

LỚP TẬP HUẤN VỀ SỬ DỤNG TCVN 6909 TRONG KHU VỰC NHÀ NƯỚC

Giới thiệu tiêu chuẩn TCVN 6909 về bộ mã các ký tự quốc ngữ 16 bit theo Unicode

Trần Lưu Chương
Ban công tác Unicode, Bộ KHCNMT

Dàn bài

1. Lược sử các bộ mã chữ quốc ngữ và việc ban hành TCVN 6909
2. Nội dung của tiêu chuẩn TCVN 6909 về bộ mã chữ quốc ngữ 16 bit theo Unicode
3. Những công cụ hỗ trợ TCVN 6909
4. Kết luận

I. Lược sử các bộ mã chữ quốc ngữ (1982-2001) và việc ban hành TCVN 6909

- ◆ **1982-1983** **Bộ mã chữ Việt cho các máy tính 8-bit (Viện KHTTĐK, nay là Viện Công nghệ Thông tin)**
- ◆ 1984-1987 Trong nước : BKED, EDVN, VIED, 3C, VietRes, BKHCM...
◆ Ngoài nước : VNLab, VNI, VietStar, VMS, VietStd, VNU/win,
◆ VNLab/Ventura , VIQR ...
- ◆ **1-1987** **Hội nghị toàn quốc lần thứ nhất về các hệ xử lý văn bản tiếng Việt tại Hà Nội**
- ◆ 6-1987 Thành lập tiểu ban Chuẩn hóa mã chữ Việt trong Chương trình cấp nhà nước về tin học, giai đoạn 1986-1990
- ◆ **7-1988** **Hội nghị toàn quốc lần 2 về chuẩn hóa bộ mã các ký tự Việt**
- ◆ **12-1991** **Bắt đầu đặt quan hệ với ISO/IEC JTC1/SC2 và Unicode**
- ◆ 12-1991 UBKHNN thành lập tiểu ban Chuẩn hóa mã chữ Việt
- ◆ 5-1992 Phiên bản số 1 của Dự thảo bộ mã chuẩn các ký tự Việt được gửi tới Tổng Cục TCDLCL để đăng ký thành Tiêu chuẩn Việt nam

I. Lược sử các bộ mã chữ quốc ngữ (1982-2001)

- ◆ **15-05-1992** **Việt nam bỏ phiếu tán thành Dự thảo bộ mã đa ngữ ISO 10646 bit với yêu cầu bổ sung thêm chữ Nôm vào bảng mã này.**
- ◆ 7-1992 Hội thảo giới thiệu bộ mã trong Tuần lễ Tin học lần thứ 3 ở Hà Nội
- ◆ 2-12-1992 Hội đồng kỹ thuật do Tổng Cục TCDLCL triệu tập đã nhất trí thông qua Dự thảo bộ mã chuẩn (trên cơ sở ý kiến thẩm định của 33 chuyên gia)
- ◆ 1-93/5-93 Theo đề nghị của nhóm Viet-Standards ở Mỹ và của một số tập thể trong nước nên Tiểu ban chuẩn đã đề nghị lui thời hạn ban hành để xem xét thêm
- ◆ **12-05-1993** **Bộ KHCNMT ký quyết định ban hành tiêu chuẩn Việt nam về bộ mã chuẩn các ký tự Việt dùng trong trao đổi thông tin TCVN 5712:1993**

I. Lược sử các bộ mã chữ quốc ngữ (1982-2001)

- ◆ 10-1994 Ban Chỉ đạo CTQG về CNTT quyết định ký hỗ trợ xây dựng bộ cài đặt TCVN 5712, sau này gọi là ABC.
- ◆ **25-12-1995 Bộ KHCNMT ban hành “Qui định sử dụng ABC trong khối các cơ quan nhà nước”**
- ◆ 7-1996 Kiến nghị về bộ mã 8 bit các ký tự quốc ngữ trên cơ sở bộ mã VN1 trong TCVN 5712:1993 (Nguyễn Chí Công)
- ◆ 1996-1997 Kiến nghị về các bộ mã 8 bit 8859-V, cp1258 và cp1129
- ◆ 20,21-5-97 Hội thảo về xử lý chữ Việt trên máy tính do Ban Chỉ đạo CTQG về CNTT tổ chức tại Hà nội
- ◆ **1999 Ban hành tiêu chuẩn TCVN 5712:1999 thay thế cho TCVN 5712:1993**
- ◆ **7-05-2001 Ủy Ban KHCNMT của Quốc Hội yêu cầu Bộ KHCNMT báo cáo về tình trạng các chuẩn chữ Việt**
- ◆ **8-08-2001 Bộ KHCNMT thành lập Ban công tác Unicode**

Khuyến cáo của Hội thảo tháng 5-1997

Khuyến khích các cơ quan nghiên cứu và công ty CNTT trong nước tìm hiểu các vấn đề thời sự hiện nay của CNTT quốc tế như : **bộ mã 16 bit Unicode**, xử lý tiếng nói và nhận dạng quang học chữ Việt, tổ chức trang Web tiếng Việt hội nhập với nhiều thứ tiếng trên thế giới, nói riêng hoà hợp được chữ Việt và chữ biểu ý trong khu vực vv...

Thuận lợi của Unicode: giải quyết trọn vẹn việc thống nhất chữ Việt trên máy tính

B1. Unicode: bộ mã đa ngữ, hỗ trợ mọi loại chữ viết, trong đó có chữ Việt la-tinh (quốc ngữ) và chữ nôm.

B2. Được nhiều quốc gia và nhiều hãng lớn chấp nhận, kể cả Việt nam (ngay từ phiên bản đầu tiên năm 1992)

B3. Đã có nhiều sản phẩm hỗ trợ bộ mã chữ quốc ngữ theo Unicode.

B4. Đã có một số tổ chức trong và ngoài nước dùng bộ mã chữ quốc ngữ theo Unicode.

Một số website dùng Unicode chữ Việt

- <http://www.vietkey.com/>
- <http://www.tinhvan.com.vn/>
- <http://www.vnn.vn/>
- <http://25th.net/>
- <http://www.vovisoft.com/>
- <http://www.omely.com/>
- <http://4vn.net/>
- <http://quartzit.com/a1/>
- <http://www.nguoi-viet.com/>
- VV....

2. Nội dung của Tiêu chuẩn TCVN 6909:2001

**Ban hành theo quyết định số
50/2001/QĐ-BKHCMNT
ngày 24-9-2001
của Bộ Khoa học, Công nghệ và
Môi trường**

2. Nội dung của Tiêu chuẩn TCVN 6909:2001 (tiếp theo)

TCVN 6909 gồm 6 điều :

- 1. Phạm vi ứng dụng**
- 2. Tính phù hợp**
- 3. Tiêu chuẩn trích dẫn**
- 4. Thuật ngữ và định nghĩa**
- 5. Các tập ký tự**
- 6. Bộ mã và mô tả ký tự**

2. Nội dung của Tiêu chuẩn TCVN 6909:2001 (tiếp theo)

1. Phạm vi ứng dụng

- Qui định bộ mã 16-bít để biểu diễn, lưu trữ và trao đổi các ký tự chữ Việt trong CNTT
- Qui định cấu trúc của bộ mã các ký tự chữ Việt
- Không qui định hình dạng, kích thước, kỹ thuật hiển thị và xử lý các ký tự này.

2. Nội dung của Tiêu chuẩn TCVN 6909:2001 (tiếp theo)

2. Tính phù hợp

- Phù hợp ISO/IEC 10646-1:2000 và Unicode 3.0 trong mặt phẳng cơ bản – BMP (Basic Multilanguage Plan) đối với các ký tự quốc ngữ. Các ký tự khác theo 2 tiêu chuẩn trên.
- Phù hợp với TCVN 6909 nếu
 - a. Có sử dụng các tập ký tự trong Điều 5, và
 - b. Mã các ký tự tuân theo đúng Điều 6

2. Nội dung của Tiêu chuẩn TCVN 6909:2001 (tiếp theo)

3. Tiêu chuẩn trích dẫn

- **ISO/IEC 10646-1:2000 – Universal Character Set (UCS), Part 1 Structure and Basic Multilanguage Plan (BMP)**
- **Unicode 3.0**
- **TCVN 6695-1:2000, CNTT : Từ vựng – Các thuật ngữ cơ bản**

2. Nội dung của Tiêu chuẩn TCVN 6909:2001 (tiếp theo)

4. Thuật ngữ và định nghĩa

Ký tự: Bit; Byte; Octet → theo TCVN 6695

Chức năng điều khiển

Bảng mã

Ký tự khả hiện

Ký tự dấu thanh

Mã ký tự

Ký tự dấu nguyên âm

Tập ký tự mã hóa

Bộ mã ký tự

2. Nội dung của Tiêu chuẩn TCVN 6909:2001 (tiếp theo)

5. Các tập ký tự

5.1 Tập ký tự cơ bản

5.2 Tập ký tự mở rộng

5.1 Tập ký tự cơ bản: 119 ký tự gồm 6 nhóm

1. Nhóm chữ cái: 33 chữ cái (12 nguyên âm và 21 phụ âm cơ bản)
hoa và 33 chữ cái thường

Bảng 1 - Các ký tự nguyên âm

Chữ hoa	A	Ă	Â	E	Ê	I	O	Ô	Ơ	U	Ư	Y
Chữ thường	a	ă	â	e	ê	i	o	ô	ơ	u	ư	y

Bảng 2 - Các ký tự phụ âm

Chữ hoa	B	C	D	Đ	F	G	H	J	K	L	M
Chữ thường	b	c	d	đ	f	g	h	j	k	l	m



Chữ hoa	N	P	Q	R	S	T	V	W	X	Z	
Chữ thường	n	p	q	r	s	t	v	w	x	z	



2 Nhóm dấu thanh: gồm 5 kí tự



3 Nhóm dấu nguyên âm: gồm 3 kí tự



4 Nhóm chữ số: gồm 10 kí tự

0	1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---	---

5 Nhóm kí hiệu: 20 kí tự

⊕	"	#	\$	%	&	'	*	+	-	/
	<	=	>	@	\	^	_		`	~

6 Nhóm dấu câu: 15 kí tự

SP	!	()	,	.	:	;	?	[]	{	}	“	”
----	---	---	---	---	---	---	---	---	---	---	---	---	---	---

2. Nội dung của Tiêu chuẩn TCVN 6909:2001 (tiếp theo)

5.2 Tập ký tự mở rộng: 120 ký tự nguyên âm mang dấu thanh gồm 60 chữ thường và 60 chữ hoa.

Bảng 8 - Các kí tự nguyên âm mang dấu thanh

Chữ hoa	À	Ả	Ã	Á	Ạ	Ằ	Ẵ	Ẳ	Ẵ	Ẳ
Chữ thường	à	ả	ã	á	ạ	ằ	ẵ	ẳ	ẵ	ẳ

Chữ hoa	À	Ằ	Ẵ	Á	Ậ	È	Ẻ	Ễ	É	Ệ
Chữ thường	à	ằ	ẵ	á	ậ	è	ẻ	ễ	é	ệ

Chữ hoa	È	Ẻ	Ễ	É	Ệ	Ì	Ỉ	Ĩ	Í	Ị
Chữ thường	è	ẻ	ễ	é	ệ	ì	ỉ	ĩ	í	ị

Chữ hoa	Ò	Ổ	Ỗ	Ó	Ợ	Ồ	Ỗ	Ỗ	Ó	Ộ
Chữ thường	ò	ổ	ỗ	ó	ợ	ồ	ỗ	ỗ	ó	ộ



Chữ hoa	Ờ	Ở	Ỡ	Ớ	Ợ	Ù	Ủ	Ữ	Ú	Ự
Chữ thường	ờ	ở	ỡ	ớ	ợ	ù	ủ	ữ	ú	ự



Chữ hoa	Ừ	Ỡ	Ỡ	Ứ	Ự	Ỡ	Ỡ	Ỡ	Ỡ	Ỡ
Chữ thường	ừ	ỡ	ỡ	ứ	ự	ỡ	ỡ	ỡ	ỡ	ỡ

Close Full Screen

2. Nội dung của Tiêu chuẩn TCVN 6909:2001 (tiếp theo)

6. Bộ mã và mô tả các ký tự

6.1 Latinh 1	0000 – 00FF
6.2 Latinh mở rộng	0100 – 01FF
6.3 Vùng dấu thanh và dấu nguyên âm	0300 – 03FF
6.4 Latinh mở rộng bổ sung	1E00 – 1EFF 2000 – 20FF

Bảng 9 - Bảng mã kí tự trong vùng 0000 - 00FF

	000	001	002	003	004	005	006	007	008	009	00A	00B	00C	00D	00E	00F
0			SP	0	@	P	`	p			À		À		à	
1			!	1	A	Q	a	q					Á		á	
2			"	2	B	R	b	r					Â	Ò	â	ò
3			#	3	C	S	c	s					Ã	Ó	ã	ó
4			\$	4	D	T	d	t						Ô		ô
5			%	5	E	U	e	u						Õ		õ
6			&	6	F	V	f	v								
7			'	7	G	W	g	w								
8			(8	H	x	h	x					È		è	
9)	9	I	Y	i	y					É	Ù	é	ù
A			*	:	J	Z	j	z					Ê	Ú	ê	ú
B			+	;	K	I	k	{								
C			,	<	L	\	l						Ì		ì	
D			-	=	M]	m	}					Í	Ý	í	ý
E			.	>	N	^	n	~								
F			/	?	O	_	o									

Unicode Character Map

Font: Times New Roman
Next

Characters to Copy:

Subset: Latin-1
Previous

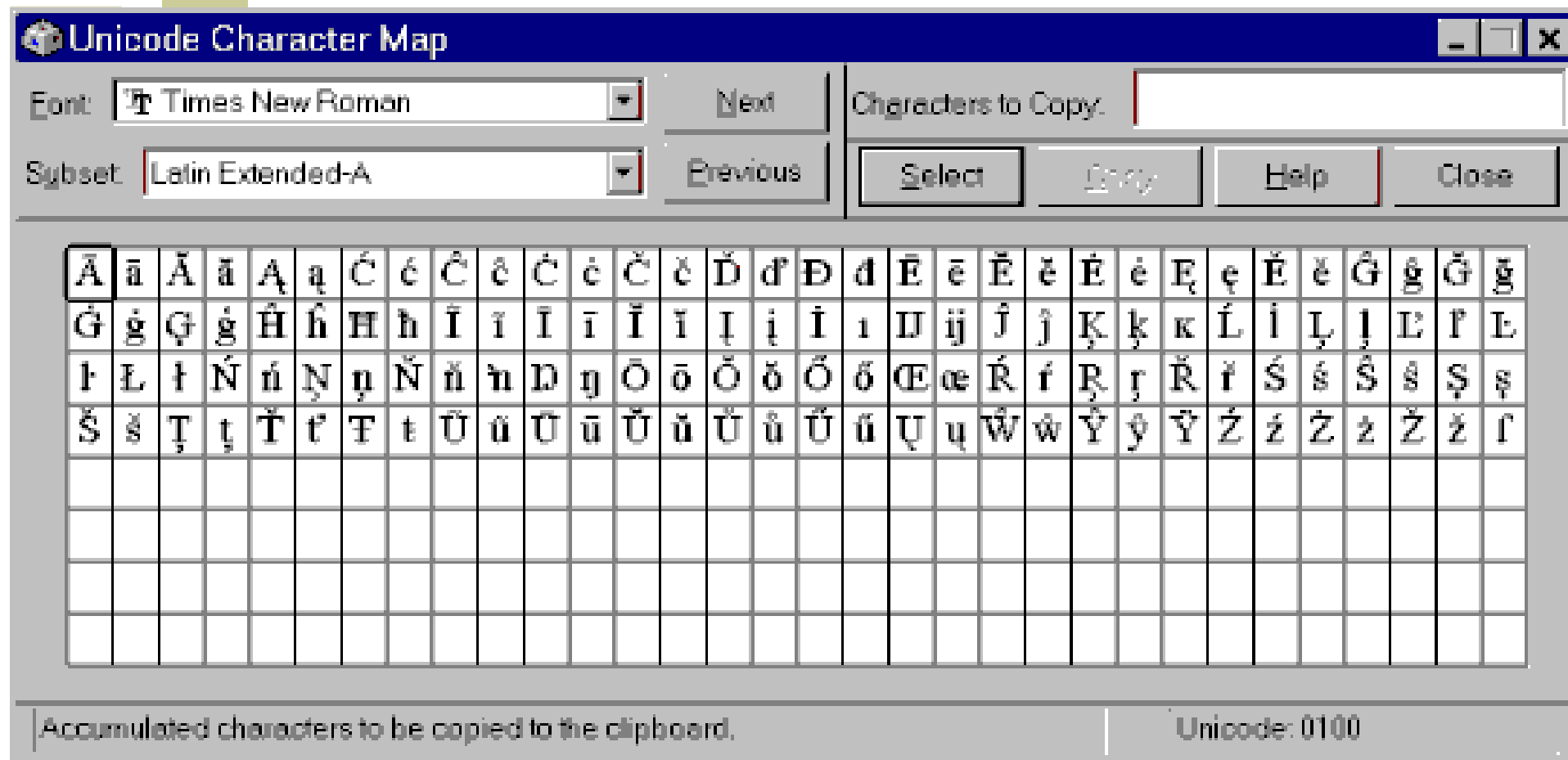
Select
Copy
Help
Close

	!	"	#	\$	%	&	'	()	*	+	,	-	.	/	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
	ı	€	£	¤	¥	¦	§	¨	©	ª	«	¬	®	¯	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿	
À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

Accumulated characters to be copied to the clipboard.
Unicode: 0020

Bảng 10 - Bảng mã kí tự trong vùng 0100 - 01FF

	010	011	012	013	014	015	016	017	018	019	01A	01B	01C	01D	01E	01F
0		Đ									Ơ	ư				
1		đ									ơ					
2	Ă															
3	ă															
4																
5																
6																
7																
8			Ĩ				Ũ									
9			ĩ				ũ									
A																
B																
C																
D																
E																
F											Ư					



Bảng 11 - Bảng mã kí tự trong vùng 0300-03FF

	030	031	032	033	034	035	036	037	038	039	03A	03B	03C	03D	03E	03F
0	0															
1	1															
2	2															
3	3		4													
4																
5																
6	6															
7																
8																
9	9															
A																
B		0														
C																
D																
E																
F																

Bảng 12 - Bảng mã kí tự trong vùng 1E00-1EFF

	1E0	1E1	1E2	1E3	1E4	1E5	1E6	1E7	1E8	1E9	1EA	1EB	1EC	1ED	1EE	1EF
0											Ạ	Ằ	Ẻ	Ó	Ỡ	Ự
1											ạ	ằ	ẻ	ó	ỡ	ự
2											Ẳ	Ẵ	Ẽ	Ồ	Ợ	Ỡ
3											ẳ	ẵ	ẽ	ồ	ợ	ỡ
4											Ẳ	Ẵ	Ẽ	Ồ	Ự	Ỡ
5											ẳ	ẵ	ẽ	ồ	ự	ỡ
6											Ẳ	Ẵ	Ẽ	Ồ	Ự	Ỡ
7											ẳ	ẵ	ẽ	ồ	ự	ỡ
8											Ẳ	Ẵ	Ẽ	Ồ	Ự	Ỡ
9											ẳ	ẵ	ẽ	ồ	ự	ỡ
A											Ẳ	Ẵ	Ẽ	Ồ	Ự	Ỡ
B											ẳ	ẵ	ẽ	ồ	ự	ỡ
C											Ặ	Ẻ	Ợ	Ờ	Ự	Ỡ
D											ặ	ẻ	ợ	ờ	ự	ỡ
E											Ặ	Ẻ	Ợ	Ờ	Ự	Ỡ
F											ặ	ẻ	ợ	ờ	ự	ỡ

▼ Full Screen

Close Full Screen

Bảng 13 - Bảng mã kí tự trong vùng 2000-20FF

	200	201	202	203	204	205	206	207	208	209	20A	20B	20C	20D	20E	20F
0																
1																
2																
3																
4																
5																
6																
7																
8																
9																
A																
B																
C		“														
D		”														
E																
F																

Close Full Screen ▾

3. Những công cụ hỗ trợ TCVN 6909

- 1. Các hệ điều hành**
- 2. Các tiện ích**
- 3. Các phần mềm văn phòng**
- 4. Các phần mềm quản trị CSDL**
- 5. Các phần mềm Internet**

3. Những công cụ hỗ trợ TCVN 6909 (tiếp theo)

1. Nhập dữ liệu (data entry -- input)
2. Truyền dữ liệu (data transfer)
3. Xử lý dữ liệu (data/information processing)
4. Lưu trữ dữ liệu (data/information archival)
5. Biểu hiện dữ liệu (data/information presentation)
6. Trao đổi, tìm kiếm dữ liệu (data/info exchange & searching)



Kết luận