

SGML과 XML의 개요

미국 GNP의 20%가 새로운 정보를 만드는데 사용되는데 그 정보의 90% 이상이 문서로 작성되며 매년 그 수는 약 920억 건이 된다. 더군다나 WWW가 전세계적인 가상 데이터 저장소가 되었으며 매우 빠른 속도로 확장되고 있다. 전세계적인 것은 배제하고 과연 사람들이 그들 회사에서 작성한 정보를 쉽게 접할 수는 있는 것인가?

또한 전통적인 워드작업이나 데스크 탑 출판 시스템에서는 작업자가 그 작업시간의 30% 까지를 정보 검색에 사용하고 또 다른 30%를 단지 인쇄한 페이지가 '종게' 보이게 하기 위하여 스타일을 적용하거나 문단을 줄이는데 사용하는 것으로 알려져 있다. 더군다나 매 18개월 정도마다 소프트웨어가 완전히 바뀌어지게 되는데 회사는 새로운 기술로 나아가기 위한 데이터 변환에 끊임없이 돈을 투자하게 된다.

어떤 집단의 미래는 그 집단의 정보를 얼마나 효과적으로 찾아내고 관리하고 사용하는가에 달려 있다고 할 수 있다. 정보관리에 있어서 최근의 경향은 전자적으로 정보를 생성하고, 분배하고 관리하는 것이다. 이러한 접근을 시도한 집단은 그들이 생성하고 사용한 정보의 비용, 정확성, 적절성, 그리고 다양성에 있어서 엄청나게 향상되었다는 것을 실감하게 된다.

이러한 움직임의 일부로써 어떤 산업계의 회사는 회사 상호간이나 수요자와의 정보교환에 대한 표준을 함께 개발하는데 동참하고 있다. 이러한 표준에 맞추는 회사는 세계 시장에서 보다 효율적으로 사업을 하거나 보다 효과적으로 경쟁할 수 있을 것이다.

◎ 회사 정보의 재생 가능한 자원으로의 변화

전통적인 문서와 그것을 다루는 방법에는 많은 제한이 있다. 인쇄된 문서는 공들인 정보 작업의 - 주로 상당한 양의 수작업에 의한 - 산물인 경우가 많다. 어떤 문서는 특정한 목적을 위하여 특정한 시간에 특정한 수신인을 위해 준비된 정보를 순간적으로 나타

내는 것이다. 일단 문서가 인쇄되면 그 내용의 출처인 전자 데이터베이스와의 모든 연결이 사라지게 된다. 정보관리에 대해 체계적으로 접근할 때는 텍스트와 그래픽 데이터를 회사의 전자정보베이스의 일부로 다루게 된다. 이렇게 함으로써 모든 사람들이 정보에 접근할 수 있다. 정보의 생성과 전달과정에 대해 폭넓게 봄으로써 사용자는 문서가 모든 종류의 정보, 즉, 데이터베이스의 쿼리, 인쇄된 서류, 온라인 진단 안내서, 부품의 설명 카타로그, 비디오 장면의 모음, 또는 WWW(인터넷)상의 홈페이지 등을 포함하는 것으로 볼 수 있게 된다.

사용자로 하여금 정보생성과 전달과정에 대해 폭넓게 볼 수 있게 한 표준은 SGML(Standard Generalized Mark-up Language)이다. 이것은 1986년에 공표된 세계 표준(ISO 8879)이다.

◎ SGML은 무엇인가?

SGML은 정보를 종이 위의 글자가 아니라 데이터 객체로 관리할 수 있게 한다. 데이터는 구별하기 어려운 비트와 바이트의 흐름이라기 보다는 정보의 구별된 요소로 그룹지어 진다. 이렇게 함으로써 정보를 효율적으로 저장하거나 재사용하고 다른 많은 사용자와 정보를 공유하고 그것을 데이터베이스 안에서 관리하는 것을 가능하게 한다. SGML은 문서 안에 기술적 마크업(descriptive mark-up)을 새겨 넣는 표준 포맷을 규정한다. SGML의 실제 가치와 능력에 중요하고 결정적인 것은 그것이 문서의 구조를 묘사하는 표준화된 방법을 상술한다는 것이다.

다른 말로 하면 SGML을 사용함으로써 우리가 만드는 문서 각각의 종류에 대한 체계적인 모델을 세우게 된다는 것이다. SGML은 장, 제목, 단락과 같은 기술적 마크업을 붙인 구조의 각 요소가 문서 내의 논리적이고 예측가능한 구조에 맞추게 한다. SGML은 매우 다양한 문서의 구조를 지원한다. 사용자는 주로 생성하려는 정보의 종류별로 정보게시판,

기술안내서, 부품 카탈로그, 설계사양서, 보고서, 편지와 메모 등 별개의 문서구조를 만들게 된다.

SGML을 사용하면 특정한 하드웨어나 소프트웨어와는 무관한 문서를 만들 수 있다. SGML 문서가 국제적 표준에 따르는 것이기 때문에 이기종사이에 이식이 가능하다. 따라서 다른 시스템을 사용하는 사용자와 문서를 교환할 수 있다.

◎ 마크업이란 무엇인가?

마크업은 문서의 내용이 아니라면 문서에 포함된 모든 것을 말한다. 마크업은 원래 설계자가 타이프로 작성된 텍스트에 첨가된 손으로 쓴 표시를 일컫는 것으로 이 표시는 타이피스트에게 자료를 작성하는 방법이나 어떤 타입페이스를 사용하는가와 같은 지식을 담게 된다. 이러한 종류의 마크업을 절차적 마크업(procedural mark-up)이라고 한다.

절차적 마크업(procedural mark-up) : 오늘날 워드 소프트웨어나 데스크탑 출판 소프트웨어 같은 대부분의 전자 출판 시스템은 절차적 마크업을 사용한다. 절차적 마크업은 주로 마이크로소프트 워드와 같은 특정한 소프트웨어마다 고유한 것이다. 각각의 소프트웨어는 그 자신에만 사용될 수 있는 고유한 마크업 코드가 있다. 이러한 마크업 코드는 문서의 텍스트 내에 섞여있는 포매팅 코드의 형태를 갖는다. 절차적 마크업 코드는 인쇄된 페이지와 같이 정보를 표현하는 한 가지 방법에만 적용되며 CD-ROM이나 인터넷과 같은 다른 미디어에 대한 정의를 내리지는 못한다.

기술적 마크업(descriptive mark-up) : 보통 일반적 마크업(generic mark-up)으로 알려진 기술적 마크업은 텍스트의 목적을 페이지의 물리적인 외양이 아니라 문서 내에 표현한다. 기술적 마크업의 기본 개념은 문서의 내용과 문서의 스타일은 구별되어야 한다는 것이다. 기술적 마크업은 문서의 구조에 기초하고 있으며 장, 단락, 또는 목차와 같은 구조내의 요소에 대해 요소가 어떻게 보이느냐 아니고 요소가 무엇인가를 표현하는 기호를 사용하여 구별한다. 기술적 마크업은 표현 정보(즉 스타일)를 구조로부터 분리함으로써 동일한 정보를 여러 가지 방식으로 표현하는 것을 가능하게 했다. 예를 들어서 하나의 소스파일로부터 종이, 온라인, CD-ROM, 인터넷에 출판하는 것이 가능하다.

◎ SGML의 작동방식

일반적으로 문서의 정의는 구조, 내용, 그리고 스타일의 세 요소로 나뉘어진다. SGML은 이 세 요소를 구분하지만 주로 구조와 내용간의 관계를 다룬다.

● 구조(Structure)

SGML 어플리케이션의 핵심은 DTD(Document Type Definition)이라 불리는 파일이다. DTD는 마치 데이터베이스의 스키마가 취급하게된 정보의 형식과 필드사이의 관계를 나타내는 것처럼 문서의 구조를 나타내게 된다. DTD는 문서를 구성하는 요소의(장이나 장의 표제, 단락, 그리고 제목 등) 틀을 제공한다.

DTD는 또한 요소사이의 관계에 대한 법칙을 정해 주는데 예를 들어서 “장의 표제는 장을 시작한 후의 첫 번째 요소가 되어야 한다” 또는 “각 리스트는 적어도 두 개의 항목을 포함해야 된다” 같은 것들이다. DTD가 정의하는 이와 같은 법칙은 문서가 일관되고 논리적인 구조를 갖게끔 도와준다. DTD는 문서에 항상 수반된다. “문서 인스턴스”는 특정한 DTD에 따라서 내용이 태그(tag)된 문서를 말한다.

● 내용(Content)

내용은 정보 그 자체이다. 내용은 제목, 단락, 목록, 표, 그림, 그리고 음성 등을 포함하는 것이다. DTD의 구조내에서 내용의 위치를 확인하는 방법을 태깅(tagging)이라 부른다. SGML 문서를 만드는 것은 내용 주위에 태그를 삽입하는 작업을 포함하는 것이다. 이러한 태그는 구조의 각 부분의 시작과 끝을 표시하게 된다. 다음의 예에서 “<par>”는 단락의 시작을 나타내며 “</par>”는 끝을 나타낸다.

<par> Content

is the information itself. </par>

You can nest elements within other elements; in the following example, the paragraph (“<par>”) is an element within the topic (“<topic>”) :

<topic><par> Content

is the information itself. </par></topic>

The structure of a particular document is revealed by the nesting of tags :

<section> <subhead> Content </subhead > <par >
Content is the information
itself. </par > </section >

다행스럽게도 작업자가 수작업으로 태그를 집어넣는
다든지 모든 태그가 올바르게 사용되었는지를 확인할
필요는 없다. 어떤 SGML 기반 저술 소프트웨어 프
로그램은 문서에서 커서의 현재 위치에서 유효한 태
그만을 보여주는 풀다운 메뉴에서 태그를 선택함으로
써 태그를 입력할 수 있도록 해준다. 이러한 프로그램
은 문서가 DTD의 법칙을 따르는지를 검사하는 파서
(parser)라 불리는 소프트웨어 모듈에 의존하게 된다.

●스타일(Style)

SGML 자체는 스타일에 대한 표준을 세우는 것과
는 아무런 관계가 없기 때문에 대부분의 시스템은
아직도 고유의 방법에 의존하고 있다. 표준에 기초한
스타일 시트를 개발하려는 노력의 결과로 mature
OS와 최근에 발표된 DSSSL이 만들어졌다.

미국방성의 CALS에서는 Output Specification(OS)
으로 알려진 독자적인 표준을 개발했다. 이 OS는 특
수한 DTD의 형태이며 사용자로 하여금 Formatting
Output Specification Instance(FOSI)를 만들 수 있
게 하는데 이것은 인쇄된 출력물이나 전자 출력물에
모두 적합하다.

FOSI는 DTD내의 각 태그를 포맷하는 방식을 결
정하는 강력한 스타일 시트이다. FOSI(스타일), 문서
(내용), DTD(구조)를 가지면 인쇄된 문서에 대한 완
벽한 교환 패키지를 갖게 된다.

1996년에 ISO는 SGML에 기초한 문서에 대한
Document Style Semantics and Specification Lan-
guage(DSSSL)의 최종적인 초안을 승인했다. 완성된
DSSSL 표준은 다양한 영역을 망라하여 그 부분집합
(subset)이 다양한 기능을 다룰 수 있도록 개발되고
있다. FOSI와 거의 유사한 기능을 갖는 부분집합이
있을 것이며 FOSI와 DSSSL 사이에서 변환이 가능
한 도구를 만드는 작업이 진행중이다.

현재 많은 수의 군사계약에서 FOSI가 요구되며
많은 방위산업이 아닌 회사에서도 또한 OS를 기꺼이
받아들이는데 그것은 OS가 완속하며 지원을 받는 표

준이기 때문이다. DSSSL과 FOSI가 가까운 미래에
도 중요한 표준으로 남게 될 것이다.

◎ XML 이란?

WWW의 놀라운 성장은 저술가로 하여금 전자 문서
를 세계의 청중에게 쉽고도 값싸게 전해줄 수 있는 인
터넷의 능력에 의하여 촉진된 것이다. 웹의 문서가 커
지고 복잡해짐에 따라 웹 제공자들은 대형 상업 출판
에 필요한 확장성과 구조, 그리고 데이터 검사를 제공
하지 않는 미디어들의 한계를 경험하게 되었다.

상업적인 웹 출판의 필요성을 역설하고 또 웹 기
술을 분산된 문서처리의 새로운 영역으로 확장시킬
수 있도록 WWW 컨소시엄(W3C)은 SGML의 이점
을 웹 상에 사용하기 쉽고 수월하게 해주는 사양
(specification)을 만들기 위해 SGML Working Group
을 만들게 되었다. W3C SGML 활동의 목표는 자
기표현이 가능한 데이터 구조를 요구하는 어플리케이션
에 어느 정도의 깊이와 복잡성을 제공하는 것이 가능
하게 하는 것이다.

이러한 노력의 초기 단계는 웹 어플리케이션에 맞
춰 특별히 고안된 SGML의 단순화된 부분집합의 사
양인 것이다. XML(Extensible Mark-up Language)
이라 불리는 이 부분집합은 SGML의 주요한 이점인
확장성, 구조, 그리고 검증을 유지하면서 전체의
SGML보다 배우기 쉽고 적용하기 쉽게 설계된 언어
인 것이다.

- 위의 내용에 대한 보다 상세한 정보는 Kais Al-Timimi에게서 구할 수 있다.

《Engineering Data Newsletter, Volume 6,
Number 9, November 1998》

본 기사는 경희대학교 김영진 편집위원이 “Engineer-
ing Data Newsletter”에서 발췌하였으며 출판사인
Datamation Ltd.의 연락처는 다음과 같다.

- Tel : +44-1223-572-579
- Fax : +44-1223-571-950
- E-mail : kais@datamation.co.uk