

TCVN

TIÊU CHUẨN QUỐC GIA

TCVN 6437 : 1998

**CÔNG NGHỆ THÔNG TIN –
BỘ KÝ TỰ DÙNG CHO NHẬN DẠNG QUANG HỌC
OCR-VN - HÌNH DẠNG VÀ KÍCH CỠ CHỮ IN**

*Information Technology – Character Set for Optical Recognition
OCR-VN – Shapes and Dimensions of the Printed Image*

HÀ NỘI – 2008

Lời nói đầu

TCVN 6437 : 1998 do Ban kỹ thuật tiêu chuẩn TCVN/JTC1 “*Công nghệ Thông tin*” biên soạn, Tổng cục Tiêu chuẩn Đo lường Chất lượng đề nghị; Bộ Khoa học, Công nghệ và Môi trường (nay là Bộ Khoa học và Công nghệ) ban hành.

Tiêu chuẩn này được chuyển đổi năm 2008 từ Tiêu chuẩn Việt Nam cùng số hiệu thành Tiêu chuẩn Quốc gia theo quy định tại khoản 1 Điều 69 của Luật Tiêu chuẩn và Quy chuẩn kỹ thuật và điểm a khoản 1 Điều 6 Nghị định số 127/2007/NĐ-CP ngày 1/8/2007 của Chính phủ quy định chi tiết thi hành một số điều của Luật Tiêu chuẩn và Quy chuẩn kỹ thuật.

Công nghệ thông tin – Bộ ký tự dùng cho nhận dạng quang học OCR-VN – Hình dạng và kích cỡ chữ in

*Information Technology – Character Set for Optical Recognition OCR-VN –
Shapes and Dimensions of the Printed Image*

1 Phạm vi áp dụng

1.1 Tiêu chuẩn này quy định các yêu cầu về kích cỡ và hình dạng in được của các ký tự OCR-VN. Tiêu chuẩn này áp dụng cho in ấn và nhận dạng quang học các văn bản quốc ngữ cũng như các văn bản của nhiều thứ chữ thuộc họ La-tin.

1.2 Tiêu chuẩn này không qui định một bộ mã ký tự. Các quy định về chất lượng in ấn và quy tắc chế bản nằm trong những tiêu chuẩn khác.

1.3 Tiêu chuẩn này cũng áp dụng cho các thiết bị in và nhận dạng quang học chữ in có thể tham khảo tiêu chuẩn này.

2 Tài liệu viện dẫn

ISO/IEC 1073-2: 1976 Công nghệ thông tin Các bộ ký tự dùng cho nhận dạng quang học

Phần 2 : Bộ ký tự OCR-B - Hình dạng và kích cỡ chữ in.

3 Khái niệm và giải thích

3.1 Bộ ký tự OCR-VN bao gồm các chữ số, chữ cái viết hoa và viết thường, các dấu câu, các biểu tượng và các dấu phụ (bao gồm dấu phụ quốc tế và dấu thanh), cho phép dùng trong in ấn và nhận dạng quang học các văn bản quốc ngữ.

3.2 Một thiết bị in (hoặc một thiết bị nhận dạng quang học chữ in) được gọi là phù hợp với tiêu chuẩn này nếu nó có thể in được (hoặc nhận dạng được) một văn bản thể hiện bằng một số hoặc toàn bộ những tập ký tự OCR-VN ở một hoặc nhiều kích cỡ đã qui định ở điều 7 của tiêu chuẩn này.

TCVN 6437 : 1998

3.3 Ký tự dựng sẵn trong tiêu chuẩn này được hiểu là một ký tự trong bảng chữ cái quốc ngữ đi kèm một dấu phụ và có hình dạng chữ thường hoặc chữ hoa tạo ra bằng phương pháp dựng sẵn.

4 Quy định chung

4.1 Tiêu chuẩn ISO/IEC 1073-2 quy định bộ ký tự OCR-B có 121 ký tự, bao gồm các dấu câu, các chữ số từ "0" đến "9", các chữ hoa từ "A" đến "Z" và các chữ thường từ "a" đến "z" theo bảng chữ cái La-tin, các biểu tượng, các dấu phụ và một số chữ cái riêng của vài thứ tiếng Âu-Mỹ.

4.2 Tiêu chuẩn này quy định bộ ký tự OCR-VN có 137 ký tự, bao hàm bộ ký tự OCR-B và các ký tự riêng của quốc ngữ không có trong OCR-B (cụ thể là 14 chữ cái: "Ă", "Â", "Ê", "Ô", "Ơ", "Ư", "Đ", "ă", "â", "ê", "ô", "ơ", "ư", "đ" và 2 dấu thanh: "hỏi", "nặng").

4.3 Đối với phần lớn các ứng dụng nhận dạng quang học, mỗi ký tự mang dấu được xác định với hình dạng và dung sai cho trước.

4.4 Đối với các ứng dụng in ấn, tùy theo khả năng kỹ thuật của thiết bị, mỗi ký tự mang dấu đều có thể được hiển thị theo phương pháp dựng sẵn hoặc tổ hợp (tức là dấu phụ được đặt ở trên hoặc dưới ký tự tương ứng trong cùng một ô chữ). Trong phụ lục của tiêu chuẩn này có đầy đủ các bản vẽ với các ký tự dựng sẵn thể hiện mọi tổ hợp giữa các dấu phụ và các ký tự tương ứng.

5 Các kiểu chữ

5.1 Có hai kiểu chữ được quy định ở đây: kiểu chữ có độ rộng nét chữ cố định và kiểu chữ có độ rộng nét chữ biến đổi. Hình dạng của một ký tự nói chung được xác định bởi đường chủ đạo (đường ở giữa nét chữ) và các đường viền của ký tự đó.

5.2 Đối với kiểu chữ có độ rộng nét chữ cố định, chỉ cần xác định các đường chủ đạo của nét chữ và các đường viền ở phần cuối nét chữ.

5.3 Đối với kiểu chữ có độ rộng nét chữ biến đổi, cần phải xác định đường chủ đạo của nét chữ và đầy đủ mọi đường viền của ký tự.

6 Các kích cỡ

6.1 Các kích cỡ trong tiêu chuẩn này được xác định sao cho có thể sử dụng được với hầu hết các loại thiết bị in ấn có chất lượng khác nhau, thí dụ: máy chữ, máy in bao bì, máy in kim, máy in phun mực, máy in offset, máy in laser, v.v. Trong thực tiễn một số công nghệ hoặc thiết bị in ấn không cho phép thực hiện các góc chữ quá sắc, tuy nhiên đối với nhận dạng quang học chữ in thì bán kính của các góc chữ không cần thiết phải chính xác đến mức nhỏ hơn 0,08 mm (0,0035 in).

CHÚ THÍCH - Các ứng dụng công nghệ thông tin thường dùng phổ biến hai đơn vị đo độ dài là mm và in. Các kích thước trong tiêu chuẩn này được làm tròn ở cả hai đơn vị đó và không thể coi là chính xác tuyệt đối. Khi áp dụng cụ thể tiêu chuẩn này chỉ cần dùng một loại đơn vị và không được dùng lẫn lộn cả hai.

6.2 Kiểu chữ có độ rộng nét chữ cố định được xác định ở ba cỡ I, III và IV. Đối với việc in chữ có bước in cố định dùng cho những ứng dụng nhận dạng quang học thì những bước in danh định là như sau:

cỡ I: tối thiểu 2,54 mm (0,100 in)

cỡ III: tối thiểu 2,54 mm (0,100 in)

cỡ IV: tối thiểu 3,63 mm (0,143 in)

6.3 Kiểu chữ có độ rộng nét chữ biến đổi chỉ được xác định ở cỡ I (cỡ nhỏ nhất). Kiểu chữ này cho phép sử dụng bước in biến đổi giữa các ký tự như thường thấy trong các nhà in.

6.4 Các đường chủ đạo trong cả ba cỡ trên có quan hệ với nhau do được thu phóng theo hai chiều dọc ngang với một tỷ lệ thích hợp. Tỷ lệ đó ở các kích cỡ III và IV so với cỡ I là như sau:

cỡ III: dọc 1,333 ngang 1,086

cỡ IV: dọc 1,500 ngang 1,500

Tỷ lệ thu phóng đường chủ đạo không được áp dụng cho các đường viền nét chữ bởi vì độ rộng nét chữ danh định không phải luôn luôn là tỷ lệ thuận với cỡ của đường chủ đạo. Độ rộng nét chữ cho mỗi cỡ trên được xác định trong phần phụ lục của chuẩn.

6.5 Trong mỗi cỡ, ký tự với đường viền có cao độ lớn nhất ở trên đường cơ sở là số "8" và ký tự với đường viền có cao độ lớn nhất ở dưới đường cơ sở là chữ "j".

Đường chủ đạo của số "8" có chiều cao là:

cỡ I: 2,40 mm (0,094 in)

cỡ III: 3,20 mm (0,126 in)

cỡ IV: 3,60 mm (0,142 in)

6.6 Ký tự rộng nhất trong mỗi cỡ là số "0" (trừ trường hợp đối với chữ "m dự bị").

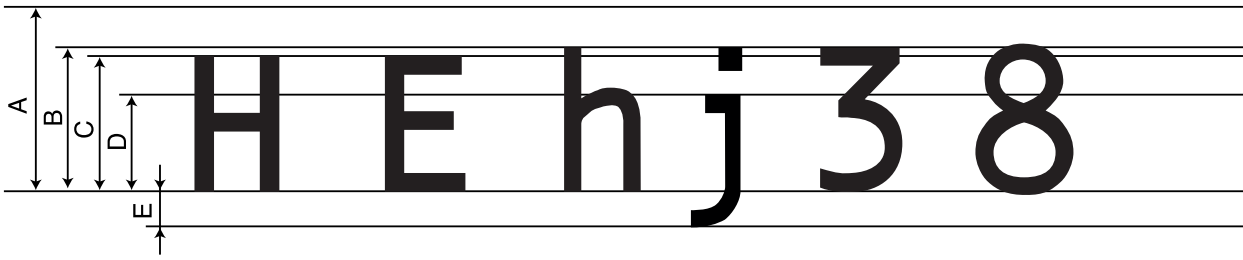
Đường chủ đạo của số "0" có độ rộng như sau:

cỡ I: 1,40 mm (0,055 in)

cỡ III: 1,52 mm (0,060 in)

cỡ IV: 2,10 mm (0,083 in)

6.7 Kích thước đặc trưng cho mỗi ký tự được xác định bởi độ rộng của ký tự (xem phụ lục) và bởi một số cao độ được đo từ đường cơ sở (xem hình 1).



Hình 1 – Các cao độ ở trên và dưới đường cơ sở

Bảng 1 – Các cao độ đặc trưng ở cỡ I

Đơn vị đo	mm					inch				
	A	B	C	D	E	A	B	C	D	E
Cỡ I	3,43	2,66	2,46	1,83	0,60	0,135	0,105	0,097	0,072	0,024

7 Các tập ký tự

Bộ ký tự OCR-VN được chia thành 5 tập ký tự như sau:

7.1 Tập ký tự tối thiểu

Tập ký tự tối thiểu gồm có 22 ký tự sau đây (kể cả dấu cách):

0 1 2 3 4 5 6 7 8 9

< + >

C E N S T X Z

! (dấu cách)

CHÚ THÍCH - Không nên sử dụng những ký tự C E N S T X Z trong nhận dạng quang học những văn bản chỉ gồm có một hoặc hai dòng.

7.2 Tập ký tự cơ bản

Tập ký tự cơ bản gồm tập ký tự tối thiểu và 25 ký tự khác, tổng cộng là 47 ký tự:

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

< + > * - = / - ,

! (dấu cách)

7.3 Tập ký tự mở rộng

Tập ký tự mở rộng gồm tập ký tự cơ bản và 51 ký tự khác, tổng cộng là 98 ký tự:

! " # \$ % & ' () * + , - . /

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 [\] ^ _

p q r s t u v w x y x { | } ~

! (dấu cách)

7.4 Tập ký tự đầy đủ

Tập ký tự đầy đủ bao gồm tập ký tự mở rộng và 37 ký tự riêng của quốc ngữ và của một số thứ chữ Âu-Mỹ, tổng cộng là 135 ký tự (không kể các nguyên âm mang dấu của quốc ngữ có thể tạo ra được bằng tổ hợp của các nguyên âm với các dấu thanh):

! " # % & ' () * + , - _ . /
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 { | } ~

¡ (dấu cách)

Ä Å Æ I J Ñ Ö Ø Ü å æ i j ø ß § ¥
Â Ă Đ Ê Ô Ơ Ư â ă đ ê ô ơ ư
“ ^ , ’ . m _

7.5 Tập các ký tự xóa

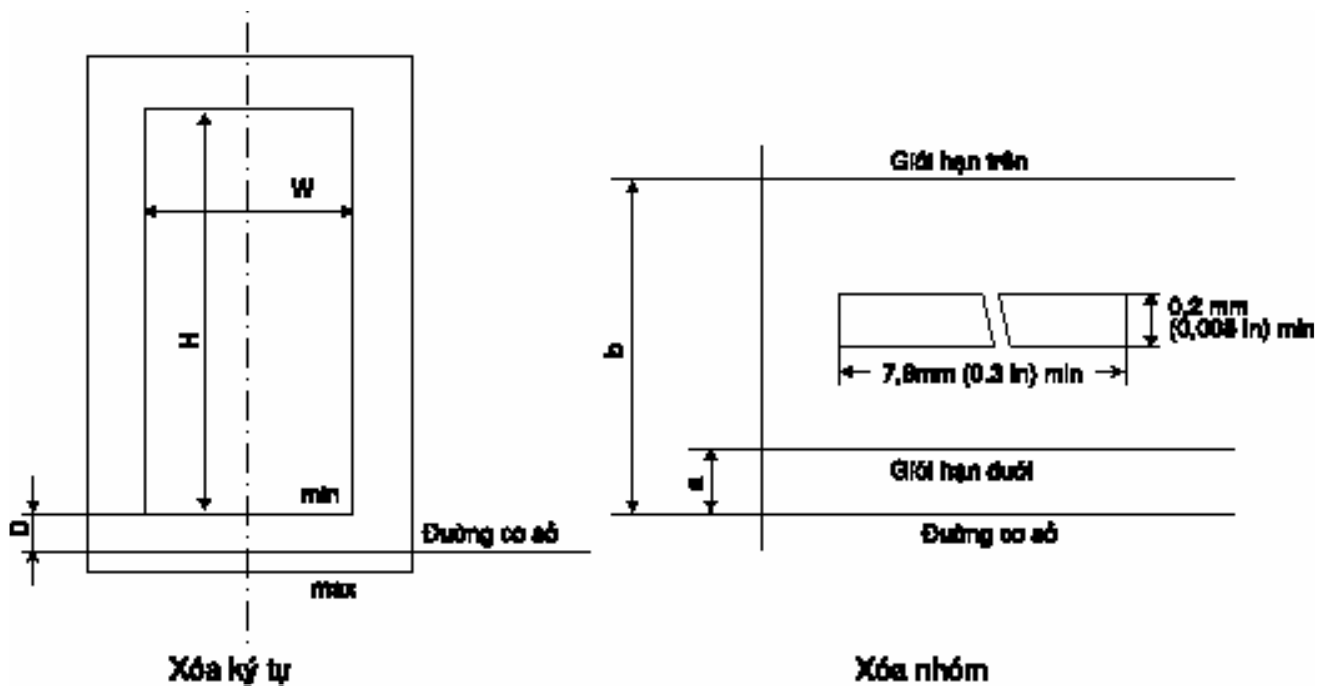
Tập này chỉ gồm có 2 ký tự có chức năng xóa:

7.5.1 Ký tự "Xóa một ký tự"



7.5.2 Ký tự "Xóa nhóm ký tự"





Hình 2 – Các ký tự xóa

Bảng 2 – Kích thước của các ký tự xóa

Đơn vị đo	mm			inch		
	I	III	IV	I	III	IV
Cỡ						
Ký tự "Xóa một ký tự":						
min. H	2,4	–	3,8	0,094	–	0,149
max. H	2,9	–	4,6	0,115	–	0,181
min. W	1,4	–	2,0	0,055	–	0,079
max. W	1,9	–	2,8	0,075	–	0,110
D	0,13	–	0,20	0,005	–	0,008
Ký tự "Xóa nhóm ký tự":						
độ dài tối thiểu	7,6	7,6	10,9	0,300	0,300	0,430
độ rộng tối thiểu	0,2	0,2	0,2	0,008	0,008	0,008
a	0,4	0,5	0,6	0,016	0,020	0,024
b	2,0	2,7	3,0	0,077	0,106	0,118

8 Bảng tra các ký tự OCR-VN

8.1 Mọi ký tự của kiểu chữ có độ rộng nét chữ biến đổi và cố định đều có ở cỡ I.

Chỉ những ký tự của tập ký tự tối thiểu và ký tự "Xóa nhóm ký tự" là có ở cỡ III với độ rộng nét chữ cố định.

TCVN 6437 : 1998

Mọi ký tự của kiểu chữ có độ rộng nét chữ cố định, trừ ký tự "dấu số dọc", đều có ở cỡ IV.

8.2 Trong bảng tra sau đây, mỗi ký tự đều có chỉ dẫn về tập hoặc những tập ký tự chứa nó và về bản vẽ chi tiết thể hiện hình dạng của nó ở phần phụ lục.

8.3 Các hình dạng của kiểu chữ có độ rộng nét chữ cố định ở cỡ IV có thể tạo ra từ những hình dạng của kiểu chữ đó ở cỡ I.

8.4 Trật tự của các ký tự trong bảng tra sau đây không liên quan gì đến trật tự của chúng trong mỗi bảng mã bất kỳ.

Cột "GHI CHÚ" trong bảng tra sau đây bao gồm một số chỉ dẫn về việc áp dụng, đặc biệt cho biết những ký tự nào là nhằm điều đích sử dụng chung và không nên dùng vào điều đích nhận dạng quang học.

Bảng 3 - Bảng tra các ký tự OCR-VN

TT	Hình ký tự	Bản vẽ số	Tên ký tự	Tập số	Ghi chú
1	1	1	SỐ MỘT	1,2,3,4	
2	2	2	SỐ HAI	1,2,3,4	
3	3	3	SỐ BA	1,2,3,4	
4	4	4	SỐ BỐN	1,2,3,4	
5	5	5	SỐ NĂM	1,2,3,4	
6	6	6	SỐ SÁU	1,2,3,4	
7	7	7	SỐ BẢY	1,2,3,4	
8	8	8	SỐ TÁM	1,2,3,4	
9	9	9	SỐ CHÍN	1,2,3,4	

10	O	10	SỐ KHÔNG	1,2,3,4	
11	A	11	CHỮ A HOA LA-TIN	2,3,4	
12	B	12	CHỮ B HOA LA-TIN	2,3,4	
13	C	13	CHỮ C HOA LA-TIN	1,2,3,4	
14	D	14	CHỮ D HOA LA-TIN	2,3,4	
15	E	15	CHỮ E HOA LA-TIN	1,2,3,4	
16	F	16	CHỮ F HOA LA-TIN	2,3,4	
17	G	17	CHỮ G HOA LA-TIN	2,3,4	
18	H	18	CHỮ H HOA LA-TIN	2,3,4	
19	I	19	CHỮ I HOA LA-TIN	2,3,4	
20	J	20	CHỮ J HOA LA-TIN	2,3,4	
21	K	21	CHỮ K HOA LA-TIN	2,3,4	
22	L	22	CHỮ L HOA LA-TIN	2,3,4	
23	M	23	CHỮ M HOA LA-TIN	2,3,4	
24	N	24	CHỮ N HOA LA-TIN	1,2,3,4	

Bảng 3 - Bảng tra các ký tự OCR-VN (tiếp theo)

TT	Hình ký tự	Bản vẽ số	Tên ký tự	Tập số	Ghi chú
25	O	25	CHỮ O HOA LA-TIN	2,3,4	
26	P	26	CHỮ P HOA LA-TIN	2,3,4	
27	Q	27	CHỮ Q HOA LA-TIN	2,3,4	
28	R	28	CHỮ R HOA LA-TIN	2,3,4	
29	S	29	CHỮ S HOA LA-TIN	1,2,3,4	
30	T	30	CHỮ T HOA LA-TIN	1,2,3,4	
31	U	31	CHỮ U HOA LA-TIN	2,3,4	
32	V	32	CHỮ V HOA LA-TIN	2,3,4	
33	W	33	CHỮ W HOA LA-TIN	2,3,4	
34	X	34	CHỮ X HOA LA-TIN	1,2,3,4	
35	Y	35	CHỮ Y HOA LA-TIN	2,3,4	
36	Z	36	CHỮ Z HOA LA-TIN	1,2,3,4	
37	a	37	CHỮ A THƯỜNG LA-TIN	3,4	Nét mạnh hơn; xem điều 13
38	b	38	CHỮ B THƯỜNG LA-TIN	3,4	Nét mạnh hơn; xem điều 13
39	c	39	CHỮ C THƯỜNG LA-TIN	3,4	Nét mạnh hơn; xem điều 13

Bảng 3 - Bảng tra các ký tự OCR-VN (tiếp theo)

TT	Hình ký tự	Bản vẽ số	Tên ký tự	Tập số	Ghi chú
40	d	40	CHỮ D THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
41	e	41	CHỮ E THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
42	f	42	CHỮ F THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
43	g	43	CHỮ G THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
44	h	44	CHỮ H THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
45	i	45	CHỮ I THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
46	j	46	CHỮ J THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
47	k	47	CHỮ K THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
48	l	48	CHỮ L THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
49	m	49	CHỮ M THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
50	n	50	CHỮ N THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
51	o	51	CHỮ O THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
52	p	52	CHỮ P THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
53	q	53	CHỮ Q THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
45	i	45	CHỮ I THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
46	j	46	CHỮ J THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13

Bảng 3 - Bảng tra các ký tự OCR-VN (tiếp theo)

TT	Hình ký tự	Bản vẽ số	Tên ký tự	Tập số	Ghi chú
47	k	47	CHỮ K THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
48	l	48	CHỮ L THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
49	m	49	CHỮ M THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
50	n	50	CHỮ N THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
51	o	51	CHỮ O THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
52	p	52	CHỮ P THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
53	q	53	CHỮ Q THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
54	r	54	CHỮ R THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
55	s	55	CHỮ S THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
56	t	56	CHỮ T THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
57	u	57	CHỮ U THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
58	v	58	CHỮ V THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
59	w	59	CHỮ W THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
60	x	60	CHỮ X THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
61	y	61	CHỮ Y THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
62	z	62	CHỮ Z THƯỜNG LA-TIN	3,4	Nét mảnh hơn; xem điều 13
63	*	63	DẤU HOA THỊ	2,3,4	

Bảng 3 - Bảng tra các ký tự OCR-VN (tiếp theo)

TT	Hình ký tự	Bản vẽ số	Tên ký tự	Tập số	Ghi chú
64	+	64	DẤU CỘNG	1, 2,3,4	
65	-	65	DẤU TRỪ	2,3,4	
66	=	66	DẤU BẰNG	2,3,4	
67	/	67	DẤU SỎ CHÉO	2,3,4	
68	.	68	DẤU CHẤM	2,3,4	
69	,	69	DẤU PHẨY	2,3,4	Nét chéo cần phải ở phía dưới đường cơ sở của các chữ hoa; xem điều 13.2 và 13.5
70	:	70	DẤU HAI CHẤM	3,4	
71	;	71	DẤU CHẤM PHẨY	3,4	Nét chéo cần phải ở phía dưới đường cơ sở của các chữ hoa; xem điều 13.2 và 13.5
72	“	72	DẤU NHÁY KÉP	3,4	Trong những ứng dụng phi nhận dạng quang học, có thể thay bằng dấu hai chấm trên (107) khi phải in cả hai dấu đó với cùng một ký tự
73	‘	73	DẤU NHÁY ĐƠN	3,4	Trong những ứng dụng phi nhận dạng quang học, có thể thay bằng dấu sắc (108) khi phải in cả hai dấu đó với cùng một ký tự
74	—	74	DẤU GẠCH DƯỚI	3,4	Đối với nhận dạng quang học, chỉ nên dùng như ký tự độc lập và không nên in dưới một ký tự khác; xem điều 10
75	?	75	DẤU CHẤM HỎI	3,4	
76	!	76	DẤU CHẤM THAN	3,4	
77	(77	DẤU MỞ NGOẶC ĐƠN	3,4	

Bảng 3 - Bảng tra các ký tự OCR-VN (tiếp theo)

TT	Hình ký tự	Bản vẽ số	Tên ký tự	Tập số	Ghi chú
78)	78	DẤU ĐÓNG NGOẶC ĐƠN	3,4	
79	<	79	DẤU NHỎ HƠN	1,2,3,4	
80	>	80	DẤU LỚN HƠN	1,2,3,4	
81	[81	DẤU MỞ NGOẶC VUÔNG	3,4	
82]	82	DẤU ĐÓNG NGOẶC VUÔNG	3,4	
83	%	83	DẤU PHẦN TRĂM	3,4	Nét mảnh hơn; xem điều 13
84	#	84	DẤU THẰNG	3,4	Nét mảnh hơn; xem điều 13
85	&	85	DẤU VÀ	3,4	
86	@	86	DẤU ĐỊA CHỈ	3,4	Nét mảnh hơn; xem điều 13
87	^	87	DẤU MŨ NHỌN	3,4	Trong những ứng dụng phi nhận dạng quang học, có thể thay bằng dấu mũ (110) khi phải in cả hai dấu đó với cùng một ký tự
88	α	88	DẤU TIỀN TỆ	3,4	
89	£	89	DẤU ĐỒNG BẰNG	3,4	
90	\$	90	DẤU ĐÔ LA	3,4	
91		91	DẤU SỎ DỌC	3,4	Xem điều 12
92	!	92	DẤU SỎ DỌC ĐỨT NÉT	1,2,3,4	Xem điều 12
93	\	93	DẤU SỎ CHÉO NGƯỢC	3,4	

Bảng 3 - Bảng tra các ký tự OCR-VN (tiếp theo)

TT	Hình ký tự	Bản vẽ số	Tên ký tự	Tập số	Ghi chú
94	Ä	94	CHỮ A HOA LA-TIN CÓ DẤU HAI CHẤM	4	Trong những ứng dụng nhận dạng quang học, nếu có thể thì thay bằng hai ký tự hoa A (11) và E (15)
95	Å	95	CHỮ A HOA LA-TIN CÓ DẤU KHUYÊN Ở TRÊN	4	Giữ cho tương hợp nhưng không nên sử dụng trong những ứng dụng mới; xem điều 9.3
96	Æ	96	CHỮ GHÉP AE HOA LA-TIN	4	
97	Ö	97	CHỮ O HOA LA-TIN CÓ DẤU HAI CHẤM	4	Trong những ứng dụng nhận dạng quang học, nếu có thể thì thay bằng hai ký tự hoa O (25) và E (15)
98	Ø	98	CHỮ O HOA LA-TIN CÓ DẤU SỎ CHÉO	4	
99	Ü	99	CHỮ U HOA LA-TIN CÓ DẤU HAI CHẤM	4	Trong những ứng dụng nhận dạng quang học, nếu có thể thì thay bằng hai ký tự hoa U (31) và E (15)
100	IJ	100	CHỮ GHÉP IJ HOA LA-TIN (HÀ LAN)	4	Trong những ứng dụng nhận dạng quang học, nên tách riêng thành ký tự I (19) và J (20)
101	Ñ	101	CHỮ N HOA LA-TIN CÓ DẤU NGÃ	4	Giữ cho tương hợp nhưng không nên sử dụng trong những ứng dụng mới; xem điều 9.3
102	å	102	CHỮ A THƯỜNG LA-TIN CÓ DẤU KHUYÊN Ở TRÊN	4	Nét mảnh hơn; xem điều 13. Giữ cho tương hợp nhưng không nên sử dụng trong những ứng dụng mới; xem điều 9.3
103	æ	103	CHỮ GHÉP AE THƯỜNG LA-TIN	4	Nét mảnh hơn; xem điều 13
104	ø	104	CHỮ O THƯỜNG LA-TIN CÓ DẤU SỎ CHÉO	4	Nét mảnh hơn; xem điều 13
105	ij	105	CHỮ GHÉP IJ THƯỜNG LA-TIN (HÀ LAN)	4	Nét mảnh hơn; xem điều 13
106	ß	106	CHỮ GHÉP S THƯỜNG (ĐỨC)	4	Nét mảnh hơn; xem điều 13

Bảng 3 - Bảng tra các ký tự OCR-VN (tiếp theo)

TT	Hình ký tự	Bản vẽ số	Tên ký tự	Tập số	Ghi chú
107	"	107	DẤU HAI CHẤM TRÊN	4	Xem điều 9 và 13
108	'	108	DẤU SẮC	4	Xem điều 9 và 13
109	`	109	DẤU HUYỀN	3,4	Xem điều 9 và 13
110	^	110	DẤU MŨ	4	Xem điều 9 và 13
111	~	111	DẤU NGÃ	3,4	Xem điều 9 và 13
116	—	116	DẤU GẠCH DƯỚI LIÊN	4	
117		Không có	DẤU CÁCH	1,2,3,4	Dấu cách là một ký tự không in được. (xem điều 11). Không phải tất cả các thiết bị nhận dạng quang học đều cần nhận biết dấu cách.
118	§	118	DẤU ĐẦU MỤC	4	
119	¥	119	DẤU ĐỒNG YÊN	4	
120	■	120	DẤU XÓA MỘT KÝ TỰ	5	Xem điều 7.5
121	--	121	DẤU XÓA NHÓM KÝ TỰ	5	Xem điều 7.5
122	.	123	DẤU NẶNG (QUỐC NGỮ)	4	Xem điều 9 và 13
123	?	176	DẤU HỎI (QUỐC NGỮ)	4	Nét mảnh hơn; xem điều 13
124	Â	163	CHỮ A HOA LA-TIN CÓ DẤU MŨ (QUỐC NGỮ)	4	
125	Ă	164	CHỮ A HOA LA-TIN CÓ DẤU TRẮNG (QUỐC NGỮ)	□	

Bảng 3 - Bảng tra các ký tự OCR-VN (tiếp theo)

TT	Hình ký tự	Bản vẽ số	Tên ký tự	Tập số	Ghi chú
126	Đ	130	CHỮ D HOA LA-TIN CÓ GẠCH NGANG (QUỐC NGỮ)	□	
127	Ê	165	CHỮ E HOA LA-TIN CÓ DẤU MŨ (QUỐC NGỮ)	□	
128	Ô	166	CHỮ O HOA LA-TIN CÓ DẤU MŨ (QUỐC NGỮ)	4	
129	Ơ	167	CHỮ O HOA LA-TIN CÓ DẤU RÂU (QUỐC NGỮ)	4	
130	Ư	168	CHỮ U HOA LA-TIN CÓ DẤU RÂU (QUỐC NGỮ)	4	
131	â	169	CHỮ A THƯỜNG LA-TIN CÓ DẤU MŨ (QUỐC NGỮ)	4	Nét mảnh hơn; xem điều 13
132	ă	170	CHỮ A THƯỜNG LA-TIN CÓ DẤU TRẮNG (QUỐC NGỮ)	4	Nét mảnh hơn; xem điều 13
133	đ	140	CHỮ D THƯỜNG LA-TIN CÓ GẠCH NGANG (QUỐC NGỮ)	4	Nét mảnh hơn; xem điều 13
134	ê	171	CHỮ E THƯỜNG LA-TIN CÓ DẤU MŨ (QUỐC NGỮ)	4	Nét mảnh hơn; xem điều 13
135	ô	172	CHỮ O THƯỜNG LA-TIN CÓ DẤU MŨ (QUỐC NGỮ)	4	Nét mảnh hơn; xem điều 13
136	ơ	173	CHỮ O THƯỜNG LA-TIN CÓ DẤU RÂU (QUỐC NGỮ)	4	Nét mảnh hơn; xem điều 13
137	ư	174	CHỮ U THƯỜNG LA-TIN CÓ DẤU RÂU (QUỐC NGỮ)	4	Nét mảnh hơn; xem điều 13

9 Sử dụng các dấu phụ

9.1 Phân loại các dấu phụ

Một số dấu phụ được chọn đưa vào bộ ký tự OCR-VN để khi cần thiết có thể dùng phương pháp tổ hợp tạo ra tất cả các ký tự quốc ngữ và phần lớn các ký tự riêng của nhiều thứ chữ khác thuộc hệ Latin.

TCVN 6437 : 1998

9.1.1 Các dấu phụ quốc tế

Dấu hai chấm trên (DIAERESIS): "

Dấu mũ (CIRCUMFLEX): ^

Dấu râu (CEDILLA): ¸

9.1.2 Các dấu phụ quốc ngữ (dấu thanh)

Dấu huyền (GRAVE): `

Dấu hỏi (HOOK ABOVE): ˇ

Dấu ngã (TILDE): ~

Dấu sắc (ACUTE): ´

Dấu nặng (DOT BELOW): ˙

a`aãáạ âầẩẩẩẩậ ăẳẳẳẳẳ
eèẻẻẻẻẻ êềềềềểểểểểể ỉỉỉỉỉ
oòỏỏỏỏỏ ôồồồồồộộộộộ ơờờờờợợợợợ
uùủủủủủ ưừừừừửửửửử yýýýý
A`A`A`A`A` Â`Â`Â`Â`Â` Ă`Ă`Ă`Ă`Ă`
E`E`E`E`E` Ê`Ê`Ê`Ê`Ê` Ỉ`Ỉ`Ỉ`Ỉ`Ỉ`
O`O`O`O`O` Ô`Ô`Ô`Ô`Ô` Ơ`Ơ`Ơ`Ơ`Ơ`
U`U`U`U`U` Ư`Ư`Ư`Ư`Ư` Y`Y`Y`Y`Y`

Hình 3 - Các thí dụ về ký tự tổ hợp với dấu thanh

9.1.3 Lưu ý

9.1.3.1 Các ký tự tổ hợp có thể in ra bằng một thao tác hoặc hai thao tác; tuy nhiên phương pháp in ra ký tự là không thuộc phạm vi của tiêu chuẩn này.

9.1.3.2 Trong các bản vẽ ở phụ lục của tiêu chuẩn này, các ký tự mang dấu phụ ở dạng viết thường và viết hoa đều được xác định theo phương pháp dựng sẵn. Tuy nhiên trong thực tiễn mỗi ký tự đó cũng có thể được xác định bằng phương pháp tổ hợp một dấu phụ với một ký tự.

9.1.3.3 Những bản vẽ trong phần phụ lục của tiêu chuẩn này thể hiện vị trí của các dấu phụ trong tương quan với các ký tự mang chúng và trong giới hạn của ô chữ.

9.1.3.4 Đối với các ứng dụng nhận dạng quang học phải xử lý nhiều dòng văn bản, khi sử dụng các dấu phụ tổ hợp với các ký tự có phần nhô lên thì cần chú ý đến khoảng cách giữa các dòng. Đặc biệt, khoảng cách thông dụng 4,23 mm (0,167 in) giữa hai dòng có thể sẽ không đủ rộng nếu dòng dưới có chứa dấu phụ ở trên mà dòng trên nó lại chứa ký tự có phần thụt xuống (như chữ 'j') hoặc chứa dấu phụ ở phía dưới (dấu "nặng", "râu") hoặc chứa dấu "gạch dưới".

9.1.3.5 Đối tượng của phần lớn các ứng dụng nhận dạng quang học là những văn bản có nhiều dòng chữ in. Việc nhận dạng văn bản sẽ được thuận tiện nếu tuân thủ một số điều kiện sau đây trong khâu in, dù các ký tự được in ra bằng phương pháp tổ hợp hoặc dựng sẵn. Trước hết dấu "nặng", dấu "râu", hoặc đuôi của những ký tự "f", "g", "j", "p", "q", "y" ở dòng trên không được dính với các dấu hoặc các ký tự nhô cao khác của dòng dưới. Ngoài ra các ký tự "Đ", "đ", "O", "o", "U", "u" không được dính sang các ký tự nằm sát bên chúng trong một chữ quốc ngữ.

9.2 Qui tắc tổ hợp ký tự quốc ngữ

9.2.1 Qui tắc chung

Trong mỗi ô ký tự quốc ngữ chỉ có thể tổ hợp nhiều nhất là một ký tự nguyên âm với một dấu thanh.

9.2.2 Đối với tập ký tự mở rộng

Có thể tổ hợp các nguyên âm chữ hoa và chữ thường trong tập ký tự mở rộng với một dấu thanh bất kỳ, trừ ký tự "i thường La-tin" chỉ có thể tổ hợp với dấu "nặng".

9.2.3 Đối với tập ký tự đầy đủ

Mọi ký tự nguyên âm thuần Việt (từ 124 đến 136 trong bảng 3) đều có thể tổ hợp với bất kỳ dấu thanh nào. Các ký tự khác thì không được tổ hợp với các dấu thanh.

9.3 Thay thế các ký tự dựng sẵn

Có thể thay thế một số hoặc toàn bộ các ký tự dựng sẵn bằng phương pháp tổ hợp các ký tự chữ cái với các dấu phụ có trong các tập ký tự OCR-VN.

Các ứng dụng về nhận dạng quang học nói chung cần nhận biết được cả hai dạng ký tự dựng sẵn và tổ hợp.

10 Sử dụng ký tự gạch dưới

Ký tự gạch dưới được sử dụng trong các ứng dụng nhận dạng quang học như một ký tự đứng riêng, và không được tổ hợp với (có nghĩa là không được in ở phía dưới) một ký tự nào khác.

TCVN_1997

Hình 4 – Ví dụ về cách sử dụng ký tự gạch dưới

11 Dấu cách

Ký tự dấu cách là một vị trí được cố tình để trống trong một dòng in. Khi in bằng các thiết bị có bước in cố định thì độ rộng danh định của dấu cách là bằng độ rộng của bước in (thí dụ 2,54 mm nếu mật độ in là 10 ký tự trên 25,4 mm). Khi in bằng các máy in có bước in biến đổi thì độ rộng danh định của dấu cách sẽ bằng độ rộng của ký tự khả thi rộng nhất.

12 Dấu số đứng và dấu số đứt nét

Dấu "số đứng" (91) và dấu "số đứt nét" (92) đều là các vạch thẳng đứng nhưng chúng khác nhau ở độ cao tối thiểu (xem bảng 4).

Bảng 4 - Kích thước của các dấu số đứng

Đơn vị đo	mm			inch		
	I	III	IV	I	III	IV
Cỡ						
Độ cao của dấu số đứng	3,2	–	–	0,126	–	–
Độ cao tối thiểu của dấu số đứt nét	3,7	5,0	5,6	0,146	0,196	0,220
Độ rộng danh định của dấu số đứng và dấu số đứt nét	0,35	0,38	0,50	0,014	0,015	0,019

13 Xác định hình dạng ký tự

13.1 Các bản vẽ tham khảo

Hình dạng và kích thước của các ký tự OCR-VN được xác định trong các bản vẽ tham khảo của phụ lục kèm theo tiêu chuẩn này.

Những bản vẽ đó thể hiện chi tiết trên giấy kẻ ô milimet hình dạng và kích thước của các ký tự OCR-VN được phóng to để dễ tham khảo. Những ký tự đó cũng có thể lưu trữ dưới dạng PostScript hoặc TrueType trong các thiết bị công nghệ thông tin nhằm tiện lợi cho xử lý và sử dụng.

Các điểm trên bản vẽ có thể xác định được với độ chắc chắn tới một nửa ô kẻ (tương đương 0,01 mm trên kích cỡ thực tế của ký tự in ra theo tỷ lệ 1:1).

13.2 Ký tự có độ rộng nét chữ cố định, cỡ I

13.2.1 Hình dạng danh định được in ra của mỗi ký tự được xác định bởi đường chủ đạo và độ rộng nét chữ danh định của ký tự đó.

Độ rộng nét chữ danh định ở cỡ I là:

0,35 mm (0,014 in) cho hầu hết các ký tự

0,31 mm (0,012 in) cho tất cả các chữ thường và ba ký tự #, %, @

13.2.2 Cần chú ý đặc biệt khi thiết kế và thực hiện chữ in để có thể đạt được bản in có đầy đủ những nét cuối và góc chữ. Điều này là đặc biệt quan trọng đối với các góc vuông của các chữ hoa B và D.

13.3 Ký tự có độ rộng nét chữ cố định, cỡ III

13.3.1 Hình dạng danh định được in ra của mỗi ký tự được xác định bởi đường chủ đạo và độ rộng nét chữ danh định của ký tự đó. Độ rộng nét chữ danh định ở cỡ III là:

0,38 mm (0,015 in) cho hầu hết các ký tự

0,33 mm (0,013 in) cho tất cả các chữ thường và ba ký tự #, %, @

13.3.2 Cần chú ý khi thiết kế và thực hiện chữ in để có thể đạt được bản in có đầy đủ những nét cuối và góc chữ như của bộ chữ tham khảo. Điều này là đặc biệt quan trọng đối với các góc vuông của các chữ hoa B và D.

13.4 Ký tự có độ rộng nét chữ cố định, cỡ IV

13.4.1 Hình dạng danh định của mỗi ký tự được xác định bởi đường chủ đạo và độ rộng nét chữ danh định của nó. Đường chủ đạo ở cỡ IV được chế biến từ đường chủ đạo tương ứng ở cỡ I (xem điều 13.2 và những hình vẽ tham khảo) bằng cách phóng đại theo tỷ lệ 1,5 lần.

VÍ DỤ: Một đường chủ đạo có độ cao là 2,40 mm ở cỡ I sẽ có độ cao ở cỡ IV là $1,5 \times 2,40 = 3,60$ mm và cứ tương tự như vậy. Độ rộng nét chữ danh định ở cỡ IV là:

0,50 mm (0,020 in) đối với hầu hết các ký tự

0,44 mm (0,017 in) đối với mọi chữ thường và ba ký tự #, %, @

13.4.2 Các nét cuối và góc chữ ở cỡ IV không thể vẽ một cách máy móc bằng cách phóng đại từ cỡ I lên 1,5 lần vì tỷ lệ giữa độ rộng nét chữ danh định của cỡ IV và cỡ I không chính xác là 1,5 lần.

13.5 Ký tự có độ rộng nét chữ biến đổi, cỡ I

Những ký tự thuộc kiểu chữ có độ rộng nét chữ biến đổi được thiết kế với những thay đổi nhỏ về độ rộng nét chữ. Tuy nhiên, các độ rộng nét chữ luôn gần với giá trị danh định là 0,35 mm (0,014 in) đối với các chữ số và các chữ hoa, và 0,31 mm (0,012 in) đối với các chữ thường và ba ký tự #, %, @.

TCVN 6437 : 1998

14 In các kiểu chữ có độ rộng nét chữ cố định và biến đổi

Muốn in các chữ có độ rộng nét chữ biến đổi và để đạt được hình thức mỹ thuật thỏa mãn nhất thì thiết bị in cần phải có khả năng in rõ nét các góc chữ và thay đổi được độ rộng nét chữ. Những đặc tính này là không cần thiết đối với những kiểu chữ có độ rộng nét chữ cố định, mặc dù cũng cần cố gắng đặc biệt để tạo ra các góc chữ rõ nét cho các chữ hoa B và D.

15 Minh họa bộ ký tự OCR-VN

```
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
* + - = / . , ; ' ' _
? ! ( ) < > [ ] % # & @ ^
¤ £ $ | | \ Ä Å Æ I J Ñ Ö Ø
Û ä æ | j ø ß § Å Ä Đ Ê Ô Ó
Ů á â ã đ ê ô ó Ů ' ' ' ^
~ . , { } m _ ■ --
```

Hình 5 - Các ký tự OCR-VN ở cỡ I với tỉ lệ 1:1

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
 * + - = / . , : ; “ ‘ _
 ? ! () < > [] % # & @ ^
 ¤ £ \$ | ; \ Ä Å Æ I J Ñ Ö Ø
 Ü å æ I j ø ß § Â Ã Đ Ê Ô Ơ
 U' â ã đ ê ô ơ u' " ' ` ^
 ~ . , { } m _ ■ --

Hình 6 – Các ký tự OCR-VN ở cỡ I với tỉ lệ 4:1

Phụ lục A
(Tham khảo)

Hình dạng và kích thước các ký tự

Hình dạng và kích thước của các ký tự OCR - VN được xác định trong 257 bản vẽ của phụ lục này. Các bản vẽ thể hiện chi tiết trên giấy milimet hình dạng và kích thước của các ký tự được phóng to để dễ tham khảo theo tỷ lệ 70:1. Những ký tự này có thể được lưu trữ dưới dạng PostScript hoặc TrueType trong các thiết bị công nghệ thông tin nhằm tiện lợi cho xử lý và sử dụng. Các điểm trên bản vẽ có thể được xác định với độ chắc chắn tới một nửa ô kẻ (tương đương 0,01 mm trên kích thước thực tế của ký tự in ra theo tỷ lệ 1 : 1).

CHÚ THÍCH:

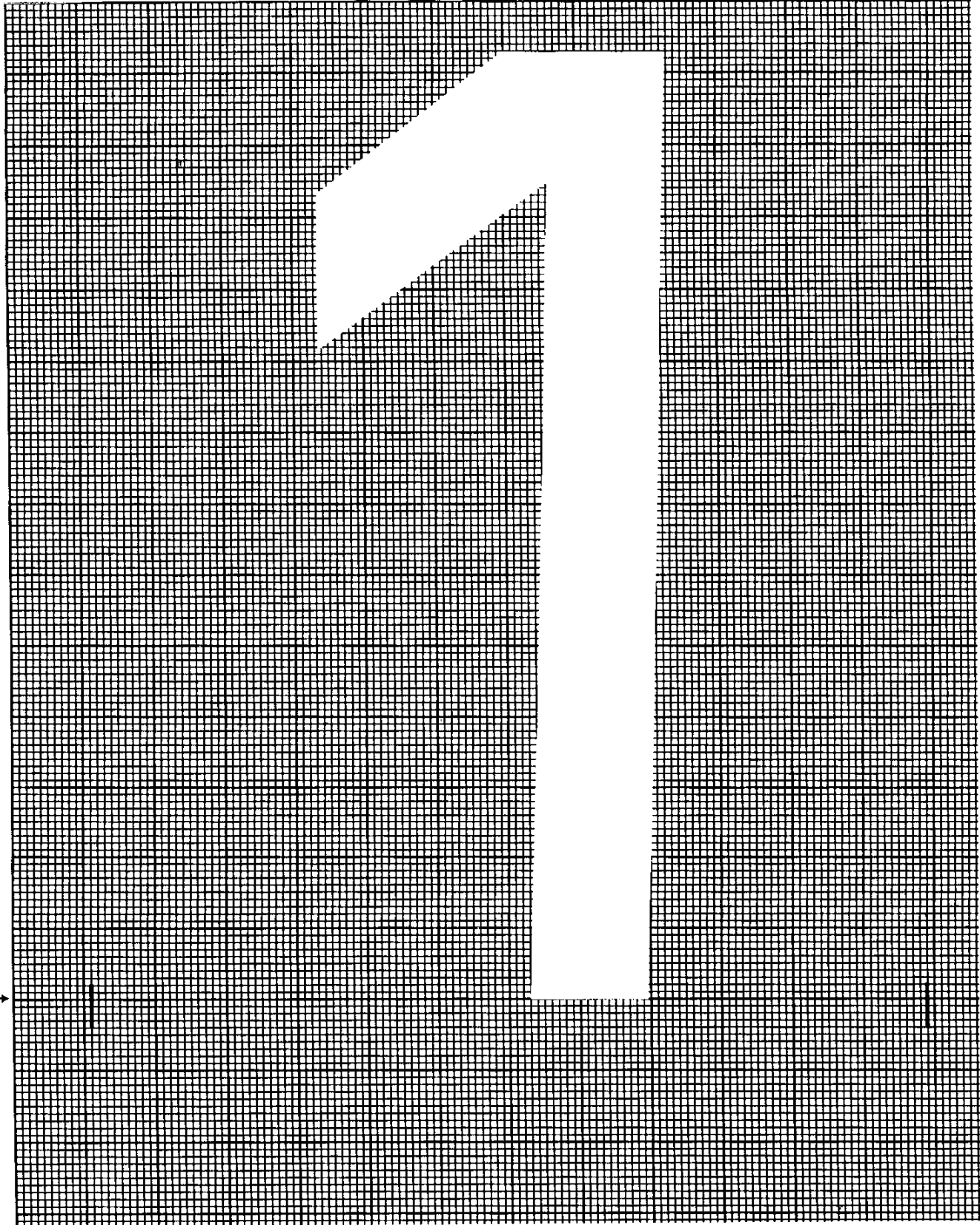
- 1) Trong các trang sau đây chỉ giới thiệu một số bản trong bộ 257 bản vẽ.
- 2) Khi có nhu cầu tham khảo trọn bộ bản vẽ, xin mời liên hệ với:

Trung tâm Tiêu chuẩn - Chất lượng

Đường Hoàng Quốc Việt - Quận Cầu Giấy - Hà Nội

ĐT : (84 4) 8 344 191

(84 4) 8 344 269



Bản vẽ số 1

