사무용문서교환형식의 [illegible] 시스템의 구현 연구

최 현 섭, 박 종 훈, 홍 온 선, 김 철 원, 최 기 호
광운대학교 전자계산기공학과

A Study on an Encoding System of Office Document Interchange Format

Hyun Sub Choi, Jong Hun Park, On Sun Hong, Chut Won Kim, Ki Ho Choi
Dept. of Computer Eng. Kwang woon Univ.

ABSTRACT

This paper presents an encoding system of Office Document Interchange Format in order to encode the ASN.1 notation of formatted mixed-mode document for data values defined in ISO 8824/CCITT X.208 into the octet stream suitable for transmission, after aplying the basic encoding rule defined in ISO 8325 CCITT X.209 and the data format defined in ISO DIS 8613.

I. 서 론

최근 사무실내의 computing환경은 PC/Personal Computer를 중심으로 모든 작업과 업무가 진행하게 됨에 따라 PC 간의 통신도 필수적으로 되면서 사무용 시스템간에 사무용 문서를 교환해야 하는 필요성이 대두되고 있다.

이러한 필요성은 PC와 고도화된 장비들의 사무자동화가 더욱 명확해졌으며 따라서, 사무실에서 사용자가 생성한 텍스트, 그래픽, 이미지, 데이터등을 포함하는 다양한 범위의 문서를 교환하기 위해서는 통신 서비스 및 사용된 프로토콜과 문서를 표현하는 정보의 엔코딩이라는 두가지 측면을 고려해야만 한다. 본 논문에서는 페이지내에 문자와 래스터정보를 포함하는 포맷화된 혼합형 문서(Formatted Mixed Mode Document)의 교환을 위한 엔코딩 시스템을 구성한다. 이를 위해 현재 ISO와 CCITT가 정한 표준화방식에 따라 사무용 문서에 대한 정보를 교환할 수 있도록 CCITT 권고안 X.208에 정의된 규정된 문서교환형식 ASN.1(Abasract Syntax Notation One)과 X.209에 정의된 Basic Encoding rule 및 ISO DIS 8613 - 5의 데이타 포맷(data format)에 따라 UNIX상의 컴파일러 생성도구인 LEX와 YACC를 이용하여 엔코딩 시스템을 생성시키도록 시스템을 모델화하고, 구성한다.

II. 시스템의 구성

1. 사무용 문서구조(ODA)

1.1 문서구조 모델

문서처리 시스템은 [illegible] 부 데이타 구조를 [illegible] 실행되는 오퍼레이션에 의해 [illegible] 이들 내부 데이타 구조는 [illegible] 다. 이상적으로는 이들 변환은 [illegible] 가능해야 하고, 이를 위해서는 [illegible] 맷이 동일 문서구조로부터 유도되면 [illegible]

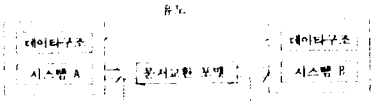


그림 1 문서구조 모델

1.2 문서구조

문서구조 (document structures)는 ODA및 교환 포맷의 가장 핵심적인 부분으로, 포맷화된 형식 문서들과 처리 가능한 형식문서들의 교환을 규정하고, 서로다른 정보유형을 포함하는 문서들의 교환을 규정하는 목적을 가지고 있다. 문서구조는 그림 2에서 보인것 처럼 문서를 이루는 구성요소들과 특성들 사이의 구조적인 관계를 기술하는 레이아웃 구조(layout structure)와 논리적 구조(logical structure)로 기술될 수 있으며, 문서 종류는 공통구조들과 특성들을 갖는 문서들의 집합을 의미하며 문서 프로파일은 전체적인 문서에 관련되는 속성들, 즉 문서의 제목과 문서의 역사및 문서의 내부적 특징을 기술한다.

문자 내용구조는 문자로 코드화된 텍스트에 속하는 내용구조들로 3가지 종류로 정의되며 ISO 6937과 6429에 정의된 제어함수들과 문자집합을 기초로 하고 있고, 라스터 그래픽 내용구조는 두 레벨(흑과 백)그림 요소들을

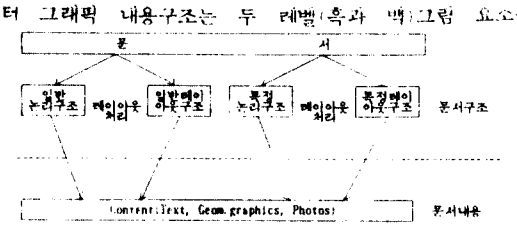


그림 2 문서구조와 문서내용

의 장방형 배열들로 구성하는 이미지나 그림등도 이루어진 내용의 구조로서, 두개의 종류로 정의되며 포맷화된 형식은 포맷화된 형식 문서들에서만 사용될 수 있다.

2. 사무용문서교환형식(ODIF)

ODIF는 ISO 8613-2에 따라 구조화된 문서를 교환하기 위해, 사용되는 데이타 스트림(data stream)의 형식을 정의하고 구조화된 문서의 오브젝트, 내용, 스타일의 표현을 정의하며, 교환된 문서에서 나타날 수 있는 속성(속성들의 집합)의 표현을 정의한다.

2.1 교환 형식

ISO 8613-2에 따라 구조화된 문서는 다음과 같은 데이타 구조들을 이루는 데이타 스트림에 의해 표현된다.

(1) 문서 프로파일 기술자(document profile descriptor)
(2) 레이아웃 오브젝트 기술자(layout object descriptor)
(3) 레이아웃 오브젝트 부류 기술자(layout object class descriptor)
(4) 논리 오브젝트 기술자(logical object descriptor)
(5) 논리 오브젝트 부류 기술자(logical object class descriptor)
(6) 표현방식 기술자(presentation style descriptor)
(7) 레이아웃 방식 기술자(layout style descriptor)
(8) 텍스트 유니트(text unit)

이러한 데이타구조를 interchange data element라고 하며 데이타 스트림내에서 interchange data element들은 지정된 어떤 규칙에 따라 교환형식 부류 A와 교환형식 부류 B가 있으며 본 논문에서는 부류 B를 사용한다.

2.2 교환형식 부류 B

교환형식 부류 B는 특정 논리구조나 일반 논리구조를 갖지않는 문서, 즉 포맷화된 문서구조 부류에 맞는 문서를 표현하는데 이용된다.

• 송수신 순서(interchange data element 순서)
(1) 문서 프로파일 기술자
(2) 레이아웃 오브젝트 부류 기술자와, 관련 텍스트유니트
(3) 표현 방식 기술자
(4) 레이아웃 오브젝트 기술자와, 관련 텍스트 유니트

2.3 기술자와 텍스트 유니트

텍스트 유니트는 다음과 같이 하위의 데이타구조와, 관련 내용부분의 속성을 나타내는 데이타항목으로 구성되는 데이타구조인 속성 필드(attribute field)와 데이타 항목이나 관련 내용부분을 이루는 내용 요소(content portion)를 나타내는 데이타 구조인 정보 필드(information field)로 구성된다.

2.4 데이타 포맷

ISO 8613에서 ODIF를 위한 데이타 포맷은 교환 데이타 유니트들, 문서 프로파일 기술자, identifier들과 extension들, 레이아웃 기술자들, 논리적인 기술자들, 표현및 레이아웃 기술자들, default값 리스트들, 텍스트 유니트 및 데이타 유니트들로 정의해 놓고 있다.

3. ASN.1

X.208은 복잡한 type을 정의하는데 이용되고, 이들 type의 값을 정할수 있게하는 표기법을 설명하고 있는데 이러한 기능을수는 표기법을 abstract syntax difinition을 위한 표기법이라 하며 CCITT의 X.208에서는 이러한 표기법으로서 ASN.1(Abstract Syntax Notation One)를 정의한다.

ASN.1은 simple type들을 tag와 함께 정의하며, 이들 type을 참조하고, 이들 type들의 값을 명세하기 위한 표기법을 설명하며 기본 type들을 이용하여 새로운 type를 구성하는 기법을 정의한다.

이 권고안에서 설명된 표기법에 의해 정의된 모든 type은 'tag'로 지정된다. 많은 여러가지 type에 같은 tag가 지정되는 것이 보통이고, 복잡한 type은 tag가 사용된 문맥에 의해서 확인된다.

기본적으로 universal, application, context-specific, private등 4가지 tag의 class가 있다.

3.1 생성 규칙

1) ASN.1 항목은 표1에 있는 문자들의 sequence로 구성 된다.

표1. ASN.1 character

A - Z
a - z
0 - 9
: , [ ]
< ' ( )

2) ASN.1의 문자로 만들어지는 모든 sequence는 ASN.1 항목으로 분류되며, 각 항목에는 이름이 있다.

표2. ASN.1 항목의 이름

BOOLEAN	INTEGER	BIT	STRING
OCTET	NULL	SEQUENCE	OF
SET	IMPLICIT	CHOICE	ANY
EXTERNAL	OBJECT	IDENTIFIER	OPTIONAL
DEFAULT	COMPONENTS	UNIVERSAL	APPLICATION
PRIVATE	TRUE	FALSE	BEGIN
END	DEFINITIONS	EXPLICIT	ENUMERATED
EXPORTS	IMPORTS	ENCRYPTED	REAL
INCLUDES	MIN	MAX	SIZE
FROM	WITH	COMPONENT	PRESENT
ABSENT	DEFINED	PLUS-INFINITY	MINUS-INFINITY

(3) 생성 규칙

생성에 의하여 ASN.1 sequence의 새로운(더 복잡한) 집합이 정의된다. 이 집합은 본 권고안에서 정의한 이름을 사용하고, 다음 두 가지중 한 방법에 의하여 새로운 sequence의 집단을 형성한다

가. 원래의 집단에 포함된 임의의 sequence로 구성된다.
나. 각 집단으로 부터 하나씩 발생할 수 있는 sequence중 임의의 sequence를 지정된 순서에 맞추어 놓는다.

4. Basic Encoding rule

Basic Encoding rule은 X.208에서 지정된 표기법을 사용하여 정의된 type들의 값을 위한 전송구문인 octet stream ODIF를 만들어 내는 규칙으로서 CCITT의 X.209와 ISO 8825에서 정의되어 있다.

4.1 엔코딩 규칙

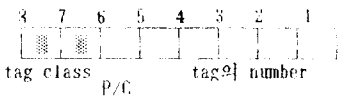
Octet stream ODIF는 다음과 같은 순서와 같은 4개의 구성요소로 된다.

- Identifier octets
- Length octets
- Contents octets
- EOC(End of Contents)

(1) Identifier octet

Identifier octet는 문서에 대한 tag class, tag number, 구조 비트로 이루어졌으며, tag class는 비트8, 비트7에 의해 표현되는데,

- 00 : universal tag class
- 01 : application tag class
- 10 : context specific tag class
- 11 : private tag class 이다. 비트6은 구조비트로서 0이면 단순형, 1이면 구조형의 문서구조를 의미한다. 비트5 - 비트1은 tag number를 표현하며 0는 EOC를 의미하고 1 - 30은 tag number이다.



0 : Primitive(simple)

1 : Constructed(structured)

(2) Length octet

현재의 tag가 영향을 주는 content octet를 나타낼 octet 수를 나타낸다. 구조비트가 0인 단순형 문서구조의 경우에는 내용정보의 길이가 된다. 구조비트가 1인 구조형일때는 종속된 tag에 대한 identifier, length, 내용들의 길이가 된다. Length octet의 형식에는 short form, long form, indefinite form이 있다. Short form은 길이가 1부터 127까지 일때 사용되고, long form은 128이상의 길이를 표현하는 방법이다. 그리고 indefinite form은 길이를 명시하지 않고, 내용의 끝에 hexa값으로 0000H의 데이타를 주는 방법이다.

(3) 내용 octet의 엔코딩

0 또는 하나 이상의 octet로 구성되며 데이타값을 엔코드하는데 이것은 type에 의존한다. 각 type에 대한 데이타값을 엔코딩하는 방법에는 다음과 같은 것들이 있다.

.Boolean 값의 엔코딩 .Integer 값의 엔코딩
.Enumerated 값의 엔코딩 .Real 값의 엔코딩
.Bitstring 값의 엔코딩 .Octet String 값의 엔코딩
.Null 값의 엔코딩 .Sequence 값의 엔코딩
.Sequence-of 값의 엔코딩 .Set 값의 엔코딩
.Set-of 값의 엔코딩 .Choice 값의 엔코딩
.Selection 값의 엔코딩 .Tagged 값의 엔코딩
.Object identifier값의 엔코딩 .Encrypted Type값의 엔코딩
.Character String Type 값의 엔코딩

5. 어휘 및 구문 분석

엔코딩 시스템은 개방할때 도와주는 tool로서 어휘분석부 생성 tool인 LEX와 구문분석부 생성 tool인 YACC를 사용한다.

5.1 LEX

LEX는 yylex라는 어휘 분석기 lexical analyzer를 만드는 프로그램으로 LEX의 입력인 LEX 규칙화일 lex.l 은 정의부와 규칙부로 구분되어 작성되며 LEX를 실행시키면 어휘분석기 lex.yy.c 가 생성된다.

본 논문에서는 ASN.1 표기법에 따라서 복잡한 [illegible] 대하여 [illegible] 시작하여 [illegible] 정의된 [illegible] 에 의해 끝나는 [illegible] 의 공간을 구분한다. [illegible] 여 인식되도록 하였다. [illegible] 리를 인식하며 [illegible] (y.tab.c)로 되돌아간다. [illegible] 부분과 C 언어로 [illegible] 로 들어오는 데이타 [illegible] 되어있다.

5.2 YACC

YACC는 yyparse라는 [illegible] 는 프로그램으로, YACC의 입력인 [illegible] 을 선언부와 규칙부분으로 [illegible] YACC를 실행 시키면 [illegible]

YACC의 규칙화일에 [illegible] 한 토큰값이 [illegible] 부분들은 [illegible] 해당 C 언어로 [illegible] 구문분석을 처리하는 [illegible] 위한 구조체가 완성된다. [illegible] 내부의 정보로 C언어로 [illegible]

6. 전체 시스템의 구성도

포맷화된 혼합형 사무용문서 교환을 위한 [illegible] 딩 시스템 구성도는 그림3과 같다.

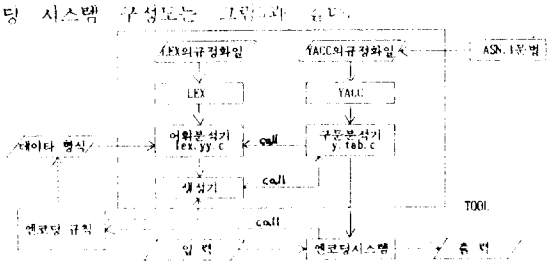


그림3 전체 엔코딩 시스템 구성도

ISO DIS 8613의 데이타 포맷과 CCITT 권고안 T.416, T.501, X.409의 데이타 포맷을 목표로 하는 엔코딩 과함께 코딩하여 lex.yy.c에서 call하여 어휘분석을 [illegible]

도록 하고, CCITT 권고안 X.208에 정의된 ASN.1을 필요한 액션과 함께 추가하여 C 언어로 코딩하여 YACC규정 화일을 얻는다. 여기서 컴파일러 생성도구인 YACC를 실행시켜 어휘분석및 구문분석을 하기위한 오브젝트 화일을 생성시키면, CCITT 권고안 X.209에서 정의한 Basic Encoding rule을 이 오브젝트 화일에 적용하여 C 컴파일러로 컴파일하여 최종적인 엔코딩 시스템을 생성한다. 이 엔코딩 시스템에 ODA로부터 출력되어 전환된 문서의 데이타 스트림을 입력시키면 전송가능한 문서의 binary octet stream이 출력된다.

III. 적용및 고찰

본 논문에서 구성한 개방 시스템간 사무용 문서 교환을 위한 엔코딩 시스템은 문서에 대한 데이타 포맷을 적용하기 위한 abstract syntax module에 ASN.1 notation을 적용하여 abstract syntax를 만든후 이 abstract syntax에 Encoding rule을 다시 적용하고 ODA로부터 출력된 문서의 데이타 스트림을 입력시키면 출력으로서 전송 가능한 presentation syntax(trasnfer syntax)를 생성 시킨다. 이 presentation stream은 연속된 binary stream이다. 그림4에 전송 가능한 문서의 생성과정을 보인다.

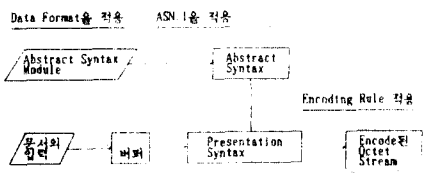


그림4 전송가능한 문서의 생성과정

IV. 결 론

본 논문에서는 신뢰성있고 시스템 독립적인 정보전송을 위해서 문자와 화상정보가 혼합된 사무용문서를 표준화된 프로토콜에 따라 교환될 수 있도록 하기 위한 엔코딩 시스템을 구성하였다.

이 엔코딩 시스템은 문자, 그림, 이미지, 그래픽, 사진 등을 동시에 디스플레이가 가능하고 혼합된 문서의 작성 및 통신이 가능한 단말장치인 혼합형 터미널(mixed mode terminal)의 개발에 기여할 수 있다.

앞으로 ASN.1과 Encoding rule에 대한 나머지 몇개의 유형과 규칙들을 완벽하게 적용하여, 교환하고자 하는 사무용 문서가 문자 텍스트와 라스터 그래픽및 geometric 그래픽을 포함하여 어떠한 형태의 정보를 포함하더라도 시스템 독립적으로 교환할 수 있도록하는 연구가 필요하고 전송되어진 정보의 완벽한 decoding과 이를 표현하는 연구도 병행되어야 할 것이다.

참 고 문 헌

1. ISO-DIS 8613, "Office Document Architecture(ODA) and Interchange Format", June. 1987.
2. CCITT Draft Recommendation X.209, "Specification of Basic Encoding Rules for Abstract Syntax Notation one", Dec. 1987.
3. CCITT Draft Recommendation X.208, "Specification of Abstract Syntax Notation One(final version)", Dec. 1987
4. J.A. Zajaczkowski, "An introduction to the CCITT/ISO standard on transfer syntax and notation", Br Telecom Technol J Vol 5, No.4, Oct. 1987.
5. Ian R. Campbell-Grant and Peter J. Robinson, "An introduction to ISO DIS 8613, "Office Document Architecture, and its application to computer graphics", Comput & Graphics Vol 11, No.4, pp325-341, 1987.
6. Marshal T. Rose, "The ISO Development Environment User's Manual", The Wollogong Group, July 1988.
7. Santa Cruz Operation, Inc. The XENIX-V Development System Programmer's Guide System Development Tools, Oct. 1985.