 및 C++로 구현된 서버 브라우저를 제공하여 별도의 클라이언트 개발 부담이 없이도 CMS 의 채용이 가능하도록 하고 있다.

본 제품이 제공하고 있는 다양한 기능을 사용하여 ‘One Source, Multiple Display’ 라는 CMS 고유의 목적을 완벽히 달성할 수 있는 솔루션이 제공될 수 있어, 본 제품은 SGMl 을 기본 문서 표준으로 채택하는 CALS 및 CITIS 솔루션을 개발하기 위한 인프라로서 활용이 가능하고, 일반 기업체에서의 문서 관리, 나아가 지식관리시

시스템의 개발을 위한 인프라로서 활용될 수 있다.

TheoCMS : 기능 목록

TheoCMS 는 3 단계의 레이어와 기타의 부가모듈을 통해 정보구조화에 가장 적합한 정보인프라를 구축할 수 있는 다양한 기능을 제공한다. TheoCMS 의 기능은 다음과 같다.

1. 논리적 분할 저장 및 복원 기능

XML/SGML 의 각 엘리먼트 단위에 의한 분할 저장 및 분할된 단위에 의한 독립적 관리, 저장된 문서의 전체 또는 일부 복원 기능을 기본적으로 제공하며, 각각의 분할된 엘리먼트들간의 논리적 연관성 및 관련된 멀티미디어 엔티티와의 연관성도 함께 저장되어 복원시 이용된다. 분할의 단위는 사용자가 필요에 따라 정의할 수 있도록 지원하고 있다.

2. 사용자 정의 가상 디렉토리

관리되는 각 정보들을 조직화하기 위한 정의가능한 디렉토리 메타포어를 제공하고 있다.

3. 접근 제어 기능

분할된 단위별로 사용자 및 그룹을 지정하여 읽기, 쓰기 등의 접근 제어 기능을 제공한다.

4. 버전 제어 기능

분할된 각 단위별로 관리할 수 있는 버전 정보를 유지하고, 필요시 이를 삭제할 수도 있으며, 버전닝 정책도 사용자가 임의로 정의할 수 있도록 하고 있다.

5. Check-out/Check-in 기능

다수 저자에 의한 동시 공동 저작을 지원하기 위해, 어떤 저자가 편집중인 부분을 Locking 하고 다른 저자가 이를 편집할 수 없도록 함으로써 정보의 무결성이 보장될 수 있도록 지원하며, 현재 편집중인 정보에 대해 한눈에 파악할 수 있는 API 를 제공하여 공동저작 환경을 보다 용이하게 지원하고 있다.

6. 다양한 스타일 적용 기능

저장된 정보의 서비스를 위해 XSL/DSSSL 등을 통한 실시간 변환 및 포맷팅이 가능한 스타일 엔진을 내장함으로써 보다 다양한 형식의 정보서비스가 가능하다. 현재 지원가능한 문서 포맷에는 XML/SGML/HTML/RTF/LaTex 등이 있다.

7. 외부 엔티티 통합 관리 기능

XML/SGML 에서 외부 엔티티는 정보 공유를 가능케하는 기본적 메커니즘이다. 즉 동일한 멀티미디어 객체를 여러 문서에서 참조하고 있을 경우, 이러한 관련성을 파악하고 이를 관리하여 데이터의 중복성을 배제함과 동시에 무결성을 보장할 수 있도록 해야 한다. 이를 위해 본 제품에서는 개개의 문서에서 참조되는 멀티미디어 엔티티들을 통합하여 관리하고, URL 을 통한 통일된 naming scheme 를 제공하고 있다.

8. 구조화 정보 검색 엔진과의 통합

본사의 구조화 정보 검색엔진과의 밀결합을 통해 보다 다양한 형태의 정보 검색이 가능하도록 지원하고 있다.

9. 개발 도구

CGI, C 및 C++, JDBC 및 ODBC 등을 이용해 통합 어플리케이션을 용이하게 개발할 수 있는 API 를 제공하며, 이를 이용해 기존의 XML/SGML 편집기등과의 용이한 연계가 가능하다.

10. 서버 브라우저

JAVA 버전 및 C++ 버전의 전용 서버 브라우저를 제공하여, 별도의 클라이언트 모듈 개발이 없이도 손쉬운 채용이 가능하다.

11. DTD 및 스타일의 등록

새로이 개발된 DTD 및 스타일을 등록할 경우 이를 파싱하고, 관련된 외부 엔티티들을 자동으로 저장, 관리하여 주므로, 시스템 개발 이후 유지, 보수가 용이하다.

12. 관리 정보 제공

분할되어 관리되고 있는 각 단위별로 사용자 정의 가능한 형식의 관리정보를

유지하여, 필요시 시스템 관리자에게 이들 정보를 제공하여 준다.

TheoCMS : 사양

작동 OS : UNIX 기반의 OS 시스템

연동 DBMS : ORACLE8, O2

작성언어: C 와 C++

연동 검색엔진 : TheoSEARCH

XML/SGML 지원 spec : W3C recommendation XML 과 ISO-8879 SGML 완벽지원

Further Information

Web Sites : <http://www.k4m.com>

E-mail : info@k4m.com

Telephone : (042) 864-1984

Address : (우) 305-333

대전시 유성구 어은동 1 번지 ETRI 창업지원센터 7114 호 (주)K4M