

SGML/XML 기반 문서 정보 시스템

주종철, 김현기, 박영찬

한국전자통신연구원, 컴퓨터·소프트웨어 연구소

자연어처리 연구부 문서정보 연구팀

요약문

SGML/XML 기반 문서 정보 시스템의 문서 포맷은 인간이 비교적 쉽게 이해할 수 있으며 기존의 문서 정보 시스템에서는 접근 불가능한 레벨까지 인간이 접근할 수 있는 통로를 제공한다. 또한, 문서처리 관점에서 보면 기존의 각 CAL/EC 관련 응용 프로그램에서 생성된 데이터는 문서 또는 파일 단위의 접근으로 업무레벨에 도움을 주는 것이 제한적인 것에 반해 SGML/XML 기반 문서 정보 시스템은 DTD(Document Type Definition)에서 설계된 엘리먼트, 속성, 엔티티 등의 단위까지 접근할 수 있는 Mechanism으로 기존의 문서 정보 시스템과 차별화 되는 지식 매체로서 지식기반 시스템의 하부 프레임워크를 제공한다. SGML/XML 기반 문서 정보 시스템의 요소로서 SGML/XML 문서 정보 검색 시스템은 표현된 계층적 논리정보를 이용하여 다양한 문서 접근점을 제공할 수 있으며, 사용자 요구에 적합한 문서의 재사용 및 동적인 문서제시를 가능하게 한다. 또한, SGML/XML 문서 정보 관리 시스템은 공동 저작을 위한 세부 논리적 단위별 check-in/check-out, 액세스 제어, 버전닝 기능 등을 제공하여 기업 내 자산 관리를 위한 새로운 패러다임을 제공한다. 본 논문에서는 구현 사례와 더불어 SGML/XML 기술이 CALS/EC 추진에 주는 영향과 향후 바람직한 CALS/EC 또는 가상기업(Virtual Enterprise)구축을 위한 기술적 방향에 대해 논한다.

I. 서론

최근 CALS/EC, 전자 도서관, 전자 정부, 기술 문서 관리 등의 다양한 응용 분야에서 ISO 국제 표준 (ISO 8879:1986)인 SGML(Standard Generalized Markup Language)[1]과 W3C 추천안인 XML(eXtensible Markup Language)[2]을 적용한 전자 문서화가 확산되고 있다. 특히, CALS/EC 분야에서 SGML/XML 관련 기술이 다양한 정보 소스를 지니는 정보 저장소간의 오브젝트 정보 공유 및 교환을 위한 향후 가상 기업을 구현하는데 중요한 역할을 할 것으로 기대되고 있다. SGML/XML 기술이 네트워크상의 기업들이 무한 경쟁 환경에서 갖는 주요한 함축적인 의미는 사람과 기계간의 연계를 촉진하는 점과 벤더 독립적 소프트웨어 개발 및 데이터 생성을 가능하게 하여 정보의 공유 및 재사용을 촉진한다는 점이다. 가상 기업 환경에서 SGML/XML 기술이 광범위한 응용분야에 활용성이 높은 구체적인 이유는 다음과 같다.

- 정보 생성에 드는 비용과 시간 절약 : 기존 워드프로세서나 DTP(Desk Top Publishing) 시스

템을 이용한 정보의 생성 시 저자는 30%의 시간을 정보를 찾는 데 소비하며, 30%는 글자의 폰트 종류, 크기 및 페이지 레이아웃 등 보기 좋은 문서를 출력하는 작업에 소비한다. 또한 매 18개월마다 소프트웨어 기술이 변하므로 기존에 생성한 데이터 변환에 비용을 투자해야 한다[3]. SGML/XML은 다수의 응용 프로그램에 의해 생성된 다양한 유형의 멀티미디어 데이터 엔티티 단위로 구성된 가상 문서 개념을 지원하기 때문에 기존에 생성된 정보의 단위들을 조합하여 새로운 정보를 쉽게 생성할 수 있다. 즉, SGML은 정보를 효율적으로 처리하기 위해서 정보를 비트나 바이트의 순차적 스트림으로 처리하지 않고 “하나의 단위로 참조될 수 있는 문자들의 집합 또는 멀티미디어 데이터들의 집합”인 엔티티(entity)를 정의한다[1]. 이와 같은 엔티티 기능은 동일한 정보를 다양한 목적에 따라 효율적으로 재구성, 재사용할 수 있게 하며, 데이터베이스 시스템을 저장소로 이용하여 사용자 요구에 의해 문서 인스턴스를 동적으로 생성할 수 있게 한다. 또한, SGML/XML 데이터는 가독성이 높기 때문에 향후 마크 업을 매개로 하여 지식베이스와 연계된 정보 생성의 가능성을 제공한다 [4,5,6,7].

- **벤더 및 시스템에 독립적인 문서 처리** : 기존의 대부분의 제품들은 논리적 구조 정보와 물리적 외양정보가 혼재되어 생성되는 데이터가 특정 응용 제품 또는 벤더에게 종속됨으로써 정보의 재활용을 위해서는 특정 응용 프로그램간의 변환 프로그램에 의해서 가능하였다. 국제 표준적 방법인 SGML/XML에 의한 문서구조화는 문서의 논리적 구조정보와 물리적 외양정보를 분리된 파일들로 생성하기 때문에 정보 변환 시 논리적 구조 정보와 물리적 외양정보가 혼재되어 있는 경우보다 매우 간단해 지게 된다. 문서의 논리적인 마크 업과 포맷팅 정보를 분리시킴으로써 특정 문서 집합에 대한 표준화된 출력과 동일 문서에 대한 다수의 출력 포맷을 사용자 특성에 따라 다양하게 지정할 수 있다. 이는 문서의 내용을 작성하는 저작 과정에서 문서의 최종 출력을 위한 포맷 정보를 작성하는 비용을 줄이는 효과를 가져올 뿐 아니라, DSSSL(Document Style Semantics and Specification Language)과 XSL(eXtensible Style Language)의 다양한 스타일시트를 이용한 개인화 된 출판을 가능하게 한다. 따라서, 이질적인 네트워크 및 분산 환경하의 여러 종류의 정보 저장소 데이터 상호간 적은 비용으로 전자 상거래 및 EDI의 정보 기술 언어로 XML의 활용성이 부각되고 있는 추세이다.
- **정보 생성자가 예측하지 못하는 다양한 활용성** : SGML/XML 문서는 DTD에서 정의된 논리적인 구조에 따라 문서가 작성되어 문서 사이클 별 다양한 활용성을 갖는다. 첫째, 정보 제시 측면에서는 DSSSL의 문서 변환 스펙은 서로 다른 DTD를 지니는 문서간의 변환이나 문서 내 논리적 구성 요소들을 선택적으로 가상 문서들을 구성하여 새로운 형태의 정보 활용을 가능하도록 한다. 둘째, 정보 검색의 측면에서는 기존 정보 검색이 문서 단위의 검색인 반면 검색의 범위를 특정 엘리먼트, 엔티티, 속성 등으로 한정할 수 있는 가능성이 가능함으로써 검색 결과의 정확도를 향상시킬 수 있다. 셋째, 정보 관리 측면에서는 기존보다 논리적 세부 단위별 라이브러리 서비스 기능(check-in/check-out, versioning), 사용자 액세스 관리 등을 제공한다. 이와 같이 문서의 생명주기 전 과정에서 SGML/XML 기반 문서 정보 시스템은 정보의 생산성, 재사용성, 지속성, 이식성 등의 측면에서 장점들을 제공한다.

II. SGML/XML 기반 문서 정보 시스템

2.1 SGML/XML 기반 문서 정보 시스템의 고려 사항

SGML/XML 문서 정보 시스템 구현 시 고려해야 될 사항들은 다음과 같다.

- **전자 문서 모델링 기능:** SGML/XML 문서를 논리적 세부 컴포넌트 별로 나누어 데이터베이스에 저장하고 원래 문서로 복원하기 위한 모델링이 필요하다. SGML/XML 문서를 저장하기 위한 데이터 모델은 저장 시스템에 삽입하기 전의 문서와 저장된 문서를 다시 복원했을 때의 데이터 손실을 방지하기 위해 SGML/XML의 모든 특성들을 고려해야 한다. 또한 문서 내에 포함된 멀티미디어 정보의 저장과 엔티티 참조에 의해 공유되는 엔티티 선언의 저장을 고려해야 한다. 이외에도 사용자가 기존의 파일 시스템을 사용하는 것과 같은 메타포어를 제공하는 가상 디렉토리 구조를 지원해야 한다.
- **라이브러리 서비스 기능:** 문서의 생명 주기(life cycle) 제어 기능, 사용자 권한에 의한 액세스 제어 기능, 다수 저자의 동일 문서에 대한 공동 저작 시 발생하는 문서의 불일치를 방지하는 check-in/check-out 기능, 문서의 생명 주기 동안의 변경 사항을 관리하는 버전닝 기능 등을 제공해야 한다.
- **공동 저작 기능:** SGML 문서 관리 개념은 SGML 문서를 공동 편집하고자 하는 요구로부터 발생된다. SGML 저장 시스템은 그룹 사용자가 동일한 문서를 동시에 문서 전체 또는 일부분에 대한 편집, 생성, 갱신 등의 경우와 같은 공동저작 기능을 제공해야 한다. 그러므로 사용자는 “Document Fragment Server”를 필요로 하며 교환과 관리의 단위는 문서 Fragment 이다 [7].
- **동적 문서 생성/교환 기능:** 사용자의 질의 또는 문서 생성 템플릿에 의해 저장소에 저장된 컴포넌트들을 재구성하여 동적으로 문서를 생성하는 기능 및 문서의 컨텍스트를 유지할 수 있는 Fragment Interchange 기능을 제공해야 한다.
- **동적 색인 기능:** 문서 관리 시스템에 임포트되는 문서 및 기존에 저장된 문서의 수정/삭제 시 구축된 색인 정보를 동적으로 갱신하는 기능을 제공해야 한다.
- **구조 정보 검색 기능:** 저장소에 저장된 SGML/XML 문서 중 사용자에게 접근 권한이 허용된 문서들에 대한 구조, 속성, 키워드에 대한 검색 기능을 제공해야 한다. 사용자는 검색 범위를 문서의 특정 부분으로 한정시켜 질의를 함으로써 검색의 정확도를 향상시킬 수 있다[6, 11].
- **문서 변환/출판 기능:** 문서 관리 시스템에 저장된 문서의 다른 포맷으로의 변환 및 웹, CD-ROM, 프린터 및 XSL, DSSSL 스타일시트에 의한 다양한 출판 기능 등을 제공해야 한다.

2.2 SGML/XML 기반 문서 정보 시스템 접근 방법

지금까지 문서정보 전자화의 주류였던 워드프로세서와 DTP 관련 제품들은 논리적 구조 정보와 물리적 외양정보가 혼재되어 특정 응용 제품에 종속됨으로써 정보의 재활용을 위해서는 특정 용

용 프로그램간의 변환 프로그램에 의해서 가능하였다. SGML/XML 에 의한 문서구조화는 특정 응용 프로그램과 시스템에 종속되지 않는 방식으로 정보의 활용성을 염두에 두고 설계된 정보의 생성으로 보다 적고 간단한 변환 프로그램으로 정보의 교환이 가능하다. 그 근거로는 SGML/XML 출판 모델은 문서의 논리적 구조정보와 물리적 외양정보를 분리된 파일들로 생성하기 때문에 정보 변환 시 논리적 구조 정보와 물리적 외양정보가 혼재 되어 있는 경우보다 매우 간단해 지게 된다. SGML/XML 출판 모델에 따른 첫 번째 단계로 논리적 구조정보에 대한 문서구조화를 위해서는 Filtered SGML 과 Native SGML 의 두 가지 접근 방법이 있는데 이에 대한 충분한 인식과 기술적 검토를 통한 선택이 필요하다.

우선 교환에 초점을 둔 Filtered SGML 은 모든 국제 표준이 공통적으로 가지는 장점으로서 기존에 이미 생성된 문서들을 표준적 형태로 만드는 자체에 의의가 높다. 주로 간단한 문서 구조에 적용되며 태그 맵핑 및 변환 검증 등의 향후 운용 비용이 많이 들고 문서구조화가 지니는 모든 효과를 기대할 수 없는 점을 유념할 필요가 있다. 그러나, Filtered SGML 접근 방법은 DTD 생성에 따른 부담을 줄이고 기존 정보 시스템과의 접목을 높이는 효과가 크기 때문에 최근 XML 에서는 Well-formed 문서라는 새로운 개념을 만들어 사용자가 자유롭게 태그를 확장할 수 있게 하고 반 자동 또는 사용자의 타스크로 태그 맵핑 또는 변환을 수행한다. 즉 구조화된 문서 객체간의 연관성에 관한 정보 처리가 시스템 측에서 거의 없는 정보 생성 또는 변환 방법이며 이에 대한 처리는 사용자의 로드로 전이하는 방법이다. Native SGML 접근방법은 충분히 검토된 문서유형 DTD 를 설계한 후 SGML 전용 편집기를 사용하여 DTD 문법에 유효한 정보를 생성하여 SGML 이 추구하는 본질에 충실한 방법이다. 이는 문서구조화 기술의 모체가 되는 SGML 이 지니는 본질이 정보 교환을 훨씬 증가하는 기대효과가 보장되는 정보 시스템의 인프라가 만들어지도록 하는데 필요한 조건으로 볼 수 있다.

2.3 SGML/XML 기반 문서 정보 시스템 구현 사례

SGML/XML 기반 문서 정보 시스템의 구현 사례로서 구조 정보 검색 시스템과 문서 정보 관리 시스템에 대해 소개한다.

2.3.1 SGML/XML 기반 문서 정보 검색 시스템

SGML/XML 문서는 계층적 구조와 이에 따른 내용으로 구분되므로 다음과 같은 검색 측면에서의 장점을 갖을 수 있다.

- **문서접근 (document access point)의 다양성** : 기존의 단순 색인어 만으로 문서에 접근하는 기능 외에 문서의 계층적 구조정보와 내용정보를 혼합하여, 다양한 각도로 문서를 검색할 수 있다. 예를 들어 특정한 키워드가 제목에 포함되거나, 저자 소속, 요약 등의 문서의 특정 부분을 지칭하여 검색의 범위를 한정할 수 있으며, 6개의 장으로 구성된 문서, 세 명 이상의 저

자가 있는 문서 등 문서의 구조 정보를 이용한 검색이 가능하다.

- **동적인 문서 제시 (dynamic presentation) 가능** : SGML/XML 은 문서의 일부만을 골라내어 자유롭게 조합하여 사용자에게 제시할 수 있다. 즉 문서전체를 보낼 필요없이 문서의 요약, 저자부분, 참고문헌만을 독립적으로 분리하여 볼 수 있으며, 스타일정보를 외적으로 추가하여 여러 형태로 조합하여 보여줄 수 있다. 따라서 불필요한 네트워크의 과부하와 시스템의 과부하를 감소시킬 수 있다. 예를 들어 문서 내에 포함된 그림만을 조합한 새로운 형태로 제시가 가능하며 제시된 문서의 재사용 또한 보장되므로 Document Re-assembly 가 가능하다 .
- **검색정보의 일관된 관리** : 일반적으로 문서에서 특정 부분을 자동 인식 (예를 들어 제목부분) 하면 오인식이 항상 개입되게 된다. 따라서 제목이 아닌 엉뚱한 부분이 제목으로 인식되어 검색될 수 있다. 그러나 SGML/XML 은 구조에 대한 유효성(validation) 검증을 마친 상태이므로 항상 일관된 정보를 바탕으로 검색에 사용될 수 있다.
- **다양한 부가 정보 제공** : 문서의 여러 부분이 계층적으로 구분되어 있으므로 검색 외의 여러 부가 정보를 얻을 수 있다. 예를 들어 가중치 계산을 위한 용어 분포 정보, 문서 요약, 분류를 위한 용어 분포정보 등을 논리적 세부 단위 별로 부가적으로 얻어 활용 할 수 있다. 부가 정보를 이용한 예로 참고문헌에 있는 논문을 이용하여, 본 논문을 참고하고 있는 논문에 대한 브라우징 및 본 논문의 저자가 쓴 다른 논문 검색 등이 단계적으로 가능하다.

SGML/XML 문서 정보 검색 시스템을 구현하기 위한 방법은 SGML/XML 이 표현하는 구조정보를 검색에서 얼마만큼 손실 없이 제공하는지에 따라 다음과 같이 4 단계로 분류하여 구현된다.

- **Phase 1: 단순 검색 방법** - 이 방법은 SGML/XML 의 마크 업을 모두 무시하고 하나의 문서를 하나의 검색단위로 하여 기존의 검색 시스템에 그대로 적용하는 방법이다. 이러한 시스템은 기존의 검색시스템을 그대로 이용할 수 있으나 SGML/XML 문서가 갖는 모든 정보를 손실하게 되어 구조검색이 갖는 장점을 제공할 수 없다.
- **Phase 2 : 독립 필드 검색 방법** - 이 방법은 SGML/XML 문서를 SGML/XML 파서를 사용하지 않고 특정 부분을 n 개의 독립된 텍스트 문서로 분리하는 방법이다. SGML/XML 문서를 구절 검색을 이용하여 검색하는 시스템이라고 할 수 있다. 이러한 시스템은 기존의 검색 시스템을 약간 수정하여 이용 가능하다. 그러나 각각의 분리된 독립된 텍스트 문서들은 개개의 문서로 취급되어 부분/전체의 통합 기능이 없으며 문서 접근점 또한 구절 검색 수준으로 제약된다. 그리고 원래의 SGML/XML 의 정보는 손실되므로 동적인 문서제시가 불가능 하다.
- **Phase 3 : 통합 필드 검색 방법** - 이 방법은 SGML/XML 파서를 이용하여 문서를 n 개의 논리적 부분으로 나누며 각 분리된 부분은 전체로 통합되어 하나의 문서전체를 이룬다. 이 방법은 Phase 2 보다 많은 검색 접근점을 제공할 수 있으며 분리된 부분 단위로 동적인 제시가 가능하며 부분과 전체의 제한된 통합기능을 제공한다. 그러나 새로운 역화일 구조의 구성이 필요하여 기존의 검색시스템으로 구현 시 색인저장 구조의 낭비가 생긴다.
- **Phase 4 : 구조 엘리먼트 검색 방법** - 이 방법은 SGML/XML 의 파싱된 결과를 모두 분리하여 저장하며 서로의 연관관계를 모두 구분하는 방법이다. 즉 SGML/XML 파싱트리 자체를 색인

구조화 하여 사용한다. 이러한 시스템은 SGML/XML 이 주는 모든 장점을 사용할 수 있으나, 상당한 색인저장 공간이 요구되며 검색측면에서 빠른 속도를 얻기 위해 여러 기술적 어려움이 존재한다.

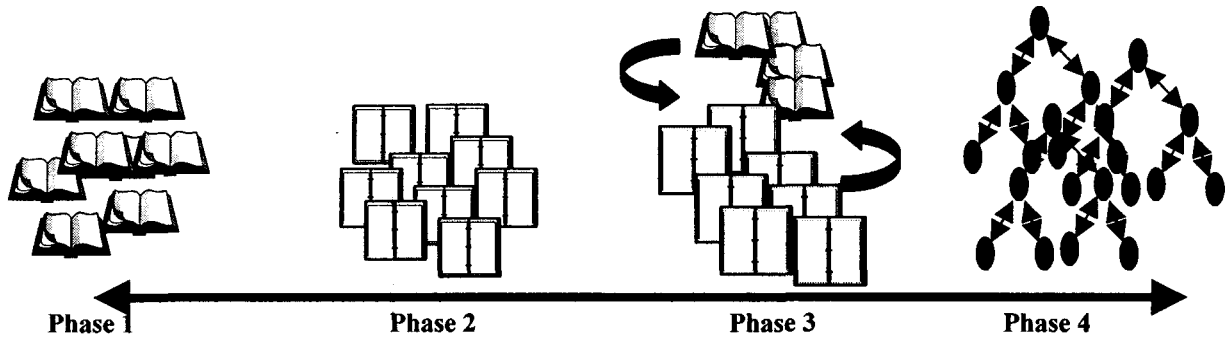


그림 1: SGML/XML 정보검색 시스템의 구현 단계

본 연구팀에서 개발된 STEER 시스템에서는 현재 Phase 3 수준의 SGML/XML 정보 검색이 가능하며 Phase 4 수준의 SGML/XML 정보 검색 시스템은 구현 중에 있다. Phase 3까지는 Filtered 접근방법에 의한 구현 방법이며 Phase 4는 Native SGML 접근 방법이다. 아래의 표는 Phase 3 수준의 SGML/XML 정보 검색 시스템 구현을 위하여 SGML 태그 명을 정보 검색 관점에서 좀더 의미 있는 태그로 변환하여 정보 검색 응용시스템을 구현하는 Meta Filter의 예를 보여 주고 있다.

FIELD	SGML_TAG	ATTR	IDX	TYPE	OCCR	OPTIONAL
&name	@FILENAME	NO	INDEX	TXT	SINGLE	MANDATORY
&title	제목	NO	INDEX	TXT	SINGLE	MANDATORY
&date	출판일	NO	INDEX	TXT	SINGLE	MANDATORY
&header	@HEADER	NO	INDEX	TXT	SINGLE	MANDATORY
&논문지정보	논문지정보	NO	INDEX	TXT	SINGLE	OPTIONAL
&영어제목	대체제목	NO	INDEX	TXT	SINGLE	OPTIONAL
&저자그룹	저자그룹	NO	INDEX	TXT	SINGLE	OPTIONAL
&한글요약	요약	한글	INDEX	TXT	SINGLE	OPTIONAL
&영문요약	요약	영어	INDEX	TXT	SINGLE	OPTIONAL
&chapter	장	NO	INDEX	TXT	MULTIPLE	OPTIONAL
&참고문헌	참고문헌	NO	INDEX	TXT	SINGLE	OPTIONAL
&이력	이력	NO	INDEX	TXT	SINGLE	OPTIONAL
&전문	@FULLTEXT	NO	INDEX	TXT	SINGLE	OPTIONAL
&그림	그림그룹	NO	INDEX	TXT	UNIFY	OPTIONAL

표 1: Filtered 접근 방법을 이용한 Meta Filter의 예

2.3.2 SGML/XML 기반 정보 관리 시스템

본 연구팀에서 구현한 SGML/XML 기반 문서 정보 관리 시스템은 임의 유형의 DTD 에 따라 작성된 SGML/XML 문서를 정보의 손실 없이 저장 가능한 Native SGML 접근방법을 수용하였다. 여기서의 정보 검색 기능은 객체지향 OODBMS 와 밀결합된 형태로서 DB 시스템과 Loosely Coupled 된 STEER 시스템과는 별도로 Phase 4 의 검색 기능을 관리 관점에서 제공한다. 하부 저장 시스템은 OODBMS 인 O2 DBMS 를 사용하였다. 또한, 논리적 세부 단위별 공동 저작을 위한 check-in/check-out 기능, 버전닝 기능, 색인 및 구조 정보 검색 기능 등을 지원한다. 문서 관리 시스템의 클라이언트 응용 프로그램으로는 저장소의 가상 디렉토리 구조를 브라우징하면서 문서 관리 작업을 할 수 있는 워크벤치인 RM Navigator 와 SGML/XML 문서의 전체 또는 일부분을 편집할 수 있는 KLEOX SGML 편집기가 있다. 일반적 SGML 파서의 출력인 ESIS(Element Structure Information Set)는 문서가 갖는 모든 정보를 포함하고 있지 않으므로 이를 이용한 문서 저장은 정보의 손실을 초래한다. 예를 들어 "IGNORE"로 선언된 마크구간(marked section)의 데이터는 파서가 파싱을 하지 않기 때문에 ESIS 을 이용하여 문서를 저장하는 경우에는 데이터의 손실을 유발한다. 또한 인스턴스에 참조되지 않은 엔티티, 표기법 등도 ESIS 에는 정보가 누락되어 있다. 따라서 저장시스템에 사용되는 파서의 출력물은 DTD 에 선언된 엔티티, 표기법 등의 정보와 "IGNORE"로 선언된 마크구간(marked section)의 정보를 모두 제공할 수 있어야 한다. 저장시스템에서 사용한 파서는 본 연구팀에서 개발한 KLESP 를 사용하였으며 KLESP 는 DTD 와 인스턴스의 파싱 결과로 메모리 포맷을 제공한다. DTD 파싱 정보는 RT(Rule Tree)로 제공되며, 인스턴스 파싱 결과는 본 연구팀에서 정의한 OXF(OX Format)으로 제공된다. 구현된 모듈별 기능은 다음과 같다.

- **DTD 임포터** : 특정 SGML 문서를 저장하기 위해서 먼저 DTD 가 시스템에 임포트되어 SGML 문서에서 참조되는 DTD 의 엔티티 선언, 표기법 선언 등의 정보를 저장한다.
- **인스턴스 임포터** : 인스턴스 임포터는 SGML 문서를 데이터 모델의 객체트리로 변환하여 저장하는 기능을 수행한다.
- **잠금 관리기** : SGML/XML 문서는 저장소에 조각화되어 저장되기 때문에 다수의 저자가 동일 문서의 특정 부분을 공동으로 저작할 수 있다. 잠금 관리기는 동일 문서에 대해 다수의 저자가 공동 저작을 하는 경우의 동시성 제어를 위한 잠금 관리 기능을 수행한다. 구현한 시스템의 공동 저작의 check-out 대상 객체는 SGML/XML 문서, 엘리먼트 부트리, 문서 형 선언 부분 집합에서 선언된 엔티티 등이다. check-out 대상 객체가 check-out 되어 잠금이 설정된 객체는 check-in 이 완료될 때까지 다른 사용자에게 의해 다시 check-out 될 수 없으며, read-only 트랜잭션일 경우에만 잠금이 설정된 객체에 대해서도 액세스가 가능하다.
- **버전 관리기** : 버전 관리기는 사용자가 문서 또는 문서 일부분에 대해 check-out/check-in 하여 문서 내용을 변경한 경우의 히스토리를 관리한다.
- **DBA 도구/사용자 관리기** : DBA 도구와 사용자 관리기는 데이터베이스 관리와 사용자 관리를 수행한다. 데이터베이스 관리 기능들로는 SGML/XML 스키마 초기화, 저장소 백업 등이 있다. 사용자 관리의 기능들로서는 사용자 등록 및 삭제, 사용자 권한 설정, 사용자 목록 관리 및 사용자 리스트 반환 등을 지원한다.

III. CALS/EC 와 SGML/XML 기반 문서 정보 시스템

“Work Anywhere, Anytime”을 위한 가상 기업 환경은 네트워크에 기반한 조직으로 극대화된 조직의 효율성을 요구한다. 이러한 환경에서의 문서 정보 시스템은 시스템간의 데이터 접근 및 교환의 효율을 높이고 미디어 독립적인 출판으로 문서 정보 기반 상호작용의 유연성을 높여 주어야 한다. 또한 CALS/EC 또는 가상기업에 적용될 문서 정보 시스템은 최소한의 네트워크 Bandwidth로 정보의 내용, 구조, 스타일, Behavior에 대한 Interoperability를 높여 주어 과거의 정보 시스템들이 특정 벤더에 의해 데이터가 종속되고 제어되는 것으로부터 자유로워질 수 있게 하여 “Software/system/language/media/vendor” 독립적인 정보 시스템으로 구축될 수 있어야 한다. 이런 측면에서 SGML/XML은 Enabling Technology로서 Interoperability를 높이는 프레임워크를 제공하지만 CALS/EC의 전분야와 전 문서 사이클을 지원하는 완벽한 SGML/XML 기반 total solution이 구현되기까지는 시간이 필요한 것은 사실이다. 기존의 PDM, EDMS 등의 정보 시스템 구축 패러다임과의 연계성을 높이며 단계별로 정보의 내용, 구조, 스타일, Behavior의 Interoperability를 높이는 방향으로 기술 개발이 지속적으로 이루어져야 한다. 1997년 XML이 발표된 이래 최근에는 Interoperability를 높이려는 목적으로 SGML/XML 기술을 적용하려는 움직임이 다각도로 나타나고 있는데 “Position Statement on Global Repository for XML Ver0.98, The XML/EDI Group, July, 1998”의 문헌에서 나타나듯이 SGML/XML 기술이 금융, 의료, 제품정보, 전자 상거래 등의 분야에서 각종 데이터들을 사용자 관점에서 하나의 일관된 비즈니스 시맨틱스로 정보를 검색하고 활용하는 일이 매우 중요한 이슈로 부각됨을 주목할 필요가 있다. 이러한 시각에서 XML을 이용한 DB 통합과 관련 응용 기술개발 경쟁이 국외에서는 활발히 진행되고 있으며 기술적으로 SGML/XML의 확장 응용 분야로 각종 다양한 정보 저장소의 정보들을 통일된 시맨틱스로 관리하는 지식관리 서버의 개발 요구가 많아지고 있다. 즉 과거의 정보 시스템에서 다루는 정보 단위보다 세분화된 논리적 단위의 정보를 일관되게 관리하고 사용자 요구에 맞는 비즈니스 콘텍스트의 정보로 유연하고도 동적으로 제시하는 하부 인프라로 SGML/XML 기반 문서 정보 시스템이 공헌할 수 있는 기대가 높아지고 있다.

IV. 결론

SGML/XML 문서 정보 시스템은 SGML/XML 문서 내에 표현되어 있는 계층적 논리적 구분을 이용하여 사용자에게 다양한 문서 접근점을 제공하며 각 논리 구분의 의미에 따라 다양한 각도로 문서에 대한 접근 점을 달리 해석하게 한다. 또한 문서의 스타일정보가 분리되어 있으므로 동적인 문서제시가 가능하여 불필요한 문서전송 없이 원하는 부분을 원하는 형태로 사용자측에 전달하고 다양한 사용자 시맨틱스로 재활용 가능하다. 이러한 SGML/XML 문서 정보 시스템을 구현하는 단계로는 SGML/XML에 표현된 정보를 얼마만큼 손실없이 유지하느냐에 따라 여러 유형의 시스템 구축이 가능하다. 중요한 점은 SGML이 추구하는 본질에 대한 타협없이 기술적인 저울질에

의한 합리적인 선택을 하여야 된다는 점이다. 장기적인 투자 의지없이 기술적인 타협에 의한 SGML/XML 문서 정보 시스템의 구축은 SGML/XML 기술이 제공하는 장점을 인식하기 전에 다수의 실망하는 사용자들을 만들 수 있다는 점에 유의해야 된다. 한 예로, 최근 국내외의 어려운 사정을 반영하여 지식 경영에 대한 관심은 높지만 그 목적을 달성하기 위한 Enabling Technology로서 SGML/XML 문서 정보 시스템이 구체적인 수단이 될 것이라는 것을 인식하는 경영자와 시스템 공급자는 많지 않은 실정이며 장기적인 관점 보다는 외형적인 시스템 구축에 급급한 점은 개선되어야 한다. 그 만큼 지식관리 차원으로 확장되기 위한 SGML/XML 문서 정보 시스템은 투자에 대한 최고 경영자 또는 정부의 의지가 중요하며 조직의 BPR (Business Process Re-engineering) 차원에서 검토되어야만 한다. 특히, SGML/XML 문서 정보 시스템의 DTD 설계 단계는 데이터베이스의 스키마로서 조직의 정보 모델링의 관점에서 접근하여 충분한 투자로 설계되는 것이 필요하다. 미국, 일본, 스웨덴 CALS/EC 프로그램에서 살펴보면 상당한 투자를 DTD 설계 및 PDM 시스템과의 연계를 위한 정보 모델링에 할애하고 있는 점을 유념하여 본받을 필요가 있다. 이와 함께 기술적으로는 SGML/XML 이 표현하는 모든 구조정보를 검색에 이용하기 위해서는 효율적인 Phase 4 수준의 검색시스템이 개발되어야 하며, 구조적 질의에 대한 연구 및 구조정보를 손쉽게 질의할 수 있는 질의 인터페이스등이 연구되어야 한다. 나아가, SGML/XML 문서 정보 시스템이 지식 관리 시스템의 하부 인프라로 정착되기 위해서는 의사결정 과정 및 타스크와 연계된 각 정보 저장소의 메타지식구조에 대한 정의, HyTime/XLL 등의 표준 스펙을 이용한 링크 관리, SGML/XML 문서 포맷 이외의 다양한 종류의 문서 포맷에 대한 통합 관리, Knowledge Brokering 을 위한 콘텍스트 관리, 워크플로우, 에이전트 등의 기능을 지원하는 차세대 문서 관리 시스템으로 확장되는 것이 필요하다.

참고문헌

- [1] C. F. Goldfarb and Yuri Rubinsky, *The SGML Handbook*, Clarendon Press, Oxford, 1990.
- [2] W3C XML Working Group, *Extensible Markup Language (XML)*, Available via <http://www.w3c.org/XML>, 1998.
- [3] Arbortext, Inc. *Getting Started with SGML*, Available via <http://www.arbortext.com/wp.html>, 1995.
- [4] V. Christophides, et al, "From Structured Documents to Novel Query Facilities," *ACM SIGMOD*, pp. 313-324, Minesota, USA, 1994.
- [5] G. E. Blake, M. P. Consens, P. Kilpelainen, P. A. Larson., T. Sinder, and F. W. Tompa, "Text/Relational Database Management Systems: Harmonizing SQL and SGML," In *Proceedings of the International Conference on Applications of Databases*, Vadstena, Sweden, 1994.
- [6] R. Sacks-Davis, T. Arnold-Moore, and J. Zobel, "Database Systems for Structured Documents," In *International Symposium on Advanced Database Technologies and Their Integration*, pp. 272-283, Nara, Japan, 1994.
- [7] Patricia Francois, "Generalized SGML Repositories: Requirements and modelling," *Computer Standards and Interfaces* 18, 1996
- [8] R.G.G. Cattell, *The Object Database Standard: ODMG-93*, Morgan Kaufmann Publishers, Inc., San

Francisco, California, 1996.

- [9] O2 Technology, ODMG C++ Binding Guide, Release 4.6, December 1996.
- [10] 최종연구보고서 SGML 테스트 데이터 구축, 전자통신연구원, 1997
- [11] 우리말 정보처리 S/W 기술 개발에 관한 연구, 정보통신부, 전자통신연구원, 1997, 1998
- [12] Brian E. Travis, Dale C. Waldt, "The SGML Implementation Guide" Springer, 1995
- [13] Charles F. Goldfarb,, "The SGML Handbook", Clarendon Press, Oxford, 1990
- [14] ISO 8879, Information Processing – Text and Office Systems – Standard Generalized Markup Language (SGML), Includes Amendment 1, 1988
- [15] ISO/IEC 10179:1996(E), Information Technology – Processing Language- Document Style Semantics and Specification Language (DSSSL)
- [16] Massimo Melucci, Passage Retrieval : A Probabilistic Technique, Information Processing & Management, Vol34, No. 1, pp. 43-68, 1998
- [17] R. Wilkinson, "Effective retrieval of Structured Documents", Proc. Of ACM SIGIR'94, pp 311-317, Dublin City, 1994