

코어 컴포넌트 기반 시맨틱 라이브러리의 설계

Using core components to design semantic libraries

정 용 규 (Yong Gyu Jung)

초 록

시맨틱 라이브러리는 의미사전을 구축함으로써 EDI 전자문서의 교환에 활용될 수 있다. 본 논문은 실무 개발자들이 메타데이터를 이용한 의미사전을 개발하는 과정에서 참고할 수 있는 시맨틱 라이브러리 설계정보를 기술한다. 시맨틱 라이브러리를 구성하는 요소로는 크게 시맨틱 요소(Semantic Element)와 시맨틱 단위(Semantic Unit), 매핑 테이블(Mapping Table) 등 3개로 구성된다. 본 논문에서는 이러한 구성요소들의 기본적인 특성과 개발 관련 설계 방법을 제안한다. 또한 이러한 구성요소와 제반 규칙을 준용하여 메타 데이터 간 의미적 교환을 위한 사전을 구축한 사례를 소개한다.

영 문 초 록

Semantic libraries can be used for EDI messages to exchange by implementing the semantic dictionaries. This paper describes the design information of semantic libraries for the field engineers to implement the semantic dictionary using metadata. The components of semantic libraries are semantic elements, semantic units and mapping tables. The basic characteristics and design methods related implementing are proposed. Also the metadata semantic dictionaries including the components and rules are introduced.

키워드: 전자문서교환, 의미사전, EDI, semantic, library, metadata

1. 서론

시맨틱 라이브러리를 개발하고자 하는 시도는 아주 오래전부터 계속되어져 왔다. 특히 자연언어처리 분야에서 상당히 많은 연구가 있어 왔으며, 각 어휘나 단어들의 시맨틱 정의와 단어들 간의 관계에 대한 규칙들을 연구해 왔다. IT 분야에서는 ISO/IEC JTC1 SC32를 중심으로 메타데이터에 대한 연구를 해왔으며, ISO TC154에서는 BSR(Basic Semantic Registry) 프로젝트를 수행하였으며 TC154내 워킹그룹1에서 이런 연구가 수행되고 있다. XML의 등장은 메타데이터와 시맨틱이 다시 주목을 받게 되는 주요 요인이었다. XML이 등장하기 이전의 메타데이터는 주로 축약어나 의미를 표현하는 기호나 구문, 어휘 등을 자체 정의하여 사용하였으며, 특히 EDI에서는 국제 표준인 UN/EDIFACT나 북미표준인 ANSI X.12는 자체 정의한 데이터 엘리먼트 또는 데이터 세트 등으로 전자문서를 개발하여 전자거래에 활용하였다. 이러한 상황에서 XML의 등장은 자체적으로 가지고 있는 표현과 의미의 일치성과 마크업 언어의 확장성 때문에 이전에 의미와 표현이 분리된 상태로 활용되었던 구문을 하나로 만들어 사용할 수 있는 기반을 제공해주고 있다.

현재 XML에 대한 데이터 표준화 작업이 여러 단체에서 진행되고 있는데,

특히 UN/CEFACT TMG 그룹에서는 코어 컴포넌트 기술규격을 만들어 XML 라이브러리를 개발하기 위한 기본 개념과 방법을 제시하고 있다. 아울러 코어 컴포넌트 기술 규격을 준용하여 OASIS의 UBL 기술위원회에서는 자체적으로 XML 라이브러리를 개발하였으며, TBG 그룹에서도 도메인별로 코어 컴포넌트를 개발하고 있으며, ATG 그룹에서는 개발된 코어 컴포넌트에 대한 XML 스키마 명명 및 설계 규칙을 개발하고 있다.

2. 코어 컴포넌트

코어 컴포넌트 기술 규격은 UN/CEFACT가 추진한 ebXML 프로젝트에서 XML Dictionary를 개발하기 위한 방법론으로 ISO/IEC 11179 표준을 기반으로 하여 개발한 규격이다. 코어 컴포넌트 기술 규격은 크게 코어 컴포넌트에 대한 개념 설명과 코어 컴포넌트 명명규칙, 코어 컴포넌트의 발견 규칙 등을 크게 다루고 있으며, XML로 메타데이터를 개발할 때 코어 컴포넌트 기술 규격을 준용할 것을 권고하고 있다. OASIS에 있는 UBL 기술위원회에서는 코어 컴포넌트 기술 규격에 기반하여 UBL 라이브러리를 개발하였으며, UN/CEFACT에서도 코어 컴포넌트 기술 규격에 기반하여 XML 컴포넌트를 개발하고 있으며, XML 스키마의 명명

및 설계 규칙 초안 작업도 추진 중에 있다. 코어 컴포넌트 기술 규격은 코어 컴포넌트와 비즈니스 정보 개체의 2가지 개념을 기반으로, 전자문서의 비즈니스 정보를 ebXML 환경에서 표현할 수 있는 지침을 제시해 준다.

2.1 코어 컴포넌트의 개념

코어 컴포넌트들은 비즈니스 도메인에서 가장 빈번하게 사용되는 비즈니스 정보 개체들로부터 추출되어 모델링 된다. 코어 컴포넌트의 목표는 국제적으로 상호운용성을 가능하게 하고 여러 다른 비즈니스 도메인에서도 서로 간에 재사용할 수 있도록 하는 것이다. 코어 컴포넌트 기술 규격은 이러한 코어 컴포넌트들을 발견하거나 생성하는 방법을 기술한다. 코어 컴포넌트 기술 규격에서 주요 개념은 코어 컴포넌트이다. 코어 컴포넌트는 모든 전자 비즈니스 메시지를 구축하는 데 기초가 되는 시맨틱 빌딩 블록이라 할 수 있다. 의미적 빌딩 블록은 올바르고 의미있는 정보 교환 패키지이다. 여기서의 정보는 특정한 개념을 기술하는 필요한 정보 단위만을 포함한다. 코어 컴포넌트는 기본 코어 컴포넌트, 연관 코어 컴포넌트, 코어 컴포넌트 유형, 집합 코어 컴포넌트, 기본 코어 컴포넌트의 표현 형태를 나타내는 데이터 유형의 5개의 서로 다른 카테고리를 가진다.

- 기본 코어 컴포넌트 : 객체 클래스를 표현하는 특정한 집합 코어 컴포넌트 내에서 단일한 비즈니스 특성을 나타내는 코어 컴포넌트이며 유일한 비즈니스 시맨틱 정의를 가지고 있다. 기본 코어 컴포넌트는 집합 코어 컴포넌트의 속성을 표현하고, 속성 값들의 세트를 정의해준다.

- 연관 코어 컴포넌트 : 객체 클래스를 표현하는 특정한 집합 코어 컴포넌트 내에서 복잡한 비즈니스 특성을 나타내는 코어 컴포넌트이다. 연관 코어 컴포넌트는 연관 코어 컴포넌트 속성과 실제 구조를 가지고 있는 연관 코어 컴포넌트로 이루어진다.

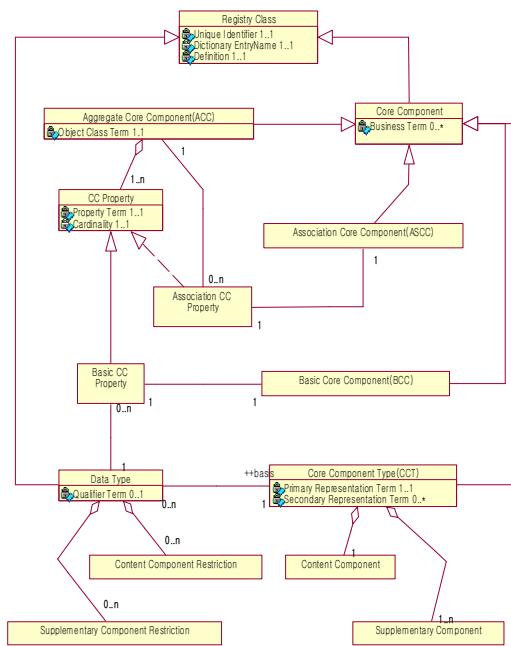
- 코어 컴포넌트 유형 : 코어 컴포넌트는 실제적인 값을 가지고 있는 하나의 콘텐츠 컴포넌트와 콘텐츠 컴포넌트의 부가적인 정의를 주는 하나 이상의 부가 컴포넌트를 가진다. 단, 코어 컴포넌트 유형은 비즈니스 의미를 지니지 않는다.

- 집합 코어 컴포넌트 : 특정한 비즈니스 컨텍스트에 독립적이면서, 명확한 비즈니스 의미를 가진 비즈니스 정보들의 연관된 개체들의 집합이다. 집합 코어 컴포넌트는 모델링 용어로 표현되며, 특정한 비즈니스 컨텍스트에 독립적인 객체 클래스로 표

현된다. 코어 컴포넌트들은 데이터 유형들로 정의된 속성들을 가진다.

- 데이터 유형 : 특정한 기본 코어 컴포넌트 속성들과 기본 비즈니스 정보 개체 속성에 사용될 수 있는 유효한 값들의 세트를 규정한다.

[그림1]은 등록저장소를 활용하여 코어 컴포넌트를 관리하기 위한 메타 모델을 표현한다.



[그림1] 코어 컴포넌트 구조

2.2 비즈니스 정보 개체의 주요 개념

코어 컴포넌트와 비즈니스 정보 개체

들의 주요한 차이점은 비즈니스 컨텍스트의 개념이다. 비즈니스 컨텍스트는 특정한 비즈니스 환경들에서의 활용에 따라 코어 컴포넌트가 한정화되고 재정의되는 메커니즘이다. 비즈니스 프로세스 정의는 메시지와 메시지 내 콘텐츠의 활용에 대한 고수준 레벨의 설명을 제공한다.

- 비즈니스 컨텍스트 : 특정한 비즈니스 환경에 대한 공식적인 설명으로 유일하게 구분해질 수 있는 서로 다른 비즈니스 환경들을 허용해주는 컨텍스트 카테고리들의 세트들에 대한 값들을 나타낸다.

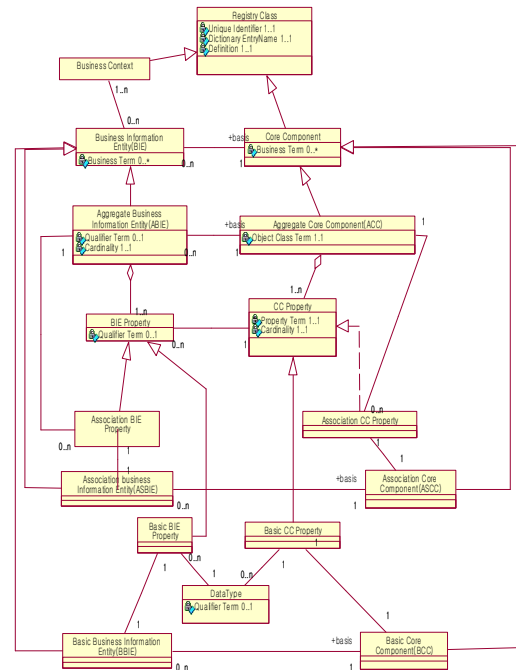
- 비즈니스 정보 개체 : 고유한 비즈니스 시맨틱 정의를 지닌 비즈니스 데이터 개체 또는 비즈니스 개체들의 그룹이다. 비즈니스 정보 개체는 기본 비즈니스 정보 개체와 연관 비즈니스 정보 개체, 집합 비즈니스 정보 개체로 구분된다.

- 기본 비즈니스 정보 개체 : 특정한 비즈니스 컨텍스트에서의 특정의 객체 클래스의 단일한 비즈니스 특성을 나타낸다. 고유한 비즈니스 의미 정의를 가진다. 기본 비즈니스 정보 개체는 기본 비즈니스 정보 개체 속성을 나타내며, 실제 값을 기술하는 데이터 유형들과 연결된다. 기본 비

즈니스 정보 개체는 기본 코어 컴포넌트로부터 파생된다.

- 연관 비즈니스 정보 개체 : 특정한 비즈니스 컨텍스트에서 객체 클래스를 표현하는 특정한 집합 코어 컴포넌트 내에서 복잡한 비즈니스 특성을 나타낸다. 연관 비즈니스 정보 개체는 연관 비즈니스 정보 개체 속성과 실제 구조를 가지고 있는 연관 비즈니스 정보 개체로 이루어진다. 고유한 비즈니스 시맨틱 정의를 가진다. 연관 비즈니스 정보 개체는 연관 코어 컴포넌트로부터 파생된다.
- 집합 비즈니스 정보 개체 : 특정한 비즈니스 컨텍스트에서 명확한 비즈니스 의미를 가진 비즈니스 정보들의 연관된 개체들의 집합이다. 집합 비즈니스 정보 개체는 모델링 용어로 표현되며, 특정한 비즈니스 컨텍스트에 있는 객체 클래스로 표현된다.

[그림2]는 등록저장소를 활용하여 비즈니스 정보 개체를 관리하기 위한 메타 모델을 표현한다.



[그림2] 비즈니스 정보 개체 구조

3. 시맨틱 라이브러리 구축

시맨틱 라이브러리는 크게 시맨틱 요소와 시맨틱 단위, 매핑 테이블 등 3개로 구성된다.

3.1 시맨틱 요소

시맨틱 라이브러리를 개발하는 과정에서 여러 가지 문제가 나타날 수 있다. 시맨틱 라이브러리를 실제로 개발하는 과정에서 제기되는 여러 가지 문제들은

아래와 같이 제시될 수 있다.

- 시맨틱 요소 항목의 의미적인 동의어와 의미적인 포함관계를 명확하게 구분 지어야 한다. 예를 들어, 물품과 제품, 품목이라는 시맨틱 요소들이 있을 경우 이 3개의 시맨틱 요소 항목간의 의미적인 포함관계를 구분 지어야 한다. 개발 실무자들마다 각자의 견해에 따라 시맨틱 요소의 의미적인 관계를 다르게 설정할 수도 있을 것이다. 어떤 사람은 물품이라는 시맨틱 한정요소를 가장 큰 범위로 둘 수 있고, 어떤 사람은 품목이라는 시맨틱 요소를 가장 큰 의미요소로 생각할 수도 있다. 이러한 문제 때문에 시맨틱 요소는 단순히 시맨틱 요소 항목들의 목록만을 개발하여서는 안되며, 시맨틱 요소 항목들간의 포함 관계, 연관 관계들까지 같이 개발하여야 한다. 아래 표는 각 요소들의 정의와 정의를 통한 요소들간의 관계를 구분짓고 있다.

[표1] 시맨틱 요소들간의 관계

시맨틱 항목	정의	동의어
물품	일정하게 쓸 만한 값어치가 있는 물건	물건, 상품, 물자
제품	원료를 써서 물건을 만듦. 또는 그렇게 만들어 낸 물품.	물품, 물건, 상품
품목	물품의 이름을 쓴 목록.	품명, 품류

위의 [표1]을 보면 우리는 물품이나

제품의 경우에는 물건이나 상품 객체를 직접적으로 지칭하는 시맨틱 요소 항목이라는 것을, 그리고 물품과 제품이 거의 비슷한 의미로 사용된 것을 알 수 있다. 품목의 경우는 직접 물건이나 상품 객체를 직접 지칭하는 것은 아니고, 물품의 이름을 쓴 목록이다. 즉, 품목과 물품, 제품은 품목이라는 연관관계를 가진다. 만약 주문 프로세스 상에서 사용되는 시맨틱 요소들을 추출할 때, 주문_물품과 주문_품목, 주문_제품 중 선택이 필요하다. 일단 주문_물품과 주문_제품은 먼저 동의어 처리를 하여야 할 것이다. 시맨틱 요소 항목으로 모두 등록할 수 있지만, 동의어 키텔에 서로 등록하도록 하며, 주문_품목은 좀더 큰 의미로 사용될 수 있을 것이다. 그래서 주문_품목에서 세부적인 물품의 속성을 언급할 때는 주문_물품이나 주문_제품 항목들을 사용하여 기술하면 된다. 시맨틱 요소 항목들은 반드시 한글사전에 기초하여 등록하여야 하지만, 전자상거래나 e비즈니스가 국제적인 범위에서 전개되기 때문에 불가피한 경우에 한해서는 영어나 외국어가 활용되도록 한다.

3.2 시맨틱 단위 개발

시맨틱 단위는 시맨틱 요소들의 조합으로 이루어지기 때문에 상당히 많은

양의 시맨틱 단위를 개발하게 된다. 그 과정에서 발견될 수 있는 경우는 아래와 같다.

- 시맨틱 단위는 ebXML CCTS의 개체를 사용하고, CCTS에서 규정하는 규칙들을 활용하여 정의한다. 따라서, 시맨틱 단위는 반드시 세가지 구성요소인 객체용어, 속성용어, 표현용어를 가지고 있어야 한다. 단, 속성용어와 표현용어가 같은 의미를 지니고 있을 경우에는 속성용어를 생략하도록 한다. 이러한 생략은 UN/CEFACT 코어 컴포넌트 기술 규격에서도 적용하고 있기 때문에 크게 문제가 되지는 않을 것이다. 예를 들어 포장_수량이라는 시맨틱 단위는 포장이라는 객체용어와 수량이라는 표현용어로 이루어져 있지만, 표현용어와 똑같은 수량이라는 속성을 생략한 채 완전한 시맨틱 단위로서 역할을 하고 있다.

- 한정된 시맨틱 객체는 원래의 시맨틱 객체가 가지고 있는 속성외에도 속성을 한정된 새로운 형태의 속성을 가질 수 있다. 예를 들면 구매_당사자라는 시맨틱 객체는 당사자라는 시맨틱 객체가 가지고 있는 속성과 똑같은 속성만을 갖는 것이 아니라 추가적으로 한정된 속성들을 가질 수 있다

- 시맨틱 단위에서 사용되는 시맨틱 요소가 동의어를 가질 경우, 시맨틱 단위에서도 동의어 처리를 하여야 한다.

3.3 매핑 테이블 개발

매핑 테이블의 개발은 대상 스펙이나 디렉토리를 분석하는 과정이라 할 수 있다. 대상 스펙이나 디렉토리의 설계구조와 각 데이터 항목에 대한 정확한 의미 파악이 중요하다. 매핑 테이블의 개발은 시맨틱 단위와 대상 스펙이나 디렉토리 간의 의미적 연결에만 한정한다.

매핑 테이블의 연결자는 대상 스펙이나 디렉토리의 구문에 맞춰서 개발되어야 한다. 예를 들어 EDIFACT나 ANSI X.12와 같은 EDI 디렉토리는 자체적인 구문 규칙을 가지고 있다. 각 한정어 코드와 데이터 엘리먼트를 조합하여 데이터를 표현하기 때문에 복잡한 매핑 테이블 연결자를 기술하여야 할 것이다. 또한 대상 스펙이나 디렉토리가 XML로 표현되었을 경우 XML의 문법에 맞게 기술되어야 한다. 여기서 주의할 것은 대상 스펙이나 디렉토리가 XML로 되었다더라도 EDIFACT와 같이 한정어 코드를 활용하여 데이터를 표현하는 XML/EDI방식일 경우에는 한정어 코드까지 같이 기술해주어야 한다.

EDIFACT와 같이 한정어 코드가 있는 경우에는 한정어 코드를 반드시 사전조건으로 지정해주고 이후에 데이터 항목을 기술하여야 한다. 사전조건일 경우에는 “[]”으로 구분하여 먼저 기술하여주도록 한다. 아래의 예는 EDIFACT 매핑 테이블의 한 예이다.

예제 : 확인수신자.구매자 할당.식별자 : NAD+ 3035[AK]+ C082:[실제 값 영역>::3055[92]

XML로 기술하였을 경우에는 XPath 규격을 활용하여 기술하도록 한다.

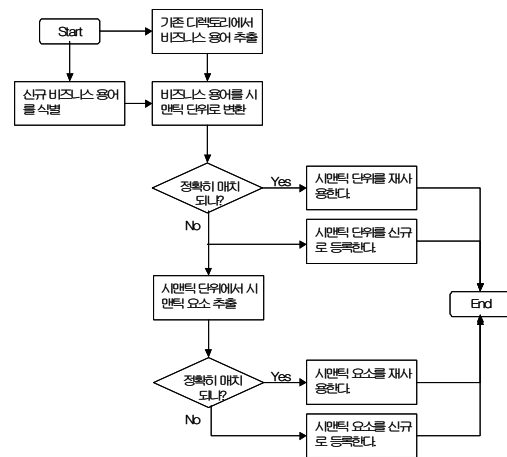
예제 : 회계_수입_부서_연락처.전화.식별자 : AccountingReceivableContact/Telephone.Text

시맨틱 단위와 매핑 테이블 연결자에 대한 개념적 동의 정도를 기술할 때는 3가지 레벨로 기술한다. 실제로 활용되는 것은 정확한 일치에 해당되는 개체들만 활용하여야 한다.

3.4 시맨틱 라이브러리 개발

아래 그림은 시맨틱 라이브러리 콘텐츠를 개발하기 위한 절차를 간단히 도식화한 것이다. 시맨틱 라이브러리 콘텐츠를 개발하는 시작점은 기존의 스펙

이나 디렉토리를 분석하여 비즈니스 용어를 추출하거나, 신규로 메타데이터나 전자문서를 개발할 때 새로이 비즈니스 용어가 나오는 경우이다. 이러한 비즈니스 용어를 먼저 시맨틱 단위로 변환한다.



[그림] 시맨틱 라이브러리 개발 절차

시맨틱 단위로 변환할 때 주의할 점은 속성과 표현용어만이 있고 시맨틱 객체가 정확히 기술되지 않은 경우에는 시맨틱 요소로만 등록하도록 한다. 시맨틱 단위나 시맨틱 요소가 정확하게 일치하는 경우에는 신규 등록하지 않고 재사용하면 되고, 만약 일치하지는 않지만, 의미적으로 동일한 의미가 있을 경우에는 신규로 등록하면서 의미적으로 동일한 항목과 동의어 처리를 하도록 한다. 그리고 만약 기존 스펙이나 디렉토리를 분석하여 시맨틱 라이브러리가 개발된 경우에는 해당 대상 디렉토리에

해당하는 매핑 테이블에 기술하여야 한다. 이러한 설계기준을 준용하여 물류업 중에 필요한 시맨틱 라이브러리를 구축하였다. 2,000여 Entry를 수집하여 각 의미적 매핑이 가능한 항목을 구성하였고 이를 통해 전자문서의 의미적 교환이 가능하도록 의미사전을 등록저장소에 Database로 구축하였다. 의미사전은 실제로 수출입 운송분야에 적용되는 EDI 전자문서의 항목을 중심으로 구성하였으며 이를 통하여 XML 문서와의 의미적 교환이 가능하도록 하였다.

4. 결론

정보 시스템(Information System)을 설계하고 구현하면서 개발자들은 파일 시스템을 만들거나 데이터베이스를 설계하고 구축하며, 기업, 조직의 시스템 간 메시지를 교환한다. 정보화나 e비즈니스가 발달하고 규모가 커지면서 이제 는 시스템 내부에서 뿐만 아니라 시스템 간 상호운용성까지 기업내부에서 사용하는 메타데이터의 상호연동을 고려하게 되었다. 이러한 메타데이터의 상호연동을 위해서는 기업 내부와 외부에서 사용하는 메타데이터에 대한 시맨틱을 명확히 하고 시맨틱 레벨에서 매핑을 할 필요성이 제기된다. 시맨틱 라이브러리는 개별 시스템 개발자들이나 설계자들이 기업 내부 시스템의 메타데이터를 개발하기 위한 공통 기준을 제시한다.

본 논문에서는 CC 및 BIE에 대한 개념을 설명하고, 이 개념을 바탕으로 시맨틱 라이브러리를 구성하는 시맨틱 요소, 시맨틱 단위, 매핑 테이블과의 연관관계를 기술하였다. 이러한 시맨틱 라이브러리 개발 및 설계 지침은 개별 실무자들이 통일적이고 일관되게 메타데이터를 개발하는데 필요한 기준 규격으로 활용하고자 하였다.

참 고 문 헌

- 김동광, 정갑주, 신호섭, 황선태. 2006. “XML Schema기반 시맨틱 데이터 통합”. 한국컴퓨터종합학술대회 논문집. 한국정보과학회
- 김현식, 이상연, 김인철. 2006. “시맨틱 웹 서비스 조합을 위한 계획 영역 지식 변환기”. 한국컴퓨터종합학술대회 논문집. 한국정보과학회
- 문희철, 심상렬. 1998. 「무역자동화와 EDI」. 무역경영사
- 오호근. 1993. 「EDI란 무엇인가」. 크라운출판사
- 임영철. 2005. “XML 전자문서의 국제 표준화 현황”. eBizKorea 2005년12월호. 한국전자거래진흥원
- 장재경. 2001. “전자상거래 표준화 패러다임의 변화”. e-Commerce 통권 31호. 한국전자거래진흥원
- 정용규, 김운섭. 2002. “UN/CEFACT 전자문서의 XML변환규칙”, 한국정보

- 과학회 학술발표논문집 29권1호
- 최중민. 2003. “시맨틱 웹의 개요와 연구 동향”. 한국어정보처리연구회 동계 튜토리얼, 2004. 2. 13
- 한국전산원. 1998. 「정부 EC 플랫폼 발전방안에 관한 연구」
- 한국전자거래진흥원. 1998. 「EDIFACT 메시지 구현 지침」
- Barbosa, D. Barta, A. Mendelzon, A. Mihaila, G. Rizzolo, F. and Gianolli, Rodriguez P. "ToX - The Toronto XML Engine". International Workshop on Information Integration on the Web, Rio de Janeiro, 2001.
- Bryan, M. 1998. "Guidelines for using XML for Electronic Data Interchange"
- Chamberlin, Donald D. Robie, Jonathan, and Florescu, Daniela. "An XML Query Language for Heterogeneous Data Sources". WebDB 2000
- Connolly, D. 1997. 「XML: Principles, Tools, and Techniques」. O'Reilly
- Cooper, Brian. Sample, Neal. Franklin, Michael J. Hjaltason, Gisli R. and Shadmon, Moshe. "A fast index for semistructured data". VLDB 2001
- Fernandez, Mary F. Morishima, Atsuyuki and Suciu, Dan. "Efficient Evaluation of XML Middle-ware Queries". SIGMOD 2001
- Moura, De Carvalho A. M. Campos, machado M. L and Barreto, C. M. 1998. "A survey on metadata for describing and retrieving Internet resources"
- W3C. 2004. "OWL Web Ontology language Overview", W3C Recommendation 10 February 2004.



현재 을지대학교 의료산업학부 교수
 1981년 서울대학교 졸업 (이학사)
 1994년 연세대학교 졸업 (공학석사)
 2003년 경기대학교 졸업 (이학박사)