

# 情報 交換用 符號의 國際 標準化

권 춘 우

(동양공업전문대학 전산사무자동화과)

## ◀ 차 례 ▶

1. 국제 표준화 기구
2. 정보 교환용 부호의 국제 표준화 동향
3. 맺음말
4. 참고문헌

## 1. 國際 標準化 機構

### (ISO : International Organization for Standardization)

국제 표준화 기구(ISO)는 세계적으로 표준화를 촉진시키고, 상품 및 서비스의 국제 교환을 용이하게 하며, 지적, 과학적, 기술적 그리고 경제적 활동 분야에 있어서 국제간 협력을 위하여 구성된 기구이다.

국제 표준화 기구(ISO)는 자발적이면서 비협정적인 단체로서 참여 국가가 지정하는 단체로 구성되어 있다. 그리고, ISO는 비정부간 기구로서 국제 연합(UN)과 관련 국제 연합 기관 및 전문 기관에의 자문적 지위를 가지고 있다. 그리고, 국제 표준화 기구 회원국의 구성을 살펴보면, 회원국의 약 1/2은 정부 기관 또는 공법으로 규정된 법인 조직이며, 나머지 1/2은 공공의 행정 기관과 밀접하게 연결되어 있다.

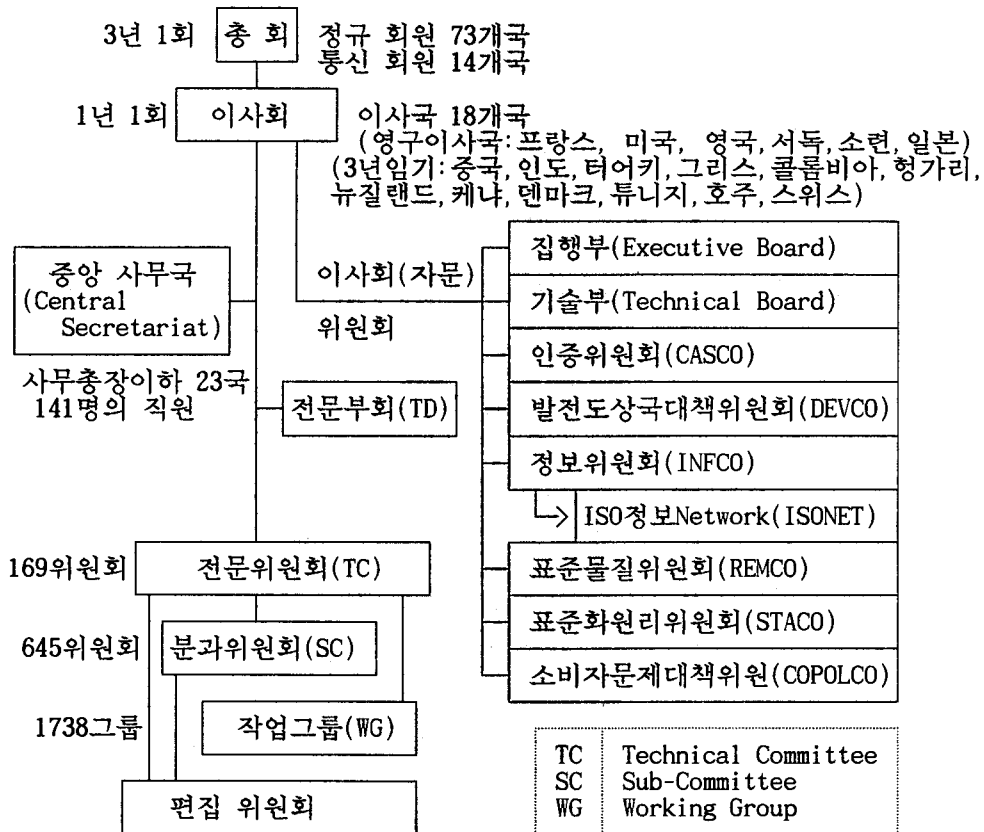
국제 표준화는 전기 및 전자 공학에 관련된 표준화로부터 시작되었는데, 이 분야의 표준화를 위하여 1906년에 국제 전기 기술 위원회(IEC:International Electrotechnical Commission)가 구성되었다. 1926년에는 기계 공학 분야에 대한 표준화를 위하여 만국 규격 통일 협회(ISA:International Standards Association)가 조직되었으나, 1930년대의 전쟁으로 말미암아 1942년에 활동이 정지되었다. 1944년에 연합국의 18개국이 중심이 되어 「공업 규격의 국제적

통일과 조정의 촉진을 목적으로 새로운 국제 기관의 설립을 추진되어, 1947년 2월 23일 유럽 국가를 중심으로 27개 회원국을 가지는 국제 표준화 기구(ISO)가 탄생하였다. 한국은 공업진흥청 표준국을 한국의 대표로 선정하여 1963년 6월 22일에 국제 표준화 기구에 가입하였다.

ISO의 활동은 원칙적으로 타 단체와 공동 연구를 통해 이루어지고 있다. ISO는 표준화에 관심 있는 450개의 국제 및 지역 조직과 연계를 취하고 있으며, 국제 전기 통신 연맹(ITU : International Telecommunication Union)과 국제 전기 기술 위원회(IEC)와는 특별히 긴밀한 관계를 맺고 있다. ISO와 IEC는 하나의 시스템으로 운영되며, 여기서 도출된 규격은 산업체와 서비스 조직에 유익하여 적용할 만하다. 이러한 대표적 공동 연구 기구로서 ISO/IEC JTC1(Joint Techn

ical Committee 1)을 들 수 있는데, ISO가 IEC와 정보(information) 기술 분야에 관한 공동 연구를 추진하기 위하여 구성된 기구이다. JTC1내에서 IEC는 하드웨어 위주의 연구를 수행하며, ISO는 소프트웨어 위주의 연구를 주로 추진하고 있다.

ISO의 조직 구성을 살펴보면 다음과 같다.



TC	Technical Committee
SC	Sub-Committee
WG	Working Group

(위원회 수는 1988년1월기준)

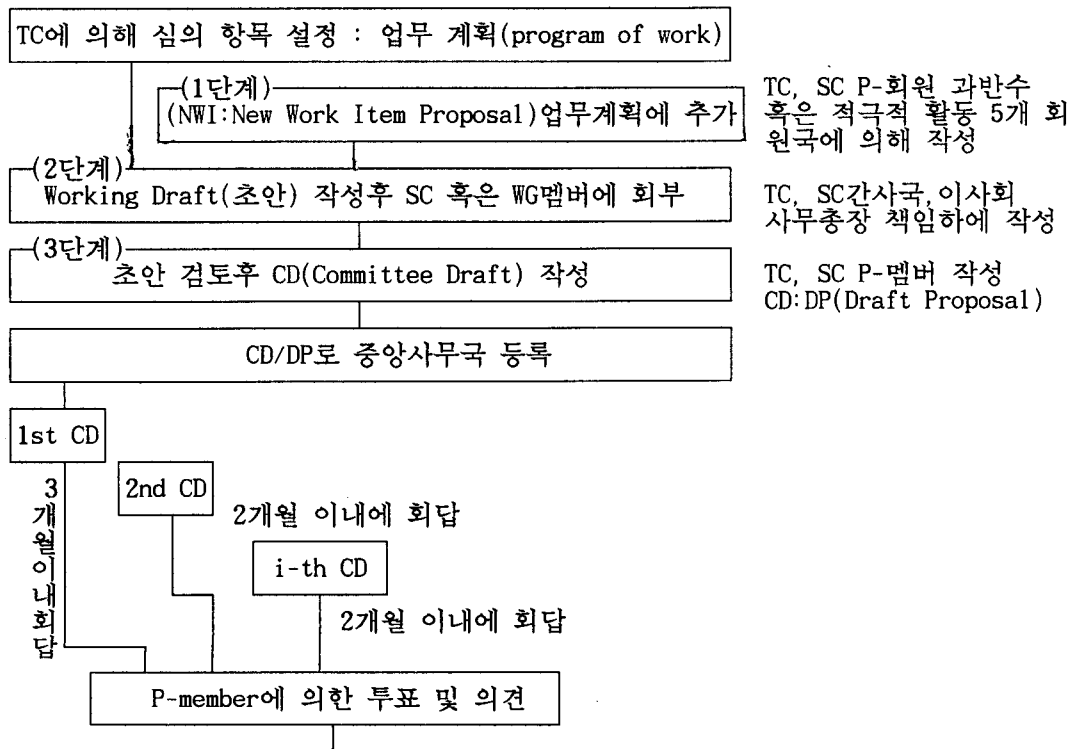
ISO/IEC의 국제 규격 제정 과정에는 약 800개의 전문 위원회(TC:Technical Committee)와 회원국 대표로 구성되는 분과 위원회(SC:Sub-committee), 그리고 기술 전문가로 구성되는 약3,000개의 작업 그룹(WG:Working Group)이 가담되고 있으며, 매년 약 1,000개의 국제 규격이 제정 또는 개정되고 있다.

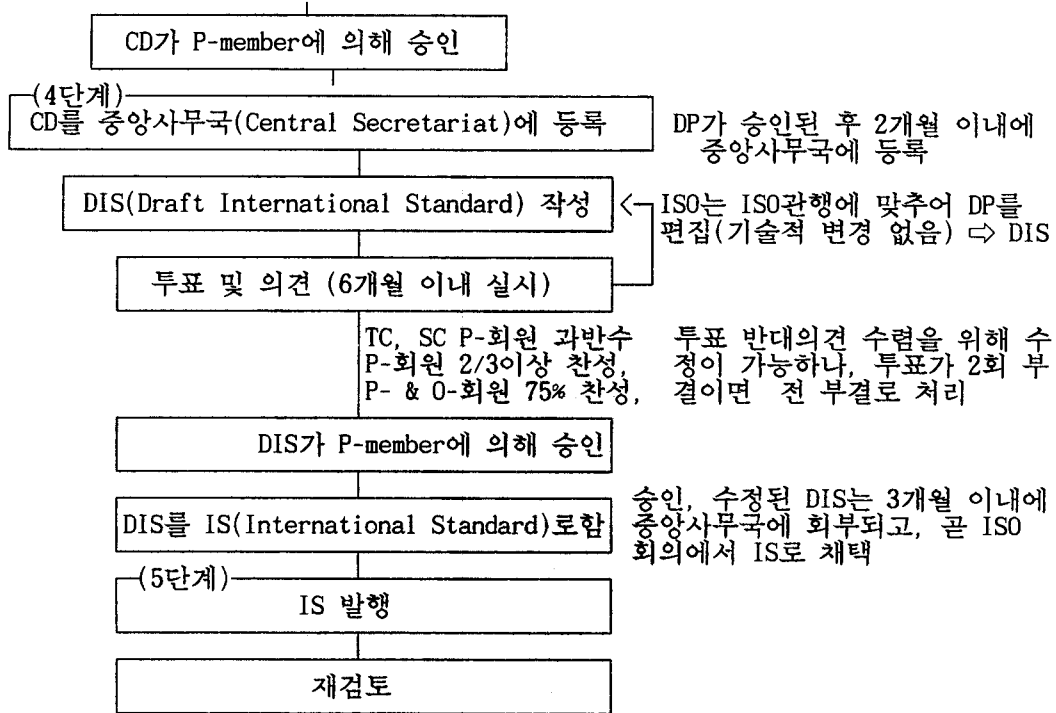
전문 위원회(TC:Technical Committee)는 국제 규격 작성의 임무를 가지고 있는 위원회로, 기술부의 권고에 따라 활동 업무 분야가 규정되며 이사회에서 설치한다. 전문 위원회의 위원은, 적극적 참가 의사 (permanent)를 표명한 회원 단체(P-member)와 업무 진행에 따른 정보 수신을 희망(observer)하는 회원 단체(O-member)로 구분한다. 간사국(Secretariat)은 P-member중 지명에서 지명되며, 의장은 간사국의 지명과 회의 출석자의 결의에 따라 임명되거나 이사회 결의에 따라 3년간 임명된다.

분과 위원회(SC:Sub-committee)는 전문 위원회의 특정 항목에 대해 조사, 심의할 책임을 지고 있으며, 작업 그룹(Working Group)은 전문 위원회 및 분과 위원회가 P-멤버 또는 관련 국제적 기관이 지명하는 전문가로 구성하는 전문가 조직으로 항목에 대한 연구를 수행한다.

편집 위원회(Editorial Committee)는 각 위원회에서 작성된 원안을 국제 편집 양식에 합치된 양식으로 작성하는 역할을 한다.

국제 규격의 입안 및 심의 단계를 살펴보면 다음과 같다. 규격 제정에 소요되는 시간은 1990년 기준으로 평균 30.4개월이며, 1996년부터는 3년 이내에 제정될 수 있도록 체제를 준비하고 있다.





국제 표준화 비용은 실제로 전세계 산업체들에 의해 주로 부담되고 있으며, 이들 산업체의 관심은 시장 성능의 개선이라는 형태로 나고 있다. 1983년부터 1993년까지 ISO 규격 기술 문서는 연간 약 15%씩 증가하였으며, 규격의 숫자도 5,000개에서 9,200개로 증가하는 등 국제 표준에 대한 관심은 현재 대단히 높아지고 있다.

국제 규격의 주요 기여도를 살펴보면 다음과 같다. 국제 규격은, 무역에 있어서 기술 장벽 제거하며, 합당한 가격에 상품 및 서비스의 품질과 신뢰도를 공급하고 있다. 사용자에게 만족과 안전을 제공하며, 상호 작용하는 상품과 서비스에 있어서 호환성 및 상호 동작성을 부여한다. 그리고, 사용자의 편리함을 위한 단순화와 환경 보호, 생산 효율 등에 기여하고 있다.

이와 같이, 국제 표준화는 무역 자유화에 대한 전세계적인 진전에 따른 필연적인 결과이며, 통신 시스템 발전의 세계화에 있어서도 반드시 필요하다. 그리고, 국제 표준화는 환경 측면의 지속 가능한 산업 발달을 이루기 위한 필요 조건이 되고 있으며, 지속적인 기술 교환을 위한 기술 이전에도 一助하고 있다. 그리고, 시장 경제로의 전환 또는 부분적인 조정 과정에 있는 경제 국들에게는 안정적인 영향력을 행사하고 있다. 따라서, 국제 표준화는 먼 장래까지도 계속해서 모든 산업에 추진될 것이다.

국내에서는, 정보 산업 분야 표준화 업무를 수행하기 위하여 공업진흥청 산하 단체로 1991년에 한국 산업 표준원 (IIS : The Institute for Industry Standards of Korea)이 발족되어, 산업계, 학계, 정부의 도움으로 ISO와 관련된 업무를 수행하고 있다. 한국 산업 표준원에서 추진 중인 ISO 관련 중점 수행 분야는 다음과 같다.

ISO/IEC JTC1	Information Technology
ISO TC46	Information and Documentation
ISO TC68	Banking and Related Financial Services
ISO TC69	Applications of Statistical Methods
ISO TC154	Documents and Data Elements in Administration, Commerce and Industry
ISO TC184	Industrial Automation Systems and Integration

## 2. 情報 交換用 符號의 國際 標準化 同鄉

### 2.1 國際 標準化 沿革

정보 교환이라 함은 서로 다른 시스템 사이에서 정보를 이용하기 위하여 하나의 시스템에서 다른 시스템으로 정보를 전달하는 것을 의미하며, 정보 교환용 부호는 이러한 정보 교환에 사용되는 부호를 말한다. 한글, 한자, 영문자를 비롯하여 현재의 세계 모든 문자는 정보 교환을 위하여 각각 서로 다른 부호 체계를 가지고 있어서 문자 정보의 교환에 많은 애로점이 있다. 이러한 애로점을 해결하는 방법으로는 각 나라마다 가지는 서로 다른 문자 부호 체계를 통일된 문자 부호로 통합하는 것이다. ISO는 이를 위하여 1984년 4월 ISO TC97/SC2/WG2에 『Two-octet Graphic Character Set』에 대한 연구 과제를 부과하였다.

ISO의 새로운 작업 항목(work item)으로 선정된 문자 부호의 국제 표준화 작업은, 1984년 11월 스위스에서 제1차 회의를 시작으로 본격적으로 「전 세계 문자를 포함하는 2바이트 문자 부호집을 위한 연구」가 진행되었다. 1987년 「2바이트의 도형 문자 부호집」에 대한 연구의 방향을 「복수 바이트의 도형 문자 부호집」으로 바꾸면서, 1987년에는 이 복수 바이트의 수를 4개로 결정하였다. 이는 현재의 국제 표준 부호로 등장한 UCS(Universal Multiple-Octet Coded Character Set)의 기본 골격이 되고 있다.

1989년 제12차 회의에서 한자에 대한 중국의 표준화 보고를 기점으로 한자에 대한 연구가 시작되었으며, 1990년 2월 「한자 부호의 국제 표준화」를 위하여 서울에서 회의가 열려 CJK-JRG(Chinese, Japanese, Korean Joint Research Group)를 결성하였다. 동아시아의 한자 문화권 국가가 참여하여 각국의 문자 표준 부호에 포함되어 있는 한자를 공동으로 검토, 연구하여, 통합된 하나의 한자 부호집을 작성하였다. 한국을 비롯하여 중국, 대만, 일본, 홍콩, 미국 등이 참여한 CJK-JRG는 1992년에 DIS 10646-1.2에 포함될 한자 20,902자를 선정하였으며, 이 한자는 1993년에 발표된 ISO/IEC 10646-1에 포함되었다. 1993년 5월, 그리스에서 열린 ISO/IEC JTC1/SC2/WG2 회의에서 CJK-JRG를 WG2 산하의 IRG(Ideographic Rapporteur Group)로 전환

하였으며, IRG의 활동 영역을 「동아시아 표의문자의 통합 문자 목록 및 배열에 관한 표준화」로 하였다. IRG에서는 CJK 한자의 사용 지침 작성과 ISO/IEC 10646-1 내의 한자 관련

내용의 보완, 수정을 주요 과제로 추진하고 있으며, ISO/IEC 10646-1에 포함된 한자를 보충하기 위해 확장 한자 세트의 연구에도 박차를 가하고 있다.

한자를 포함하여 전 세계 거의 모든 나라의 문자를 포함하는 4바이트 길이의 통한 문자 부호집은, 1993년 5월 ISO/IEC 10646-1(UCS)의 이름으로 국제 규격화되었으며, 현재 많은 시스템에서 이를 채택해 가고 있는 추세이다. 이 부호집의 도움으로, 전세계의 모든 문자는 통합된 하나의 문자 부호를 사용할 수 있게 되었으며 정보 교환용 시스템의 표준화가 가능하게 되었다.

## 2.2 UCS (Universal Multiple-Octet Coded Character Set) : ISO/IEC 10646-1

1977년 ANSI에 의해 공인되고 ISO 646으로 채택된 7-비트 체계의 ASCII (the American Standard Code for Information Interchange)가 현재의 정보 처리 교환용 부호로 표준의 역할을 하고 있으나, 전 세계의 모든 문자를 처리할 수 있는 코드의 필요성이 증가되면서 새로운 형태의 표준 부호의 등장이 요구되었다. 1988년 4월 전세계 표준을 위한 기구인 ISO (International Organization for Standardization)와 IEC (the International Electrotechnical Commission)가 정보 기술 관련 표준을 위한 JTC1(Joint Technical Committee 1)을 구성하였으며, ISO/IEC JTC1은 1992년 1월에 DIS 10646-1.2<sup>1)</sup>을 발표한 다음 약간의 수정과 보완을 거쳐, 1993년 5월에 국제 규격(IS)인 ISO/IEC 10646-1을 정식으로 제정하였다. 국제 규격(IS)인 ISO/IEC 10646-1은 전세계의 언어 및 부호의 표현, 전송, 교환, 처리, 저장, 입출력에 적용되며, UCS라는 이름으로 1993년 국제 표준으로 채택되었다.

4차원 구조의 UCS는 그룹(Group)-자면(Plane)-행(Row)-셀(Cell) 등으로 되어 모두 4개 바이트로 이루어진다. UCS는 128개의 3차원 그룹으로 구성되며, 각 그룹은 256개의 자면, 각 자면은 256개의 행을, 각 행은 다시 256개의 셀을 가지며, 하나의 문자는 1개의 셀에만 할당되도록 하였다. UCS에서는 문자의 표현 방법을 크게 두 가지로 나누고 있는데, UCS-4와 UCS-2가 그것으로, 각 옥테트(octet)의 값은 16진수로 표시한다.

UCS-4는 4-옥테트 표준 형태 (four-octet canonical form)로 'LATIN CAPITAL LETTER A'의 경우 '0000 0041'과같이 4개 바이트로 표현한다. 이 방법은 32-비트 프로세서(processor) 구조와 연관성을 엮을 수 있으며, 전세계의 모든 문자를 표현하기에 충분한 코딩 영역이 제공되고 있다. 그룹 00의 최초 자면(Plane 00)을 BMP(Basic Multilingual Plane)라고 하며, 다른 자면을 보조자면(Supplementary Plane)이라고 한다. BMP는 영문자(alphabetic), 음절 문자(syllabic), 표의문자(ideographic scripts), 기호(symbol) 및 숫자(digit) 등을 포함하고 있으며, 문자의 특수 용도를 위해 제한 영역(RU zone: Restricted Use zone)을 두고 있다. 보조자면 가운데는 국제 표준 (IS:International Standard)에서 특별히 셀의 내용을 정하지 않아서, 개인이 사용할 수 있는 자면이 있는데, 자면 E0에서 FF까지의 32개이며, 이를 PU 영역 (Private Use zone)이라 한다. PU 영역으로 Group 60에서 Group 7F의 32

1 DIS(Draft International Standard)는 IS(International Standard)의 초안을 의미한다.

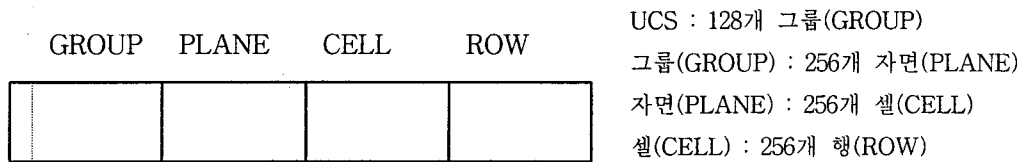
개 Group도 할당되어 있다. 이렇게 UCS-4는 그룹-옥테트, 자면-옥테트, 행-옥테트, 셀-옥테트 등으로 부호화 된다.

다른 문자 표현 형태로 UCS-2가 있는데, BMP영역을 사용하며 2개의 octet로 한 개 문자를 표현한다. 예를 들어, 'LATIN CAPITAL LETTER A'는 '0041'로 표현된다.

UCS는, UCS 이전의 국제 표준인 ISO 2022를 따르고 있는 통신 시스템과의 정보 교환을 위해 필요한 UCS와 ISO 2022 간의 코드 변환 알고리즘을 제시하고 있는데, 이를 UTF (UCS Transform Format)라 한다.

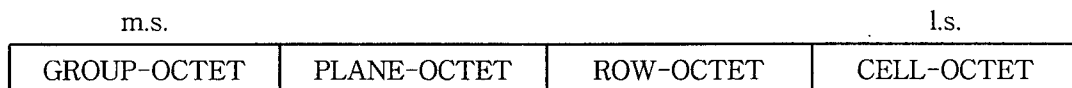
**2.2.1 基本 構造**

(1) 구조 : 4차원 코딩



- \* Group-octet(most significant octet)의 bit8은 장치의 내부 처리용으로 사용한다. 즉,
  - bit8=0 : character-coded-data-element 용
  - bit8=1 : internal processing 용

(2) 문자의 부호화 : 문자를 부호화할 때는 16 진수를 사용하여 표현한다.



m.s. : most significant

l.s. : least significant

(예) 0000 0030 --> DIGIT ZERO

0000 0041 --> LATIN CAPITAL LETTER A

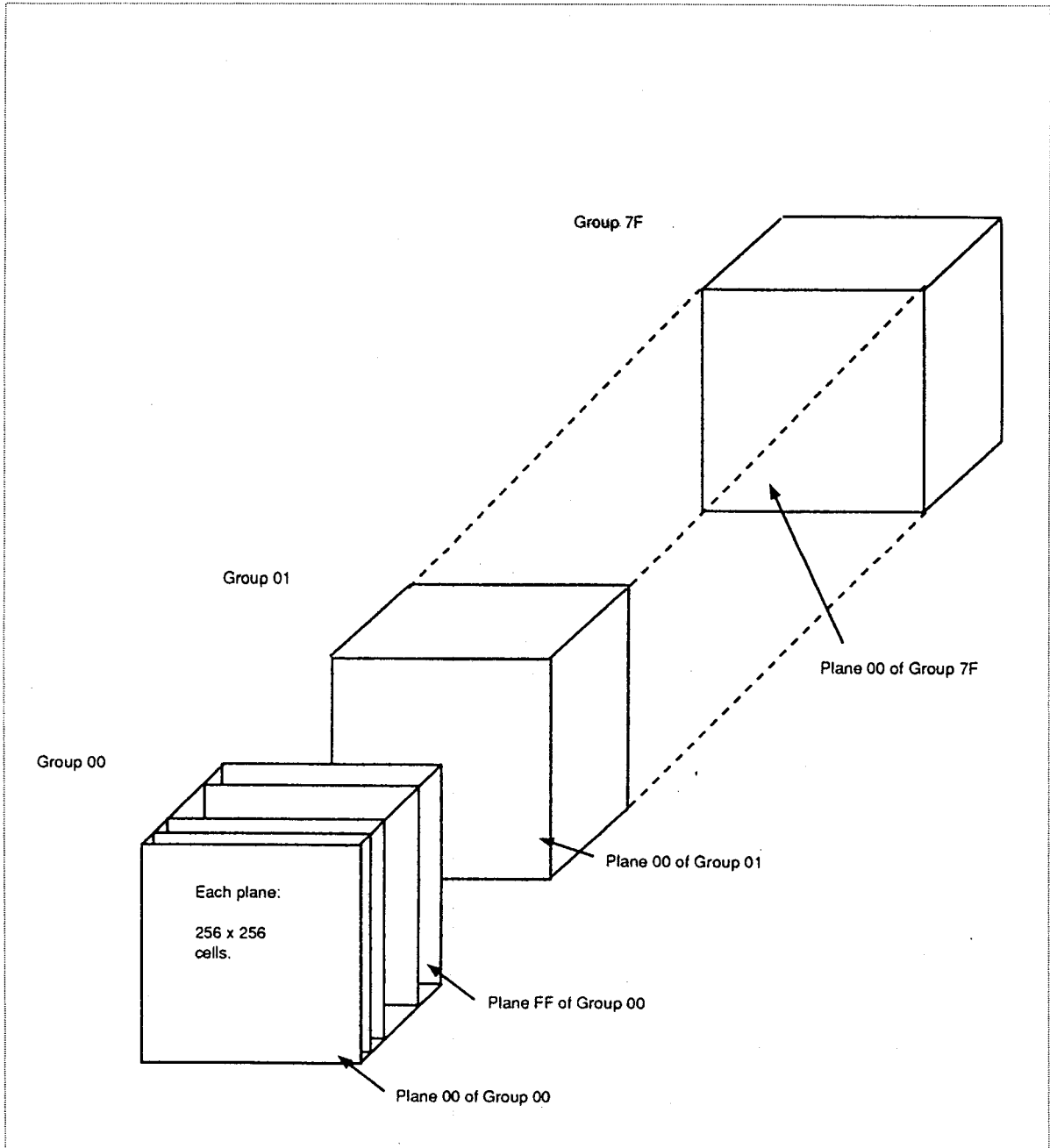
(3) Octet 순서

문자를 표현하는 옥테트의 순서는 일반적으로 상위 옥테트(a more significant octet)가 하위 옥테트(a less significant octet)보다 선행하여 처리한다. 그러나, 일렬적으로 표현하지 않을 경우에는 송·수신자의 상호 약속에 의해 새로이 규정될 수 있다.

**2.2.2 符號 領域의 特徵**

(1) P-, R-, C-옥테트는 00에서 FF의 영역을 가지며, G-옥테트는 00에서 7F의 영역을 갖는다. 그리고, 모든 자면(Plane) 상에서의 FFFE, FFFF 부호 값은 사용하지 않는다.

-FFFE : 'signature' 용 - 문자열의 시작점에 위치하여 문자열이 UCS-2, UCS-4로 부호화 되어 있는지, 그리고 각 문자 표현을 위한 옥테트(octet) 순서 등을 표시하는데 사용한다.



<UCS 문자 부호의 영역>



-FFFF : 내부 처리용 - 테이블 종료, 문서의 끝(end-of-text) 등과 같이 문자 부호가 아닌 숫자를 요구하는 경우나, 이진(binary) 혹은 연속(sequential) 검색에서 인덱스(index)의 최종 값으로 사용하는 경우에 사용한다.

- (2) IS에 할당되어 있지 않은 모든 부호 값은 미래 표준을 위해 예약되어 있으며 다른 용도로의 사용이 제한된다. 단 PU(Private Use) 영역은 예외이다.
- (3) 동일 도형 문자는 한 개의 부호 값을 갖는다.
- (4) 호환 문자(compatibility character) : 정보의 손실 없이 양방향(two-way) 부호 변환이 가능하도록 하기 위한 문자로서 기존 부호와의 호환성을 제공한다.

**2.2.3 字面(Plane)의 構成**

그룹(GROUP)	자면(PLANE)	내 용
00	00	BMP(basic multilingual plane)
	01-DF	미래 표준을 위한 예약
	E0-FF	PU(private use)
01-5F	00-FF	미래 표준을 위한 예약
60-7F	00-FF	PU(private use)
80-FF	00-FF	내부 처리용(internal processing)



**2.2.4 BMP (Basic Multilingual Plane)**

ZONE (영역)	ROW (행)	내 용	비 고
A-ZONE (영숫자 영역)	00	ISO-646, Latin, Greek, Cyrillic, Armenian, Hebrew, Arabic, Devanagari, Symbol, Dingbat, ..	A-ZONE : 19,903자
	30	CJK Symbol And Punctuation, Hiragana, Katakana	ゝ 〃 ㉿ ○ 『 』 [ ]
	31	Bopomofo, Hangul Compatibility Jamo, CJK Miscellaneous	丨 一 二 三 四 上 中 下
	32	Enclosed CJK Letters and Months	㊤ ㊦ ㊧ ㊨ ㊩
	33	CJK Compatibility	일본 단어(word)의 부호화
	4D	한글, 한글 보조 문자	
I-ZONE (한자 영역)	4E 9F	CJK Unified Ideographs (한자 20,902자)	I-ZONE : 20,992자
O-ZONE (예약 영역)	A0 DF	(미래 표준용으로 예약)	O-ZONE : 16,384자
R-ZONE (제한 영역)	E0 F8	Private Use Area	사용자 정의 문자
	F9 FA	CJK Compatibility Ideographs (CJK 호 환용 한자)	重出字
	FE	CJK Compatibility Forms	丨 ˆ ˘ ˙ ˚ ˛ ˜ ˝
	FF	Halfwidth and Fullwidth Form	R-ZONE : 8,190자

※ 제어 코드 (control code) 0000 - 001F  
007F - 009F

### 2.2.4 Basic Multilingual Plane (BMP)의 부호 구성

00	Basic Latin		Latin-1 Supplement	
01	Latin Extended-A		Latin Extended-B	
02	Latin Extended-B	IPA Extensions	Spacing Modifier Letters	
03	Combining Diacritical Marks		Basic Greek	Greek Symbols and Coptic
04	Cyrillic			
05	Armenian		Hebrew (Basic and Extended)	
06	Basic Arabic		Arabic Extended	
09	Devanagari		Bengali	
0A	Gurmukhi		Gujarati	
0B	Oriya		Tamil	
0C	Telugu		Kannada	
0D	Malayalam			
0E	Thai		Lao	
10	Georgian			
11	Hangul Jamo			
1E	Latin Extended Additional			
1F	Greek Extended			
20	General Punctuation	Super-/Subscripts	Currency Symbols	Comb. Diacritical Marks for Symbols
21	Letterlike Symbols	Number Forms	Arrows	
22	Mathematical Operators			
23	Miscellaneous Technical			
24	Control Pictures	O.C.R.	Enclosed Alphanumerics	
25	Box Drawing	Block Elements	Geometric Shapes	
26	Miscellaneous Symbols			
27	Dingbats			
30	CJK Symbols And Punctuation	Hiragana	Katakana	
31	Bopomofo	Hangul Compatibility Jamo	CJK Miscellaneous	
32	Enclosed CJK Letters and Months			
33	CJK Compatibility			
34	Hangul			
3E	Hangul Supplementary-A			
44	Hangul Supplementary-B			
45				
4D	CJK Unified Ideographs			
9F				
A0				
DF	Private Use Area			
E0				
F8	CJK Compatibility Ideographs			
F9				
FA				
FB	Alphabetic Presentation Forms			
FC	Arabic Presentation Forms-A			
FD				
FE	Comb. Half Marks	CJK Compat. Forms	Small Form Variants	Arabic Presentation Forms-B
FF	Halfwidth and Fullwidth Forms			Specials

 = reserved for future standardisation  
 = not graphic characters

**2.2.5 制限 領域 (Restricted Use zone)**

(1) PU (Private Use) 문자

ISO/IEC에서는 별도로 제한하지 않은 문자로서, 표의 문자 사용자의 공동 요구 사항인 사용자 정의 문자 (user-defined character)로 사용하거나, 응용 프로그램의 DRCS (dynamically-redefinable characters set)로 사용한다. 사용자 정의 문자가 의미 있는 전달되기 위해서는 송·수신자간의 사전 약속(agreement)이 있어야 한다.

(2) 표현 형식 (presentation form) 문자

기본 모양 이외에 또 다른 모양을 가지는 도형 문자에 있어서, 이들을 사용하는 기기와의 호환성 유지를 위해 표현 형식(presentation form) 문자를 사용하는데, Arabic 문자에 많이 나타난다. 보통 문자에 대치(substitution), 겹침(superimposition), 조합(combination) 등의 기법을 사용하여 표현 형식(presentation form) 문자로 전환한다.

(3) BMP의 호환(compatibility) 문자

F9행에서 FA행 사이에 위치하는 CJK 호환용 한자(CJK Compatibility Ideographs)의 용도는, 한국 표준 한자 세트와 같이 동일 한자에 대해 여러 개의 국내 표준 부호 값을 가지는 경우에 사용된다. CJK 호환용 한자는 ISO/IEC 10646-1과 자국 내의 국가 표준과의 정보 교환시 발생할 수 있는 정보 손실을 방지하여 정보의 양방향 변환이 가능하도록 한다. 이는 ISO/IEC 10646-1의 한자 배열은 『部首書數順』에 의한 배열이기 때문에 발생하는 것으로서, 한국의 경우 KS C 5601은 『漢字音順-한국』이기 때문에 268자의 重出字가 부호집 내에 존재한다. 이 경우, ISO/IEC 10646-1에서는 1개의 부호 값을 가지는 重出字가 KS C 5601의 부호 값으로 변경될 경우, 여러 개의 KS 부호 값 가운데 어느 부호 값으로 寫像(mapping)되어야 하는지가 곤란 한다. 즉, KS C 5601로 작성된 문서를 ISO/IEC 10646-1로 변환한 다음 다시 KS C 5601로 재변환 할 경우에, CJK 호환용 한자를 사용하지 않으면 KS C 5601로 표현된 최초의 문서와 ISO/IEC 10646-1로부터 변환된 문서는 동일하지 않게 된다. 정보의 손실이 발생하게 된다. 그러나, CJK 호환용 한자를 사용하면 이러한 정보 손실을 방지할 수 있게 되며, 따라서 양방향 변환이 가능하게 된다. 예를 들어 漢字「樂」은 ISO/IEC 10646-1에서는 부호 값「0000 6A02」의 값을 가지지만 KS C 5601-1987에서는 「악(6837)」, 「낙(4966)」, 「락(5305)」, 「요(7289)」 등과 같이 4개의 부호 값을 가진다. 이 경우 CJK 호환용 한자의 사용 방법은 다음과 같다.

ISO/IEC 10646-1		(양방향 변환)	KS C 5601-1987	
문자	부호값		문자	부호값
樂	6A02	←————→	樂(악)	6837
樂★	F914	←————→	樂(낙)	4966
樂★	F95C	←————→	樂(락)	5305
樂★	F9BF	←————→	樂(요)	7289

(주) ★ 는 제한 영역의 CJK 호환용 한자 임.

**2.2.6 部分 符號集(Subsets)**

송·수신 장치는 UCS 코드의 일부분을 문자 세트로 사용할 수 있는데, Limited Subset, Selected Subset 등 두 가지 방법이 있다.

(1) 제한 부분 부호집(Limited Subset)

UCS 또는 UCS와 다른 코드를 사용하는 공동 장치나 응용에 있어서, UCS내의 제한된 일부 도형 문자를 부호집으로 지정하여 사용할 수 있는데 이를 제한 부분 부호집(Limited Subset)이라 한다. 이 제한 부분 부호집은 국제 표준(IS:International Standard)가 정하는 도형 문자의 이름이나 부호 값의 나열로 호출한다.

(2) 선택 부분 부호집(Selected Subset)

선택 부분 부호집(Selected Subset)은 국제 표준(IS)이 정하는 도형 문자의 모음 목록으로 구성되는데, 이들 목록을 나열하여 호출, 사용한다. 이 부분 부호집은 그룹-00, 자면-00, 행-00의 셀-20에서 7E까지(0020-007E)는 자동으로 포함한다.

**2.2.7 UCS의 符號 表現 形式**

IS는 UCS-2, UCS-4 등 두 가지 표현 형식을 제시하고 있다.

(1) 2-옥테트 BMP 형식(form) : UCS-2

BMP 영역의 문자를 사용하며, 행-옥테트, 셀-옥테트 등 2개의 옥테트로 표현한다.

(2) 4-옥테트 표준 형식(canonical form) : UCS-4

4 octet로 표현하는 UCS의 표준 형태로 그룹-, 자면-, 행-, 셀-옥테트 등으로 표현한다.

**2.2.8 適用 레벨 (Implementation Level)와 組合 文字 (Combining character)**

ISO/IEC 10646에는 문자가 다른 문자에 주는 相互 動作에 따라 文字 세트를 크게 3가지로 구분하고 있다. 基本文字(base character), 組合文字(combining character), 한글字母(Hangul Jamo) 등이 그것인데, 조합 문자는 기본 문자와 결합하여 새로운 모양의 도형 문자를 만드는 기능을 가지는 문자로 혼자서는 사용되지 않는 문자를 말한다. 일반적으로 유럽의 영문자인 판독 기호(combining diacritical mark, 예: á â ã ä å ì ö ø 등) 조합 문자는 모음(vowel)으로 기본 문자인 자음(consonant)에 합쳐져서 새로운 문자를 만들어 낸다. 반면에, 한글字母는 이들 한글字母간만의 結合을 통해 한글 音節을 형성시키는 문자로서, 基本文字와 결합하는 組合文字와 달리 구분하고 있다.

이렇게 구분된 3가지의 문자 가운데 基本文字는 어느 경우에도 適用되고 있으나 組合文字나 한글字母는 適用水準(implementation level)에 따라 선택적으로 적용된다. 適用水準 1 (Implementation Level 1)에서는 조합 문자나 한글字母를 포함하지 않고 기본 문자만으로 구현하는 수준이며, 適用水準 2 (Implementation Level 2)는 기본 문자와 조합 문자 가운데 일부 文字(표에서 밑줄 친 문자)를 제외한 조합 문자로 적용하는 수준을 말한다. 適用水準 3 (Implementation Level 3)은 適用水準 2에서 제외된 일부 組合文字까지를 포함하는 문자, 즉 ISO/IEC 10646이 제공하는 모든 문자를 포함하여 적용하는 수준을 말한다.



**2.2.11 漢字 文字 (CJK unified ideograph)**

UCS 구조 내에 한자를 포함시키기 위해, 한국, 중국, 일본이 중심이 되어 대만, 홍콩, 미국 등이 참여한 CJK-JRG (Chinese/Japanese/Korean Joint Research Group)가 결성되어 통합 한자 세트를 연구, 개발하였다. 각국은 자국의 표준 한자 세트를 제시하고, 공동으로 한자 선정, 배열하여 1991년 12월에 통합 한자 세트 1차 버전으로 20,902자를 발표, DIS 10646에 포함시켰다.

① 사용된 각국 표준 부호

국가	한자 선정	한자 표준 부호	글자수	
한국	Hanja K source	K0	KS C 5601-1987	4,888
		K1	KS C 5657-1991	2,856
중국	Hanzi G source	G0	GB 2312-1980	6,763
		G1	GB12345-1990 + Hongkong 58 + 한국 「吏讀」 92	2,352
		G3	GB 7589-1987 (unsimplified form)	7,237
		G5	GB 7590-1987 (unsimplified form)	7,039
		G7	현대 중국어용 Hanzi list	42
		G8	GB 8565-1989	290
대만	Hanzi T source	T1	TCA-CNS 11643 第1字面	5,410
		T2	TCA-CNS 11643 第2字面	7,650
		TE	TCA-CNS 11643 第14字面	6,568
일본	Kanji J source	J0	JIS X 0208-1990	6,356
		J1	JIS X 0212-1990	5,801

KS: Korean Standards(한국)

GB : Guojio Biaozhun(중국)

CNS: Chinese National Standards(대만;자유 중국) JIS : Japanese Industrial Standards(일본)

TCA: Taipei Computer Association

② 통합/분리 基準

- 이체자 (異體字) : 同字이나 字體를 달리하는 文字로 분리한다.
- 3차원 Model에서 Y-axis의 문자를 통합함을 기본으로 한다.

- ┌ X-axis : ideographic character
- ├ Y-axis : ideographic forms
- └ Z-axis : type face difference

③ 配列 方法 : 部首, 劃數에 따라 배열하되 다음 辭典을 기준으로 삼는다.

- ① 共通 - 康熙字典
- ② 日本 - 大漢和辭典 (諸橋轍次, 昭和 61年, 修訂版)
- ③ 中國 - 漢語大字典 (湖北辭書出版社, 1986)
- ④ 韓國 - 大字源 (張三植, 三省出版社, 1990, 2版)

### 2.2.12 UCS 와 ISO 2022의 變換

UCS로 작성된 문서 데이터를 현재의 통신 시스템을 통해 전송할 때, ISO 2022에서 도형 문자로는 사용하지 않는 C0, C1, SPACE, DEL 등은 배제되어야 한다. 이렇게 UCS에 대응되면서 ISO 2022를 고려한 코드 형태를 UTF(UCS transformation format)라 하는데 이는 여러 옥테트(octet)로 구성되며, UCS에서 제공하는 규칙에 따라 변환된다.

그 변환 규칙을 요약하면 다음과 같다.

- ① 16진수로 표시하며, octet 경계는 ';' (semicolon)으로 한다.
- ② modulo (%), 정수 나눗셈(/), 멱수(Mn) 등의 연산자를 사용하며, 연산 우선 순서는 멱수 > 나눗셈 > modulo 순이다.
- ③ UTF octet수는 1-octet에서 5-octet를 사용하며, 4-octet는 3-octet의 사용 영역을 최대한으로 하기 위해 사용하지 않는다.
- ④ 함수 T(z)는 다음으로 정의한다.

$$\begin{aligned}
 z = 00 \dots 5D & \quad T(z) = z + 21 \\
 z = 5E \dots BD & \quad T(z) = z + 42 \\
 z = BE \dots DE & \quad T(z) = z - BE \\
 z = DF \dots FF & \quad T(z) = z - 60
 \end{aligned}$$

⑤ 함수 U(z)는 다음으로 정의한다.

$$\begin{aligned}
 z = 00 \dots 5D & \quad U(z) = z + BE \\
 z = 21 \dots 7E & \quad U(z) = z - 21 \\
 z = 7F \dots 9F & \quad U(z) = z + 60 \\
 z = A0 \dots FF & \quad U(z) = z - 42
 \end{aligned}$$



## 2.3 Unicode

### 2.3.1 背景

1987년, Xerox Palo Alto Research Center의 Joe Becker, Lee Collins와 Apple사의 Mark Davis는 보다 간단하며 일관성 있는 코드 개발에 착수하여, ISO 10646에 대응하는 Unicode의 표준화 작업에 들어갔다. 그후 1991년 1월에는 Unicode의 사용을 촉진하기 위해 기업체 중심의 Unicode 컨소시엄(consortium)을 설립하였는데 참여 주요 회사는 다음과 같다.

- |                                |                           |                          |
|--------------------------------|---------------------------|--------------------------|
| . Adobe                        | . Aldus                   | . Apple                  |
| . Borland                      | . DEC                     | . Ecological Linguistics |
| . GO                           | . IBM                     | . Lotus                  |
| . Metaphor                     | . Microsoft               | . NeXT                   |
| . Novell                       | . Pacific Rim Connections |                          |
| . The Research Libraries Group |                           | . SUN                    |
| . WordPerfect                  | . Xerox                   |                          |

Unicode의 특성은 아래와 같으며 Unicode란 말도 그 특성에 연유되었다.

- universal                      세계 모든 언어 포함
- uniform                        16-bit의 길이를 가지는 효율성
- unique                         문자 코드의 중복을 최소화

1991년 9월에 Unicode 1.0은 완료, 발표되었으며, 1993년에 ISO/IEC 10646(UCS)이 발표된 후에는 UCS와 통합된 하나의 문자 부호집으로 통일되게 되었다.

### 2.3.2 Unicode 制定 原理

- Completeness : 텍스트(text) 작성에 필요한 모든 문자 포함한다.  
     ⇨ Sanskrit등 死(dead)文字도 포함한다.
- Simplicity 와 efficiency : 동일 길이의 코드이다. (16 비트)
- Unambiguity : 각 부호는 한 개의 문자를 표현한다.
- Correctness : 부호화되는 모든 문자는 어학 전문가의 인정을 통한 실제 문자이다.
- Fidelity : 기존 문자로의 변환시 데이터의 손실이 없다.

## 2.4 REACC (RLIN East Asian Character Code)

미국 연구 도서관 네트워크, RLIN (Research Libraries Information Network)이 학술 자료의 처리를 위해 Harvard, Yale, Columbia 대학교가 주축이 되어 1974년 구축되었다. 그후 36개의 대학 도서관과 미국 의회 도서관이 이 기구에 가입하면서 한.중.일 도서 목록 등 동양 자료 처리의 필요성이 제기되었다.

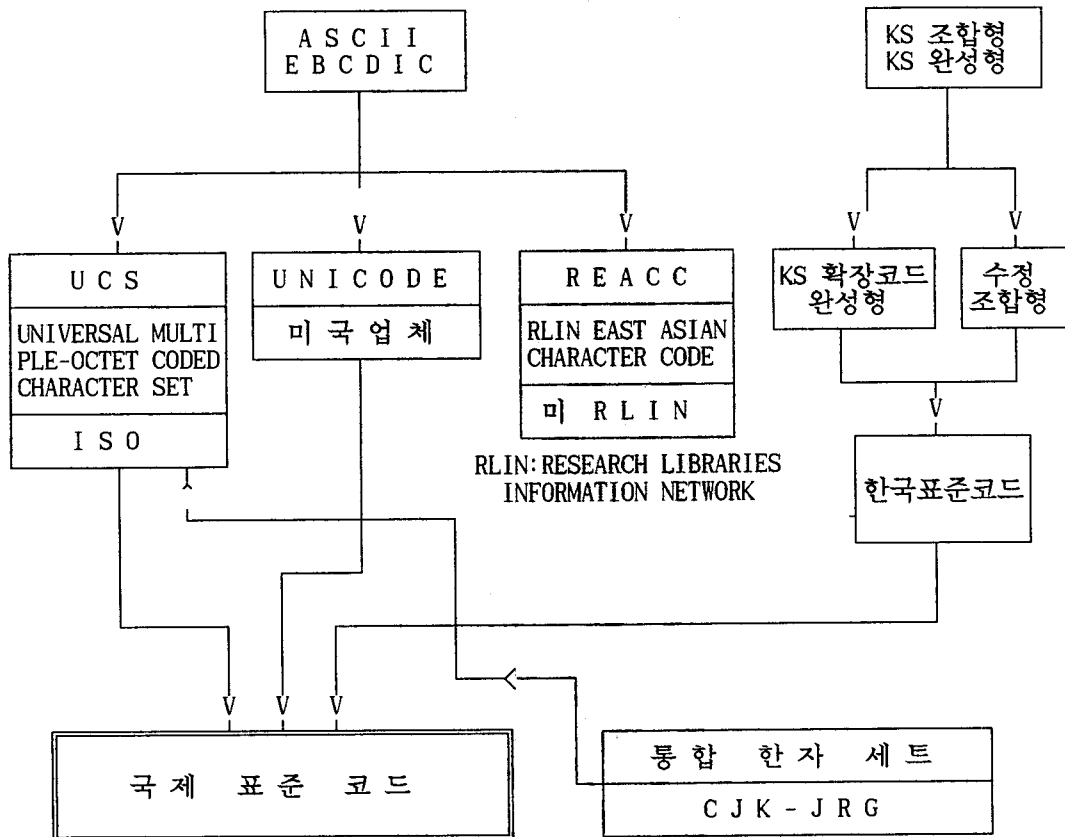
1983년 RLIN에서는 대만의 한자 코드로 3-바이트 체계 (Plane/Section/Position)를 가진

CCCII를 기본으로 하여, 한국의 KIPS(Korean Information Processing System) 코드, 일본의 JIS C 6226, 중국의 GB 2312-1980과 CTC(Chinese Telegraph Code)등을 포함하는 코드 체계를 발표하였는데 이를 REACC (RLIN East Asian Character Code)라 한다. 이 REACC에는 한자 14,063자, 한글 1996자, 일본 가나 174자가 포함되어 있다.

1987년 ANSI (American National Standard Institute)에서 이를 수용하여 미국의 표준으로 사용되었으며, 1989년은 REACC를 ISO에 국제 표준으로서의 수용을 요구하였다. 그러나, 한국, 중국, 일본의 소극적인 자세와 당시 ISO에 검토 중인 DP 10646 (Draft Proposal)로 인해 ISO에서의 수용은 실현되지 않았다.

### 3. 맺음말

#### 3.1 情報 交換用 符號의 要約



### 2.3 情報 交換用 符號의 國際 標準化에 對한 所見

(1) UCS 動向 : 현재의 ASCII 부호 체계는 1977년 ANSI에 의해 공인되고 ISO 646(대부분 동일한 내용임)으로 채택된 이후 실질적인 情報 交換用 國際 標準 부호로 자리잡고 있다. 그러나, ISO 및 Unicode에서 부호의 표준화 작업이 진행되면서, UCS (ISO/IEC 10646-1) 부호 체계는 전세계의 모든 문자를 지원한다는 강점을 가지고 ASCII 부호의 獨走를 제지하고 있다. 초기의 UCS는 복잡한 각국의 이해 관계와 Unicode와의 부호 배열의 불일치라는 문제점이 있었으나, 이러한 문제점이 모두 해소되어 1993년에 새로운 국제 표준(IS:International Standard)으로 발표되었다.

(2) UCS와 ASCII : IBM, Apple, Xerox, Microsoft 등 정보 처리 산업의 우수 업체가 이들 정보 교환용 부호의 국제 표준화 작업에 적극적으로 參與하면서, 새로운 부호 체계를 適用한 소프트웨어 및 하드웨어 開發이 進行하고 있으며, 일부는 完成 段階에 있다. 또한, 한국, 중국, 일본 등 비 ASCII 부호 체계의 국가에서는 다소 拒否感이 없는 것은 아니나, 대체적으로 이러한 표준화에 肯定的으로 參與하고 있다. 따라서, ISO의 표준으로 1993년에 UCS가 확정 발표됨으로서 빠른 속도로 현행 부호 체계를 代替할 可能性이 있다고 생각된다.

(3) UCS에의 對應 : 새로운 표준 부호의 등장은 현행 부호와의 부호 이식(code migration) 그리고 호환성이 중요한 要件인 만큼 이를 대비하는 자세가 필요하다. 정보 교환 시스템을 다루는 産業界나, 學界, 使用者 團體는 이러한 변화에 신속히 대응하여, 정보 산업에 있어서 主人 役割을 하여야 할 것이다. 특히, 산업계에 있어서 소프트웨어와 하드웨어의 國產化 作業은, 지금까지 국내 업체에서 의해 主導的으로 수행되어 왔으나, 앞으로는 이러한 한글화가 한국만의 獨占物이 될 수 없음을 인식하여야 할 것이다. 그런가 하면 한편으로는, 국내에서 개발된 우수한 워드프로세서, 스프레드시트, DBMS, 디스플레이 컨트롤러(display controller), 통신 시스템 등의 제품은 국내뿐만 아니라 외국에서도 그대로 사용될 수 있는 면이 있는 만큼, 이는 우리가 가지는 새로운 기회라고 할 수 있다. 따라서, 단순한 한글화 작업을 뛰어넘는 관련 분야의 核心 技術 確保야말로, 새로운 정보 교환용 부호의 국제 표준화와 더불어, 정보 산업에 있어서의 우리의 防禦的 위치를 세계화라는 攻擊的 위치에 설 수 있도록 하는 새로운 계기가 될 것이라고 생각한다.

參考文獻

---

- (1) "ISO/IEC 10646-1 : Information technology - Universal Multiple-Octet Coded Character Set (UCS) -", ISO, 1993
- (2) "KS C 5601-1992 : 정보 교환용 부호(한글 및 한자)", 한국표준협회, 1992년10월15일
- (3) "KS C 5657-1991 : 정보 교환용 부호 확장 세트", 한국표준협회, 1991년12월31일
- (4) "KS C 5620-1977(1990 확인) : 정보 교환용 부호의 확장법", 한국표준협회, 1977년12월23일
- (5) "GB/T 12345-90 : 信息交換用漢字編碼字符集 輔助集", 國家技術監督局, 1990年6月13日
- (6) "GB 7589-87 : 信息交換用漢字編碼字符集 第二輔助集", 國家標準局, 1987年3月27日
- (7) "GB 7590-87 : 信息交換用漢字編碼字符集 第四輔助集", 國家標準局, 1987年3月27日
- (8) "CNS X5012(11643) : 中文標準交換碼", 經濟部中央標準局印行, 民國 75年8月4日
- (9) "JIS X 0208-1990 : 情報交換用漢字符號", 日本規格協會, 平成 2年9月1日
- (10) "JIS X 0212-1990 : 情報交換用漢字符號-補助漢字", 日本規格協會, 平成 2年9月1日
- (11) "ISO 2022-1986 : Information processing - ISO 7-bit and 8-bit coded character sets - Code extension techniques", ISO, 1986
- (12) "ISO 2375-1985 : Data processing - Procedure for registration of escape sequences", ISO, 1985

권춘우(權春佑) 정회원

---

- 1958년생 경북 안동
- 1981년 2월 서울대학교 공과대학 전자공학과 졸업
- 1983년 2월 서울대학교 대학원 전자공학과(공학석사)
- 1983년~1992년 금성사 중앙연구소 연구실장
- 1992년~1995년 1월 현재 동양공업전문대학 전산사무자동화과 교수