

Office Document Architecture

鄭會京, 李壽淵

光云大學校 電子計算機工學科

I. 서 론

정보화 시대로 되어감에 따라 컴퓨터의 보급 및 이용이 늘어가면서 컴퓨터, 워크스테이션(workstation) 혹은 워드프로세서(wordprocessor) 등에서 사무문서의 작성이 행해지고 있다. 그러나 문서는 문자뿐만 아니라 도형정보, 화상정보도 포함하고 있어 이들 문자, 도형, 화상등의 문서표현 미디어와 문서의 논리 및 배치구조를 체계적으로 규약하며, 네트워크의 발전에 따라 이 기종간의 효율적인 문서 교환을 실현시킬 목적으로 문서의 교환형식을 국제 표준화하는 것이 절실히 요구되고 있다. 이에따라 국제 표준화로서 규정한 것이 ODA(office (open) document architecture : 사무문서 체계)이다.

ODA의 시작은 1982년에 ECMA TC 29의 2 번째 모임에서 논의되기 시작하여 ODA 표준인 ECMA-101이 1985년에 나왔다(ECMA는 European Computer Manufactures Association이며 TC 29에서는 사무용 문서에서 사용되는 텍스트의 작성, 처리, 교환 분야에서 표준안을 만들기 위해 설립된 기술 위원회이다). 한편 텔레마틱 서비스를 연구하는 CCITT(International Telegraph and Telephone Consultative Committee) SG VII은 팩시밀리 텔레텍스를 통합한 혼합 모드(mixed-mode : 권고 T. 72)와 통신을 위해 문서표현과 통신방법을 규약하는 문서교환 프로파일(profile : 권고 T. 73)을 1984년에 권고 하였으며, 비데오텍스(videotex)와 같은 회화형 서비스등 텔레마틱 전반에 적용하기 위해 문서교환 프로토콜(protocol)을 특정 서비스에 의존하지 않게 구성하고, 수신단에서 재처리를 가능하게 하는 논리구조를 도입하여 도형정보등 문서표현 미디어를 표현 가능하게

하는 것을 SG VII회의에서 합의하였다. 또한 ISO의 ODA 표준과 상호 호환성을 갖고 정합하기 위해 권고 T. 73을 권고 T. 400시리즈로 표준화를 진행하였다. 한편 텍스트와 사무용 시스템(text and office system)을 연구하는 ISO(International Organization for Standardization) TC97/SC18에서는 문서의 논리구조와 문서표현 미디어의 체계화 연구가 진행되어 문서구조와 교환형식(ISO 8613)이 초안화 되었다. 이렇게 정보처리 즉 텍스트와 사무용 시스템에 인식을 같이하여 공동 연구한 결과 1988년에 국제표준(IS : International Standards)으로 ODA의 표준안이 나오게 되었다. 국제표준인 ODA는 문서구조, 내용구조 및 교환양식을 기술하기 위한 기본 표준안(basic standard)으로써 수정가능한 형태(processable form)의 문서 및 포맷된(formatted) 수정가능한 형태를 교환할 수 있도록 하는 것이며 문서표현 매체로는 문자, 화상, 도형과 같은 여러가지 형태의 정보를 가지고 있는 문서를 교환할 수 있게 하는 것이다. 그러나 이후의 ODA 표준은 문자, 화상, 도형등의 정적인 미디어와 동시에 모화, 음성, 애니메이션(animation)등의 시변요소를 동반하는 동적 미디어와 수치 데이터 등의 처리 미디어를 통합적으로 표현하는 구조로 발전하는 것으로 예상되며 이에 대한 연구가 진행중이다. 이런 의미에서 문서체계는 문서표현을 위한 체계에서 미디어의 물리적 및 논리적인 구조면에서 통합적으로 정보를 표현하는 전자문서 체계로 발전하는 것이 고려되고 있으며, 여기에 통신 프로토콜을 조합함에 의해 먼거리 시스템에서 액세스(access)하고 문서의 저장, 검색 서비스와 분산된 정보요소의 결합 공용화에 의한 멀티-미디어(multi-media) DB

(data base)의 상호이용 서비스에도 적용가능한 것으로 생각된다. 또한 계산기와 프린터간에서 교환되는 인쇄 서식의 기술언어로의 적용등 전자인쇄 표준으로의 적용도 예상되 이에 따라 멀티-미디어 통합을 지향한 기능확장 연구가 진행중에 있고, ODA를 기본 표준으로 문서구조, 미디어, 통신기능의 사용규정을 규약한 응용 프로파일의 표준개발이 활발히 진행되고 있으며, OSI의 보급을 추진하는 기관 단체등에서 ODA를 기본 표준으로한 기능표준(functional standards)의 연구개발이 진행되고 있다.

본고에서는 2장에서 ODA의 전반적인 개요에 대해 ODA의 문서체계 및 문서처리 모델을 중심으로한 처리방식과 내용체계에 대해 설명하고, 3장에서는 현재 진행중인 ODA에 대한 국제 동향과 국내 연구활동에 관해 설명하기로 한다.

II. ODA의 개요

본장에서는 ODA 표준의 기본이 되는 구조화 개념 및 문서의 유형, 문서 구성요소의 공통화와 문서처리 모델에 따른 문서처리를 설명하고 내용체계와 데이터 구조에 대해 설명한다.

1. 논리구조와 배치구조

문서체계의 기본개념은 구조화이다. 그림 1에 표시한 것처럼 문서구조는 트리구조의 계층구조로 표현되고 마디(node)는 대상(object)이라 하고 가지(branch)는 대상의 종속대상으로의 분할을 표시한다. 현재 문서체계 표준은 논리구조와 배치구조로 불리는 2개의 구조표현이 가능하다. 논리구조는 문서를 대상의 의미(예: 장, 절, 그림, 단락, 표제, 각주등)에 기본적인 계층구조화이다. 한편 배치구조는 문서를 대상의 물리적 구성(예: 페이지, 프레임, 블록)에 기본적인 계층구조화이다.

대상에는 기본대상과 복합대상이 있다. 분할되지 않는 최소의 대상이 기본대상이고 이외를 복합대상이라 한다. 각 구조의 근에는 문서배치근과 문서논리근이 있다. 배치대상은 기본대상인 블록과 페이지가 있고 복합대상에는 프레임, 페이지, 페이지 집합이 있다. 페이지는 기본대상과 복합대상의 쌍방의 성질이 있다. 이것은 팩시밀리와 텔레텍스와 같은 단일 매디어를 취급하는 경우 페이지가 직접 데이터를 종속하는 것이 가능한 것처럼 하기 때문이다. 페이지 집합을 제외한 각 배치대상은 표시 매체상의 직사각형 영역에 대응한다. 논리대상은 기본대상과 복합대상

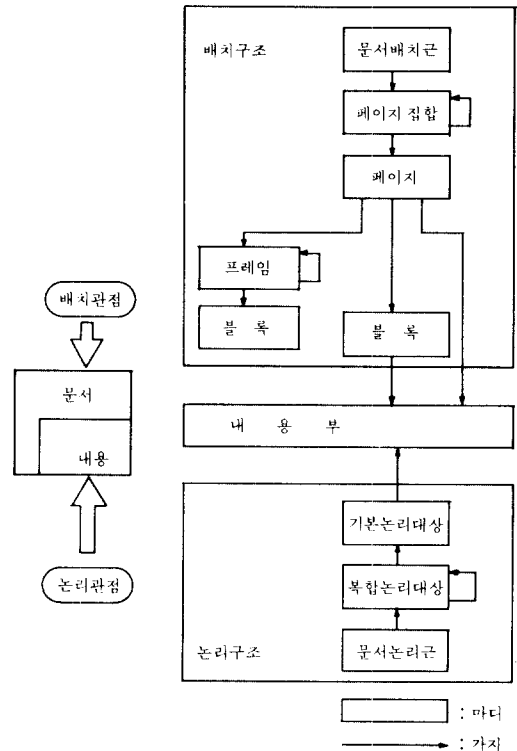


그림 1. 논리구조와 배치구조

만이 정의되고 있다. 장, 절, 단락등의 특정 논리대상 종류는 언어체계등에 의존하기 때문에 문서체계 기본 표준에는 규약하지 않고 응용에서 정의하는 것으로 했다. 따라서 응용 프로파일에서는 특정의 논리대상 종류와 이들 계층관계를 규약하고 시스템 상호간에서 동일하게 해석 가능하다. 논리구조와 배치구조는 기본적으로는 독립적이며 문서의 논리구조는 편집 처리에 의해서 되며 배치처리는 배치처리에 의해 결정된다. 논리구조로부터 되는 문서를 작성자의 의도에 의해 자동적으로 배치하기 위해 배치 지시속성이 정의되고 이 배치 지시속성은 논리구조와 관련됨에 따라 배치처리를 제어하는 것이 가능하다.

ODA에서는 논리구조와 배치구조를 사용하여 응용에 따라 3가지 유형 즉 포맷된 형식, 수정가능한 형식, 포맷된 수정가능한 형식으로 문서를 표현하고 있다.

포맷된 형식은 문서가 작성자의 의도대로 프린트

또는 디스플레이 되도록 하는 형식으로 문서 배치와 표현에 관한 정보를 지정한다. 이의 문서는 제어 정보가 문서내에 나타나 있지 않기 때문에 수정 및 변경이 어렵다. 수정가능한 형식은 논리 구조를 지정하고 있기 때문에 문서가 전달되었을 때 수신자가 작성자의 의도에 따라 편집등의 처리를 할 수 있는 형식이며 이는 문서 프로파일, 특정 논리구조, 공통 논리구조 및 공통 배치구조를 지정한다. 이 문서는 쉽게 수정 변경될 수 있으며 실제 배치는 문서가 포맷될 때만 결정된다. 따라서 제어 정보는 문서 내용과는 독립적으로 변경할 수 있다. 포맷된 수정 가능한 형식은 편집할 수도 있고 그 일부를 다른 문서에서 사용할 수도 있으며 이 형식은 문서 프로파일, 특정 논리구조와 특정 배치구조, 공통 논리구조 및 (또는) 공통배치 구조를 포함한다.

2. 속성

속성(attribute)은 문서 특성, 논리대상, 배치상대 및 내용부의 특성을 기술하기 위해 사용된다. 문서를 전체로서 취급하기 때문에 문서 특성에 관한 속성 집합은 문서 프로파일이라 한다. 문서 프로파일은 작성일자, 작자, 표제, 판수등의 문서관리 정보와 문서 본체의 구성, 사용한 미디어등의 문서 특성을 기술하는 것에 사용한다. 논리대상 및 배치대상을 기술하는 속성은 대상의 형에 의존한다. 즉 기본 논리대상, 복합 논리대상, 페이지 집합, 페이지, 프레임, 블록에 대해서 다른 속성이 적용된다. 이들 대상을 기술하는 속성은 문서구조 속성이라 하고 각 대상으로의 적용 규칙이 규약되고 있다. 문서구조 속성에 추가로 기본 논리대상과 기본 배치대상에 적용되는 속성이 있고 이것을 표현속성(presentation attributes)이라 한다. 표시속성은 대상에 적용되는 문자, 화상등 내용의 형에 의존하고 다른 표시속성이 정의된다. 문서구조 속성은 대상에 적용되는 내용의 형과는 독립이다.

3. 특정(specific)구조와 공통(generic)구조

동일 성질을 갖는 대상 및 문서는 대상부류(class) 및 문서부류라 한다. 문서부류는 보고서, 의사록, 논문등과 같은 유사한 형식을 갖는 문서에 대해서 문서의 논리구성과 양식등의 배치구성과 같은 문서에 공통의 특성을 기술한다. 대상부류는 표제, 본문, 장, 절, 각주등의 문서 구성요소의 공통특성을 기술한다. 논리대상 부류와 배치대상 부류의 구조적 관련은 공통 논리구조와 공통 배치구조라 한다. 이것에 대해

대상으로 하는 문서에 대해서 고유의 구조를 특정 논리구조와 특정 배치구조라 한다. 그림 2에 대상부류와 문서부류를 사용한 문서 처리 예를 나타낸다. 문서등의 정형적인 문서작성에는 논리적인 공통 문서로서 기술하고 이것을 사용해서 내용작성, 배치처리를 행하는 것이 가능하다. 그림 2에서 문서는 표제, 본문으로 구성되고 더우기 본문은 2개의 단락으로 되는 구성을 문서부류에 의해 기술한다. 다음에 문서부류, 배치대상 부류에 준해 표제, 본문의 내용작성과 이들 내용의 페이지상에 배치를 행하는 배치처리를 행한다. 이때 표제, 본문등의 배치구성과 표제부의 내용을 공통적으로 쓰이는 것이 가능한 경우 배치대상 부류로서 참조한다. 문서부류에 기준해 내용작성, 배치처리를 행한 결과로서 특정 배치구조가 생긴다. 본 예에서 보고서는 2개의 페이지로 구성되고 각각의 페이지는 2개의 블록으로 구성된다. 제1 페이지의 최초의 블록에는 표제부가 다음 블록에는 본문을 구성하는 최초 단락의 일부가 할당된다. 제2 페이지에는 최초 단락의 나머지와 이어지는 단락이 할당된다. 문서 교환을 행하기 위해 문서는 기술자와 텍

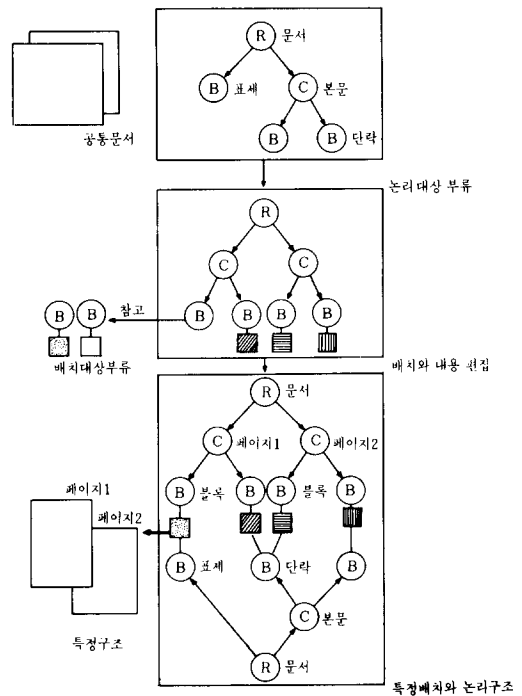


그림 2. 대상부류와 문서부류에 의한 문서처리 예

스트(text)단위라는 데이터구조에 의해 표현된다. 기술자는 문서 프로파일, 대상, 대상부류를 기술하고 텍스트 단위는 내용부를 기술한다. 논리구조 및 배치구조의 경우 대상간의 계층관계는 대상 식별자, 종속대상으로의 참조, 내용부 식별자, 종속 내용부로의 참조에 의해서 표현한다.

4. 문서처리 모델

문서처리 모델의 목적은 주처리 즉 편집, 배치, 표시(imaging)처리 및 이들 처리에서 문서의 다양한 요소의 역할을 기술하는 것이다. 여기서는 세가지 처리에 의해 문서가 어떻게 처리되는가를 개념적으로 기술한다. 그림 3에 이들 관계를 보여주고 있다.

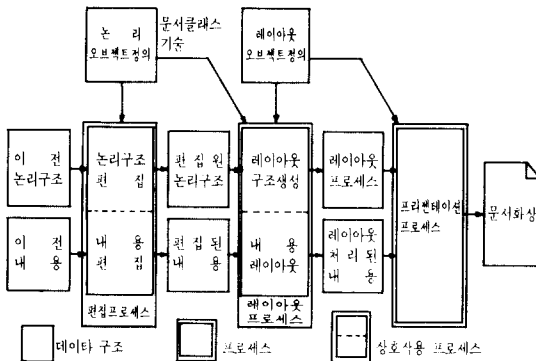


그림 3. 문서처리 모델

1) 편집처리

편집처리에서는 내용편집과 논리구조 편집 처리를 포함하고 있다. 내용편집은 새로운 내용을 만들거나 이미 만들어져 있는 내용을 수정, 변경하는 처리이며, 논리구조 편집은 특정 논리구조의 생성 또는 이미 생성된 특정 논리구조를 변경하기도 하며, 기본 논리대상에 내용을 연결시키는 것등을 포함하고 있다. 편집처리 결과는 배치처리에 입력된다.

(1) 공통 논리구조

공통 논리구조는 편집처리에서 특정 논리구조를 생성하기 위하여 사용되는 상위개념의 구조로써 예를 들면 편지류, 보고서류와 같은 유사한 속성을 갖는

문서부류를 정의하고 있다.

(2) 특정 논리구조

입력, 수정 처리중의 문서부류에 대응하는 공통 논리구조에 따라 현재 처리중인 문서를 위한 특정 논리구조를 생성, 변경하는 것으로 논리대상의 생성 및 삭제와 특정 논리구조내 대상의 위치를 변경하고 대상기술과 관련된 속성을 추가, 삭제, 변경함으로써 대상의 특성을 변경한다.

(3) 내용

내용을 편집하기 위해서는 내용자체가 수정가능한 형식 또는 포맷된 수정가능한 형식이어야 한다. 변경내용은 다음과 같다.

- 하나 또는 그 이상의 내용요소를 추가, 삭제, 변경한다.
- 문서내용에 포함된 제어함수(embedded control function)를 추가, 삭제, 변경한다.

(4) 공통 배치구조, 특정 배치구조, 배치양식(layout style), 표현양식(presentation style)은 편집처리에서 직접적인 역할을 하지 않는다.

2) 배치처리

배치처리는 문서배치 처리와 문서내용 배치처리로 이루어진다. 문서배치 처리는 문서내용이 배치될 특정 배치구조의 생성에 관한 것이다. 이것은 공통 배치구조와 배치방침(directive)이라는 속성집합에 의해서 제어된다.

문서내용 처리는 표현양식이라는 속성을 이용하여 문서 배치구조내에 그래픽 요소(테스트, 화상, 도형)를 배치시키는 과정이며, 문자만을 대상으로 한 경우 이를 포맷팅 처리라고 한다. 문서배치 처리는 페이지 집합, 페이지, 프레임등과 같은 배치대상을 만들기 위한 것이고 문서내용 배치처리는 페이지나 프레임속에 내용이 배치될 영역 즉 블록을 만드는 것이다. 배치처리의 결과는 표시처리(imaging process)에 입력된다.

(1) 공통 논리구조

배치처리 동안에 방침의 값과 공통 내용부를 제공한다.

(2) 특정 논리구조

특정 논리구조내에 있는 논리대상들의 순서에 의해서 문서내용이 결정된다.

(3) 배치 양식

문서배치 처리를 지시하는 배치 방침값을 제공한다

(4) 내용

배치처리 동안 내용은 블록등과 같은 기본 배치대

상에 할당되고 문서내용 배치처리가 표시처리를 용이하게 하기 위해 내용에 제어함수를 삽입할 수도 있다.

(5) 공통 배치구조

공통 배치구조와 특정 배치구조의 관계는 공통/특정 논리구조와의 관계와 유사한 것으로 특정 배치구조를 결정하기 위해서는 '완전(complete)'공통 배치구조이어야 한다.

(6) 특정 배치구조

특정 배치구조는 문서배치 처리와 문서내용 배치처리가 특정 논리구조, 내용, 공통 논리구조, 공통 배치구조, 배치 양식, 표현양식에 적용된 결과이다.

3) 표시처리

기본 배치대상의 표현속성과 배치내용에 삽입된 제어함수의 조절하에서 표현매체(종이, 화면)상에 문서 화상을 만들어 내는 것이다. 화상처리는 그 자체가 local 처리로 간주되고 있으므로 구체적인 것은 IS 8613에서 정의되고 있지 않다. 이것은 사용되고 있는 디스플레이 방법에 종속적이기 때문이다.

(1) 내용

사람이 인지할 수 있는 문서 화상을 만드는데 사용한다. 배치후에는 '포맷된 형식 또는 포맷된 수정 가능한 형식'으로 될 수 있다.

(2) 공통 배치구조

공통 배치구조는 내용의 표시를 지시하는 속성과 공동부를 제공

(3) 특정 배치속성

내용과 함께 사람이 인지할 수 있는 문서 화상을 만들기 위해 사용, 특정 배치구조에서의 대상 순서는 문서내용을 표시하는 우선순위를 결정한다.

(4) 배치 양식

표시처리 동안 아무런 역할도 하지 않는다.

(5) 표현양식

표현양식의 표현속성이 내용의 표시 생성에 영향을 미친다.

5. 내용체계

내용체계는 그래픽 요소, 제어함수, 내용과 관련된 부호화 방법, 표현속성으로 구성된다. 내용 체계에 있어서 중요한 것은 첫째, 배치대상에 있는 그래픽 요소를 위치지정(positioning)하는 규칙을 정의하고 두 번째는 수정가능한 형식이 어떻게 포맷된 형식으로 변환되는가를 기술한다. 그래픽 요소는 다음과 같다.

a) 문자요소 : ISO 6937과 CCITT 권고 T. 61에서

정의된 8비트 부호화 환경과 그래픽 문자에 기초한 문자 텍스트이다.

b) 화상요소 : CCITT 권고 T. 4, T. 5, T. 6에서 정의된 대로 픽셀밀리 방법으로 만들어지는 2차원 scanned image이다.

c) 도형요소 : ISO 8632에 따라 정의되고 부호화되는 그림기술 요소이다.

이들을 처리 중심으로 설명하고자 한다.

1) 문자내용체계(character content architecture)

(1) 표현속성과 제어문자

내용체계는 기본 대상의 표현속성, 제어함수, 사용되는 그래픽 문자를 정의한다. 문자내용은 그래픽 문자, space 문자, 제어문자이다. 표 1과 표 2에 표현속성과 제어함수를 각각 나타내었다.

a) 논리표현속성 : 기본 배치대상에만 적용되고 배치처리 동안에만 유효하다.

b) 배치표현속성 : 배치대상에만 적용되고 배치처리 동안 유효하다.

c) 공용(shared)표현속성 : 기본 논리대상과 기본 배치대상 모두에 적용한다.

d) 논리제어함수 : 배치처리에만 적용

e) 배치제어함수 : 배치처리에 의해 생기고 표시처리에 입력된다.

f) 공용제어함수 : 양쪽 처리에 모두 적용

(2) 문자내용 배치처리

문자내용 배치처리는 두개의 단계로 행해진다.

a) 문자경로, 행진행방향, 첫 행 인덴테이션(indentation)은 행길이를 결정하기 위해 사용된다.

표 1. 표현속성

| Shared Attributes | Layout Attributes |
|---|---|
| <ul style="list-style-type: none"> • Alignment • Character Font • Character Orientation • Character Path • Character Spacing • First line format • Graphic character subrepertoire | <ul style="list-style-type: none"> • Alignment indicator • Initial offset |
| <ul style="list-style-type: none"> • Graphic rendition • Kerning rendition • Line layout table • Line progression • Line spacing • Precision | <p>Logical Attributes</p> <ul style="list-style-type: none"> • Indentation • Leading • Orphan size • Pairwise kerning • widow size |

표 2. 제어함수

| Shared Attributes | Layout Attributes |
|---|---|
| <ul style="list-style-type: none"> •CR Carriage Return •GCC Graphic Character Composition •IGS Identify Graphic Subrepertoire •LF Line Feed •PLD Partial Line Down •PLU Partial Line Up •SGR Select Graphic Rendition •SHS Select Graphic Spacing •SPI Spacing Increment •SRS Start Reverse String •STAB Selective Tabulation •SUB Substitute Character •SVS Select Line Spacing •VPB Line Position Backward •VPR Line Position Relative Code Extension Control Function | <ul style="list-style-type: none"> •BS Backspace •HPB Character Position Backward •HPR Character Position Relative •JEY No Justify •SACS Set Additional Character Separation |
| | Logical Control Function |
| | <ul style="list-style-type: none"> •BPH Break Permitted Here •NBH Not Break Here |

b) 문자열 내에서 행 브레이크(break)를 결정한다. 행 브레이크는 논리대상 내용내에 있는 SC(space character), BPH(break permitted here)와 hyphenation 알고리즘의 구현등에 의해 결정된다. 배치처리 결과가 표시 형식일 때 매 행 브레이크에 하드행 종료자(hard line terminator)가 삽입되고 포맷된 수정가능한 형식일 때는 소프트행 종료자(soft line terminator)가 삽입된다. 정렬(alignment)이 지정됨에 따라 맨처음 그래픽 문자의 위치가 제어함수 SHS 또는 SI와 같은 제어함수를 삽입함으로써 결정된다.

(3) 문자내용 표시처리

표시처리는 문자 box의 집합으로 만들어지는 행 box 집합으로써 기본 배치대상의 표시를 형성하고 임의의 폰트(font)로 문자를 표시한다. 여기서의 box는 단지 텍스트의 위치지정을 기술하기 위해 개념적으로 도입한 것이지 실제 배치대상은 아니다. 위치지정은 내부 좌표 시스템에 관련하여 행해진다. 이것의 원점은 배치표현 속성 'initial offset'에 의해서 기본 배치대상의 좌표 시스템과 관련해서 지정되고 이것의 축을 문자경로와 행 진행방향에 의해 결정된다.

라인 box는 일정한 간격을 가지고 행 진행방향으로 기본 배치대상내에서 연속적으로 위치한다. 문자 box는 일정한 간격 또는 비례적인 간격을 갖는 문자경로를 따라서 행 box내에 위치된다. 문자 box의 폭은 표시될 현재의 폰트와 그래픽 문자의해서 결정된다.

2) 라스터 그래픽 내용체계(raster graphic content architecture)

라스터 그래픽 내용체계는 그림을 데이터로 표현함으로써 그림정보의 기술적인 표현을 제공한다. 이것은 CCITT 권고 T.6과 FAX 코딩(coding) 구성과 G4 FAX를 위한 제어 체계에 기초한다. 현재 ODA 라스터 그래픽 내용체계는 monochrome 표시(black/white)로 한정되고 있으나 colour, gray 표현을 위한 작업은 진행중에 있다. 라스터 그래픽 내용체계에서는 다음 3가지의 형식을 제공하고 있다.

- Final 형식 : 라스터 그림의 크기를 정확하게 지정하기 위한 형식이며, 배치구조와 함께 사용되는 형식이다. 이것을 특별히 ODA 포맷된 형식으로 사용된다.
- 고정된 크기(dimension) 형식 : 논리구조와 함께 사용하기 위한 것이다. 이 형식을 라스터 그래픽 내용이 페이지 배치를 갖는 문서내에 포함할 수 있으나 고정된 크기를 가진다. 따라서 수정 가능한 형식에 적용될 수 있다.
- Scaleable 형식 : 논리구조와 함께 사용되나 실제적인 라스터 그래픽 표시의 크기를 제한하고 있지 않고 배치처리가 이를 결정한다. 따라서 라스터 그래픽 내용중 가장 유연성 있는 표현이며 수정가능한 형식에 적용될 수 있다.

(1) 표현속성

- Clipping 표현속성 : 라스터 표시로써 사용될 펠 배열(pel array)의 부분영역을 결정짓는다. 이 속성에 의해 정의된 정방향 영역을 벗어나는 배열상의 모든 펠은 표시되지 않는다.
- 표시제한 표현속성 : 내용을 포함하기 위해 만들어진 블록에 의해 허용할 수 있는 최대 크기를 지정한다.

(2) 내용 부호화 구성

- 비트맵(bitmap) 코딩 : 펠 배열이 비트맵으로 표현된다. 즉 0과 1의 값을 갖는 비트의 2차원 배열로써 0을 background 색조, 1은 foreground 색조로 하게 된다. 이것은 현재 고해상도 비트맵 워크스테이션에서 사용되고 있는 것과 같다.
- T.6 부호화 : 2차원 부호화 구성이며 내용의 압

축된 형식과 압축되지 않은 형식 모두를 제공한다. 이것은 G4 FAX의 효율적인 교환을 위해 CCITT가 권고했던 것이다.

- T. 4 부호화 : CCITT 권고 T. 4에서 정의된 1차 /2차원 부호화 구성을 따른다.

(3) 배치처리

라스터 그래픽 내용체계의 배치처리는 블록이 논리내용을 포함하고 표현 속성상의 제한을 해결하기 위해서는 블록 크기가 어떻게 결정되어야 하는가를 나타낸다. 여기에 사용되는 기본 정보는 다음과 같다.

- 라스터 이미지가 놓여질 수 있는 페이지상 영역
- 내용이 어떻게 clip될 것이며, 펠과 행의 진행방향 그리고 블록에 대한 최대 크기를 나타내는 표현속성
- 내용이 고정크기 내용 유형인지 scalable 크기 유형인지에 대한 정보

(4) 표시처리

라스터 그래픽 내용체계에서 배치처리는 다른 내용 체계에 비해 비교적 간단하다. 라스터 그래픽 내용을 포함하는 각 배치 블록은 펠로 이루어져, 표시될 펠의 값이 주어져 있으므로 표시처리는 해당 색조로 각 펠을 표시하기만 하면 된다.

3) 도형 그래픽 내용체계 (geometric graphic content architecture)

도형 그래픽 내용체계는 line, circle, arc와 같은 도형 요소가 혼합된 도형, 그래픽, 도표, 묘화 등을 ODA 문서에 포함시킬 수 있는 기능을 제공한다. 도형 그래픽 내용체계의 목적은 기존의 그래픽 응용들과 동일한 방법으로 기능을 제공하는 것이며 기존의 그림을 ODA 문서에 통합시킬 수 있게 하는 것이다. 이것은 ISO IS 8632-CGM (computer graphics metafile)에 기초한 것이다. CGM은 그림기술 정보의 저장, 검색, 교환에 적합한 형태를 갖추고 있다. 도형 그래픽 내용체계에서는 그림의 내용이 되는 도형 요소가 CGM 표준규칙에 따라 가상장치 좌표 시스템 (VDC : virtual device coordinate system)이라는 좌표로 위치 지정된다. 그래서 내용체계에서는 이러한 내용 위치지정의 좌표 시스템을 ODA 배치구조에 의해 설정되는 영역과 어떻게 관계지을 것인지를 명시해야 한다. 내용의 CGM 코드된 형식은 가상장치 좌표범위 (extent)라는 요소를 가지고 있는데 이것은 무한한 가상 공간에 그림영역을 지정하는 것이다. 또 그림 방향이라는 속성이 있어서 VDC 범위영역이 내용을 포함하는 블록에 어떻게 매핑될 것인가를 지정

한다. 그림 3에 VDC 범위와 ODA 블록간의 관계를 도시하였다. VDC 범위의 좌측 하단 좌표 (0, 0)가 블록의 좌측 하단과 매핑되고 있다.

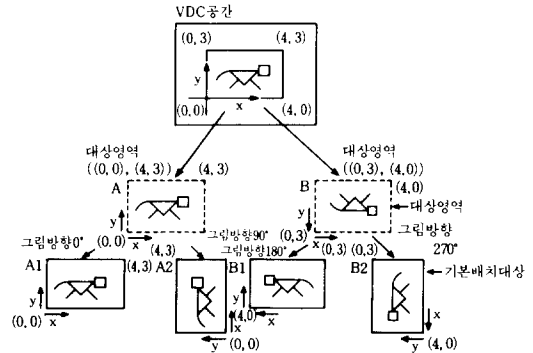


그림 4. VDC 범위와 ODA간의 관계

(1) 표현속성

도형 그래픽 표현속성은 도형 그래픽 내용을 포함하고 있는 논리대상과 배치대상을 배치하고 표시하기 위한 초기 조건을 지정한다. 이것은 내용에 표현될 수 있는 CGM 요소의 속성과도 일치한다. 도형 그래픽에는 'metafile default 속성'과 'geometric coding announcer'라고 하는 두개의 표현속성이 있다. 전자는 배치와 표시처리에 사용될 정보를 저장하기 위해 사용되며 파라메타 집합으로 구성된다. 후자는 부호화된 metafile defaults 속성을 번역하는데 필요한 정보를 지정한다. 이것은 실수와 정수값의 precision과도 같은 정보이다.

(2) 그림 크기

도형 그래픽 내용체계에서는 배치구조에 논리 도형 그림의 크기를 제한하는 논리대상 속성을 정의하고 있는데 이것을 그림크기 속성이라고 한다. 이 속성을 사용함으로써 논리 그림은 표현 매체에 배치될 때 폭과 높이가 실제 크기의 범주에 들 수 있도록 지정할 수 있다. 그림크기 속성은 수정가능한 문서에 포함될 수 있는 그림에 대해, 특히 공통 배치규칙이 수신자에 의해 지정될 수 있을 때 필요하게 된다.

(3) 배치처리

배치처리는 특정 배치를 하기 위해 공통 배치구조

와 논리구조의 속성이 어떻게 결합되어야 하는가, 즉 내용을 갖는 블록의 위치와 크기를 어떻게 결정 지을 것인가를 구체화 시킨다. 도형 그래픽 내용체계에서는 배치처리를 위해 다음과 같은 정보를 사용한다.

- 논리속성의 값, 그림크기, 그림방향, metafile default replacement
- VDC 범위 값
- 배치경로 값
- 이들 정보로부터 배치처리는 블록의 크기와 이것의 위치 그리고 블록내에 있는 그림의 방향을 결정한다.

(4) 표시처리

표시처리는 배치구조에 관한 정보와 내용부의 그래픽 및 제어요소가 최종적으로 보이는 이미지를 만들기 위해 어떻게 사용될 것인가를 지정한다. 또 내용의 블록이 어떻게 표시될 것인가, 특히 같은 영역에 중첩되는 내용을 결합하기 위해서 배치속성 투명도를 어떻게 번역할지에 대해서도 지정하고 있다.

6. ODIF (Office Document Interchange Format)

ODIF는 IS 8613에 따라 구조화된 문서를 교환하기 위해 사용하는 데이터 스트림(stream)의 형식을 정의하고 구조화된 문서의 대상, 내용, 양식의 부호화를 정의하며 교환된 문서에서 나타낼 수 있는 성분(속성들의 집합)의 표현을 정의한다.

(1) 데이터 스트림의 구조

데이터 스트림인 일련의 기술자와 텍스트 유니트(unit)로 구성되어 있다. 기술자는 문서 프로파일, 대상부류, 대상의 속성을 나타내는 기본 데이터 항목과 서브데이터 항목으로 이루어지는 복합적인 데이터 항목으로써 다음과 같은 것이 있다.

- a) 문서 프로파일 기술자
- b) 배치대상 기술자
- c) 배치대상부류 기술자
- d) 논리대상 기술자
- e) 논리대상부류 기술자
- f) 표현양식 기술자
- g) 배치양식 기술자

이러한 구조를 교환 데이터 요소라 하며 이들은 지정된 규칙에 따라 순서가 정해지는데 여기에는 두가지 형식이 있으며 주어진 데이터 스트림에 어떤 규칙이 적용되는가는 문서 프로파일 기술자에 의해 정해진다. 텍스트 유니트는 내용부를 나타내는데 이것

의 속성을 표현하는 기본 데이터 항목과 서브데이터 항목 그리고 내용부의 그래픽 요소를 표현하는 하나 이상의 기본 데이터 항목 집합으로 구성된다. 텍스트 유니트는 교환 형식에 따라 데이터 스트림상에 나타나는 순서가 다르며 교환 형식은 다음과 같다.

a) 데이터 스트림 A : 기술자 단위에서의 전송이 가능하고 화일전송에 사용된다.

- 문서 프로파일 기술자
- 배치대상부류 기술자
- 논리대상부류 기술자
- 공통 내용부에 대한 텍스트 유니트
- 표현양식 기술자
- 배치양식 기술자
- 배치대상 기술자
- 논리대상 기술자
- 특정 내용부에 대한 텍스트 유니트

b) 데이터 스트림 B : 특정 논리구조나 일반 논리구조를 갖지 않는 문서 즉, 포맷된 문서구조에 맞는 문서를 표현하는데 이용된다. 이는 대상과 텍스트 단위가 관련되어 전송되며 혼합모드와 같은 실시간에서 지면으로의 인쇄 가능한 단위로의 전송에 적당하다.

- 문서 프로파일 기술자
- 배치대상부류 기술자와 관련 텍스트 유니트
- 표현양식 기술자
- 배치대상 기술자와 관련 텍스트 유니트

(2) 데이터 스트림 형식의 형식지정

형식지정은 IS 8824에서 정의된 ASN.1(abstract syntax notation 1)에 기초했으며 CCITT 권고 X.409에서 지정된 기본적인 개념과 같은 내용이다. Syntax 표기는 데이터 흐름의 포맷과 이것의 성분을 정의하기 위하여 사용되는 집합 또는 시퀀스(sequence)로 정의한다.

(3) Coded 표현

코딩은 IS 8825에서 정의된 ANS.1에 대한 부호화 규칙에 따라 행해진다. 최종적인 문서표현은 octet 시퀀스이다. 기술자나 텍스트 유니트를 구성하는 각각의 데이터 항목에 대한 coded 표현은 3개의 필드 즉, 식별자, 길이, 값으로 구성된다. 관련 데이터 항목이 복합데이터의 경우, 식별자 필드는 데이터 항목에 대응하는 속성 또는 그 그룹을 나타내고 길이 필드는 값 필드의 길이를 지정하며 값 필드는 하나 이상의 triplet로 구성된다. 이 때 각각의 triplet은 3개의 필드로 구성된다.

7. 통신기능

문서체계 표준은 OSI 참조 모델 제7층의 응용 프로토콜이다. 이하에서는 CCITT에서 연구되고 있는 문서체계 표준에서 표현된 문서를 어떻게 전송하는지를 말하는 통신 프로토콜 체계 표준화 동향에 대해 서술한다.

(1) 텔레마틱 서비스

CCITT SG VIII에서는 텔레마틱 서비스를 위한 문서의 통합적인 통신 조작환경을 계획하기 위해 end-end 통신서비스, center-end 통신서비스에 공통적으로 적용 가능한 프로토콜의 연구를 DTAM(data transfer access and manipulation)에서 진행하고 있다. 문서체계 표준을 쓰는 end-end 통신서비스에서는 G4와 혼합모드 같은 문서의 bulk 전송응용과 문서회의 같이 이미 전송된 문서로의 구조 조작에 의한 문서편집, 검색, 표시등의 회화형 응용이 있다. 한편 center-end 통신서비스에 대해서도 먼거리 시스템으로의 문서의 저장, 검색, MHS(message handling system) 등으로의 적용등이 bulk 전송응용과 비데오텍스, 멀티-미디어 DB로의 액세스 같은 문서 구성요소를 부분적으로 전송하기도 하고 갱신, 소거등을 행하는 회화형 응용이 있다. 이들 텔레마틱 서비스 전반에 적용하기 위한 통신기능으로서 이하의 기능이 연구과제이다. 문서의 bulk 전송기능, 통신 프로토콜을 넣은 문서의 내용구조를 조작하는 먼거리 문서구조 조작기능, 먼거리 시스템에 어느 문서 파일로의 액세스 기능등이다.

(2) DTAM 서비스와 프로토콜

DTAM은 OSI 참조 모델의 응용층에 위치하며 DTAM의 프로토콜 모델은 DTAM 서비스 요소, 응용 제어 서비스 요소로 구성된다. DTAM 서비스 요소는 문서체계 표준에서 표현된 문서구조의 생성, 해석을 행함과 동시에 문서의 bulk 전송구조 조작, 관련 제어등을 실현하기 위한 통신기능을 사용자에게 제공한다. DTAM이 제공하는 기능 단위의 조합에 의해서 텔레마틱 서비스를 위한 통신기능을 실현 가능한 것처럼 범용적인 응용층으로의 개발이 진행되고 있으며 기능단위의 조합을 서비스 부류라 하고 이것이 한개의 통신서비스를 제공한다.

Ⅲ. 국제동향과 국내 연구활동

이 장에서는 현재 진행중인 문서 응용 프로파일(DAP : document application profile)과 여러나라에서

ODA 실현 노력들 및 ODA기능 확장에 대해 설명하고 국내 연구현황에 대해 설명한다.

1. 문서 응용 프로파일

DAP은 ISO 8613이 규정하고 있는 방대한 속성 및 구성요소등을 송신단 및 수신단의 능력을 고려하여 몇가지 기능적인 단계를 정하여 이 단계에 따라 ISO 8613의 속성 및 구성요소의 선택에 제한을 가하여 작성된 부분 집합을 뜻한다. ODA 기본 표준에 대해서 서비스에 대응하여 문서구조, 미디어, 통신기능의 사용규정을 규약한 응용 프로파일의 표준 개발이 여러 나라에서 진행되고 있다. 또한 OSI의 보급을 추진하는 기관, 단체등에서 ODA을 기본 표준으로한 기능 표준의 연구개발이 다음과 같이 응용 프로파일 개발과 같이 진행되고 있다. ODA 기능 표준 연구결과를 각 조직에서 발표된 것을 살펴보면 표 3 과 같다.

표 3. 여러 나라의 ODA 기능 표준

| LEVEL/Area | AOW | EWOS | NIST | core set |
|------------|------------------|------|----------|----------|
| 1 | AE1111 AE1116 | Q111 | | core11 |
| 2 | | Q112 | | core26 |
| 3 | AE1126 | Q113 | NIST DAP | core36 |

표3에서 각각 AOW(Asia-Oceania Workshop), EWOS(European Workshop for Open Systems), NIST(National Institute of Standards and Technology:미 국립 표준국)에서 여러 프로파일을 발표했는데 이들 프로파일들을 국제적으로 조화된 ODA 프로파일을 생성하고 전세계에서 문서를 교환하기 위해 'core set'을 제안하여 PAGODA(Profile Alignment Group on ODA) 모임을 만들어 1990년에 ISO 국제 표준 프로파일을 발표할 목표로 진행중이다. PAGODA회의에는 각 지역을 대표해서 EWOS, AOW, NIST 워크숍에서 참여하며 우리나라에서는 우리나라를 대표하여 ETRI에서 참여하고 있다. 주요 논의대상은 언어와 문화에 의존적인 면이 되고 있다. Core set은 core 11, 26, 36으로 나누어서 작업을 진행하고 있는데 이들 각각은 계층적 트리구조로 이루어져 있어 core 26은 core 11를 포함하고 core 36은 core 26을 포함한다. 각 core set은 다음과 같다.

- Core 11 : 기존의 wordprocessing system간의 문서 교환을 지원하고 내용에서 문자만 지원하는 small function set로 AOW를 중심으로 논의되고 있다.

논리구조 : 단락, 단락 그룹으로 이루어 지고 각 그룹에 대해 다른 배치와 표현 특성 적용 가능
배치구조 : 한 컬럼(column) 문자만 지원하며 각 페이지에 선두어(header), 후미어/footer) 내용 포함

- Core 26 : Advanced 문서처리 시스템간의 멀티-미디어 문서 교환을 지원하며 문자, 라스터 그래픽, 도형 그래픽 내용을 처리하고 core 11에서 제공되는 모든 특성을 제공한다. Core 11과 core 36의 중간 기능을 갖추며 EWOS를 중심으로 논의되고 있다.

논리구조 : Chapter, section의 자동번호 부여, 각주의 자동번호 부여, 문서 내용에서의 각주 참조 등

배치구조 : 멀티컬럼지원, 라스트 또는 도형 그래픽 내용으로 구성된 도표의 자동번호 부여

- Core 36 : DTP (desktop publishing)용 문서처리 시스템간의 멀티-미디어 문서 교환을 지원하며 문자, 라스터 그래픽, 도형 그래픽 내용을 처리하고 core 26에서 제공되는 모든 특성을 제공한다. 이는 medium function set로 NIST 워크숍을 중심으로 논의되고 있다.

논리구조 : Core 26보다 더많은 논리단계를 포함하고 문서내용에서 chapter, section, figure들의 자동 참조

배치구조 : Core 26에 비해 더 자유로운 배치기능 제공

그러나 우리나라에서 문서를 작성할 때 한글 영문 혼용시 문자폭이 달라 빈번히 사용되는 문자 폭 변환 제어 코드의 양이 많아지고 또한 우리나라에서 현재 많이 쓰이는 문자 폭이 7.5 pitch(160BMU)이지만 이 값이 core 11, core 26에 포함되지 않아 한글 표현이 어려운 실정이다. 기본 표준의 SHS 값에도 160 BMU가 없으므로 국내에서는 SCS를 사용할 수 밖에 없는 실정이다. 이 역시 텍스트 코드의 양을 증가시키는 요인이 된다. 이에 대한 연구가 시급하다 하겠다.

2. ODA 확장에 대한 노력

문서 구성요소가 정적인 미디어에서 동적인 미디어

로 변하며 수치 데이터등의 처리미디어를 통합적으로 표현하고 칼라(colour), DB에서 액세스등 여러가지 고려하여 ODA 표준에 첨가될 새로운 항목들이 계속 연구중에 있다. ODA의 미래 개발을 위한 기본 원칙은 현재 ODA에 기능이 첨가되는 것이며 삭제되지는 않고, 개방 시스템 환경중에서 정의되고, 공통 구조는 특별한 응용에 대해 조정 가능하게 디자인 하고, 다른 표준들과 연관관계를 갖어야 한다는 것이다. 증가되는 기능들을 살펴보면 다음과 같다.

(1) 배치와 표시(layout and imaging)

- Intersection 및 overlaying
- Table 배치
- 배치 스트림 확장 : 리스트(list), 주 텍스트, 인덱스(index) 등이 독립적으로 포맷되도록 다양한 종류 내용을 구별하는 방법을 제공
- 사각형이 아닌 배치
- 배치 내용이 표시되는 실제 영역을 결정하기 위한 마스크(mask) 제공
- 칼라
- Typographical quality : Diacritics, underline 등
- Imposition : Folding, trimming 등
- 그림의 분리 : Tiling
- 변수화한 배치 등

(2) 구조적인 면

- 주석
 - 다중 언어 문서
 - Hypertext 등
- ### (3) 조직적인 면
- 인덱스 생성
 - 상호 참조(cross reference) 등

(4) Interface

- 먼거리 참조, 먼거리 편집
- 다른 응용(EDI, CAD/CAM, 3-D graphics)과의 interface
- Hyphenation, spell check

(5) 기타

- Form과 letter 지원
- 문서에서의 데이터
- Spreadsheet
- Business graphics
- 수학적, 화학적, 음표 표현
- 시간에 의존한 내용(voice, music, 동화, 묘화)
- 비디오텍스트
- Security

- 분산 시스템에서 여러 사용자에 의해 생성, 수정이 가능

이렇게 ISO 8613의 내용이 확장됨에 따라 확장된 문서 처리 모델에 필요하며 이는 현재 ISO 8613-1에서 정의된 모델의 superset이 된다. 여기에 evaluation, resolution, manipulation, interchange와 control 개념이 포함된다.

3. ODA 실현 노력

- EXPRES (Carnegie Mellon University & University of Michigan) : ODA tool kit, translators
- NBS (Electronic publishing laboratory) : EXPRES와 교환 동반자이며 ODA → Postscript translators, Wordperfect ↔ ODA
- McDonnell Douglas Corporation : Interleaf ↔ ODA translators
- PODA (ICL, Bull, Olivetors, Siemens) : ODA tool kit, translators, 응용 프로파일, interchange 데모 (H'87, H'88)
- XEROX : ODA → interpress translator
- AT&T/Bell Labs : ODA 환경에서 작업이 진행중이며 ASN.1 parser 생성기
- QMC interactive system : ODA에 근거한 에디터

4. 국내 활동

국내에서도 ODA의 중요성을 인식하여 OSIA (Open System Interconnection Association; 개방형 컴퓨터 통신 연구회)를 중심으로 활동하고 있으며 EWOS, NIST, AOW와 같은 지역적 워크숍 및 PAGODA 회의에 참여하여 능동적으로 적응하며 한국형 문서 응용 프로파일의 정의를 위해 OSIA TG8을 중심으로 몇몇 연구소, 대학, 학회에서 진행중이지만 행정 전산망 등을 고려할때 ODA의 연구 및 구현에 대한 노력이 매우 절실한 것으로 생각된다.

IV. 결 론


ODA의 표준화 경과, ODA의 개요, ODA의 국내의 활동등에 관해 설명했다. ODA는 문자, 화상, 도형 등에 추가로 음성, 동화, 묘화와 수치 데이터등의 각종 미디어를 통합적으로 표현하는 체계의 개발이 필요하다. 이에 따라 문서 구조표현도 전자문서의 작성, 편집, 검색, 표시등에 알맞는 구조표현의 고도

화 연구가 중요하며 ODA에 의해 기술되는 문서의 통신 기능도 멀티-미디어 문서 DB로의 등록, 검색, 갱신등을 문서 구성요소 단위로 가능하게 하는 프로토콜을 개발하고 먼거리 문서의 검색, 편집, 표시, 인쇄등의 문서구조 조직 프로토콜도 병행해서 연구할 필요가 있다. 이들 기본표준의 발전 및 사용자 요구에 기본적인 서비스에 필요한 미디어와 문서구조 및 통신기능을 조합한 기능표준의 연구개발이 중요하다. 위에서 서술한 것처럼 장래 전망을 본 정보통합 표현 체계의 개발, 프로토콜 개발, 응용 프로파일 개발 및 사용자 인터페이스(in. erface)면을 고려해서 처리하는 기술의 개발을 병행해서 진행함에 따라 다른 시스템간에 자유롭게 정보교환을 가능하게 되는 것으로 생각된다.

외국에서는 이에 대한 투자를 많이 하고 있으며 1990년초에 ODA를 이용한 문서처리 시스템의 시작품이 나올 것으로 예상되고 있다. 그러나 우리나라에서 OSIA를 중심으로한 연구가 진행되고 있지만 행정 전산망등을 고려할 때 ODA 연구 및 구현에 대한 더욱 많은 관심을 기울여야 할 것이다.

参 考 文 献

- [1] "ISO IS 8613 Information Processing Part 1-8", ISO, 1988.
- [2] "ISO IS 8632/1 Computer Graphics," ISO, 1987.
- [3] Karen Bishop, "Graphics," Top Technical Review Committee 1986. 12. 17
- [4] Karen Bishop, "Top Document Interchange Recommendation Top Technical Review Committee," 1986, 11. 26
- [5] Lan R. Campbell-Grant and Peter J. Robinson, "An Introduction to ISO," Document Architecture and Its Application to Computer Graphics," Computer & Graphics, vol. II, no. 4, pp. 325-341, 1987.
- [6] Jon Barrett Kirk Reistroffer, "Designing a Raster Image Processor," BYTE pp. 171-180, 1987. 5
- [7] 이수연, 김차중, "Document Processing," 전자공학회, 컴퓨터 기술, vol. 5, no. 2, 1988. 12
- [8] 이수연, "ODA 표준화 동향," 한국정보과학회, 정보통신기술, vol. 1, no. 1, 1987. 7
- [9] Nobour Sonehrra, "Present and Future of Office Document Architecture," Japan IE 87-96.

- [10] Lan R. Campbell-Grant "Framework for future extensions to ODA," ISO/IEC JTC1/SC18/WG5 N706, 1989. 2
- [11] CCITT Recommendation T. 400, "Document Transfer, Access and Manipulation (DTAM) -General Introduction,"
- [12] A. J.H.M Peels, "Document Architecture and Text 6 Formatting," ACM Trans., Office Information System, vol. 3, no. 3, 1982. 9
- [13] L.J. Zeckendorf, "Informal Introduction to ODA," ISO/TC97/SC18/WG3 N764
- [14] ISO 8824, 8825, "ASN. 1: Abstract Syntax Notation One and Basic Coding Rules" 

筆者紹介



李 壽 淵

1946年 10月 24日生
 1969年 광운대학교 전자통신
 공학과 졸업
 1977年 연세대학 대학원 전자
 공학과 졸업
 1983年 일본 교토대학 정보
 공학과 (박사)

1973年 3月 ~현재 광운대학교 전자계산기
 공학과 교수

주관심분야: 문서화상, Graphics, Multi-media
 정보처리



鄭 會 京

1961年 1月 28日生
 1985年 광운대학교 전자계산기
 공학과 졸업
 1987年 광운대학 대학원 전자
 계산기 공학과 졸업
 (B. S)
 1989年 3月 ~광운대학 대학원
 전자계산기 공학과
 박사과정 재학중

주관심분야: 문서화상, Graphics, Multi-media
 정보처리