UTKAST

2001-01-19

# Stöd för samiska språk i svensk databehandling

Support for Sami languages in Swedish data processing

Innehållsförteckning

This publication ia available from the Statskontoret web page (www.statskontoret.se).

Stöd för samiska språk i svensk databehandling
Support for Sami languages in Swedish data processing

1 Omfattning

Denna Statskontorets tekniska norm anger speciella krav på datorer för hantering av alfabetena för de samiska språk som används i Sverige. Normen är i första hand avsedd som referensdokument för upphandlingssituationer.

Kraven är tillämpliga på tre vanliga datormiljöer, nämligen ISO/IEC 8859-kompatibel (t.ex. flertalet Unix-system), Microsoft "Windows" och Apple "Macintosh", för såväl 8-bits som fler-oktett kodningsscheman.

Anmärkning: "Macintosh" är ett varumärke ägt av Apple Computer, Inc. "Windows" är ett varumärke ägt av Microsoft Corporation. Dessa varumärken är registrerade i Sverige och andra länder.

2 Motivering

Statskontorets tekniska norm nr 35 (TN 35) specificerar 8-bits kodningsscheman för samiska språk vid databearbetning och informationsutbyte. För implementeringar måste flera faktorer bestämmas utöver kodningsschemat, framförallt tangentbordslayout. Vidare behöver databehandling i fler-oktettmiljöer behandlas.

2.1 Speciella faktorer för tangentbord

För tangentbordsanvändning vid skrivning av samiska språk i Sverige är två olika situationer aktuella:
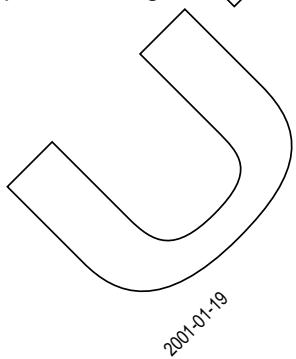
1 Scope of standard

This Statskontoret Technical Standard defines special computer requirements for handling the alphabets of the Sami languages used in Sweden. The Standard is primarily intended as a reference document for purchasing situations.

The requirements are applicable to three common computer environments, namely ISO/IEC 8859-compatible (e.g. most Unix systems), Microsoft "Windows", and Apple "Macintosh", for both 8-bit and multi-octet coding schemes.

Acknowledgment: "Macintosh" is a trademark of Apple Computer, Inc. "Windows" is a trademark of Microsoft Corporation. These trademarks are registered in Sweden and other jurisdictions.

2 Rationale

The Statskontoret Technical Standard no. 35 (TN 35) specifies 8-bit coding schemes for Sami-language data processing and information interchange. For actual computer implementations several factors have to be decided in addition to the coding scheme, most important among them the keyboard layout. Also data processing in multi-byte environments needs to be dealt with.

2.1 Special keyboard factors

For Sami-language keyboard input in Sweden, two different situations must be considered:

1. Production of Sami-language text. In this case a keyboard suitable for easy entry of all the Sami letters is desirable. It must also be possible to input – although not necessarily in a completely convenient manner – all the Swedish letters that do not exist in the Sami alphabets. Further it should be possible to easily input the Norwegian letters Ææ and Øø, due to Sami cross-border aspects.

   This computer situation is the normal one for some Sami-speaking individuals, but it also applies to authorities needing to input larger amounts of texts in Sami.

2. Production of Swedish-language text, but with the need to write also Sami names, personal and/or geographical. This requires the addition of the specifically Sami letters to the normal

2001-01-19

Swedish keyboard, and further the letters Ææ and Øø.

This computer situation applies to most authorities located in the Sami-language area, and to several authorities outside it.

Both cases could in theory be handled by a single keyboard layout. Considering ergonomy, as well as established keyboard situations, different keyboard layouts are however needed.

3 Sami languages in Sweden

Different Sami languages exist in Sweden, and within them different dialects. From an orthography point of view three languages/dialects need to be considered, namely Southern Sami, Lule Sami and Northern Sami.

The orthographies of these languages are not officially defined in Sweden. Authoritative sources, used in the production of textbooks, however all base the repertoire of letters on the Swedish alphabet, although some of its letters – like Q and W – are not used in Sami words. Special Sami letters are also needed:

For Southern Sami: No special letters
For Lule Sami: Áá Ńń
For Northern Sami: Áá Čč Đđ Ŋŋ Šš Ŧŧ Žž

The need for the letter Ńń in Lule Sami is the subject of some discussion. Since it represents the same sound as written by Ŋŋ in Northern Sami there are recommendations to use that letter instead. Also, in data processing the letter Ññ is sometimes used in place of Ńń.

It can be noted that the recent Swedish legislation on minority languages does not make any difference between the Sami languages. From a data processing point of view this is also not necessary: when the Swedish alphabet is complemented with the Northern Sami letters, the Southern and Lule alphabets are also covered (with the exception of the letter Ńń which, as described above, can be represented in alternative ways).

It can further be noted that the letter requirements of the other recognized regional minority languages in Sweden, Finnish and Meänkieli ("Tornedalsfinska") are covered by the Swedish alphabet when complemented with the Northern Sami letters.

2001-01-19

4 Requirements

4.1 Character set

The repertoire to be handled by the operating system is required to contain as a minimum:

A: The characters of registration no. 6 of the ISO "International register of coded character sets to be used with escape sequences", ISO-IR 6.

and

B: The letters Áá Čč Đđ Éé Ŋŋ Šš Ŧŧ Üü Žž Ææ Øø Åå Ää Öö

For a specification according to ISO/IEC 10646-1 of the letters, see Annex B.

Notes:

1. The coding scheme of ISO-IR 6 is identical to that of ASCII.

2. At the time of publication of this standard, the register is available at the Internet site www.itscj.ipsj.or.jp/ISO-IR

In a purchasing situation, characters in addition to those of items A and B will normally be specified.

Such a specification can take the form of en enumeration of the additional characters, or a reference to a specific coding scheme e.g. "The repertoire of Statskontoret TN 35, Windows-environment" (the characters of items A and B form subsets of all the three coding schemes of TN 35).

4.2 Character coding

4.2.1 8-bit coding

For 8859-type environments, coding according to one of the following schemes is required:

ISO/IEC 8859-4   (Latin-4)
ISO/IEC 8859-10 (Latin-6)
ISO-IR 6 + ISO-IR 204   (see Statskontoret TN 35)

For Windows and Macintosh environments, coding according to the schemes specified in Statskontoret TN 35 is required.

4.2.2 Multiple-octet coding

Coding according to ISO/IEC 10646-1 is required.

Note: To the extent relevant for normal Swedish data processing, this scheme is identical to "Unicode" from its version 2.

4.2.3 Identification of character coding

In data interchange situations, identification of the coding used will normally be done by mutual agreement between sender and receiver of data, or by separate documentation accompanying the data.
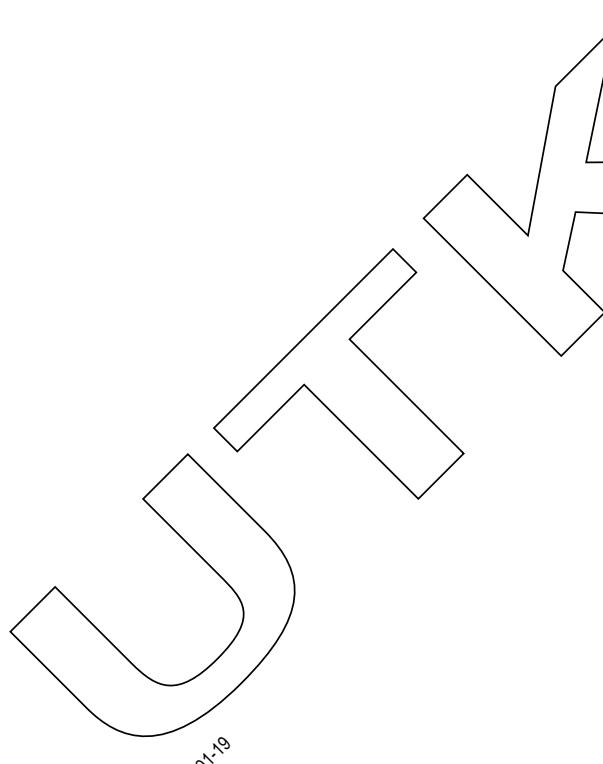
2001-01-19

If however automatic identification of the character set used for a specific computer file is desired, the methods of ISO/IEC 2022 or ISO/IEC 4873 are recommended for the ISO-standardized 8-bit character sets; see those standards for detailed information. The following identification sequences will then apply (ESC designates the escape character, hex 1B; complete hex sequences in parentheses):

ISO/IEC 8859-4:   ESC (B ESC - D   (1B 28 42 1B 2D 44)
ISO/IEC 8859-10:   ESC (B ESC - V   (1B 28 42 1B 2D 56)
ISO-IR 6 + 204:   ESC (B ESC - h   (1B 28 42 1B 2D 68)

The TN 35 Windows- and Macintosh-environment schemes have no identification sequences (since they are not conformant with ISO coding standards).

ISO/IEC 10646 coding will have different identification sequences depending on what coding form and implementation level that has been chosen (see the standard for further explanations):

UCS-2 level 2:

UCS-2 level 3:

UCS-4 level 2:

UCS-4 level 3:

4.3   Keyboards

4.3.1   Background on keyboard layouts

A large number of different keyboard layouts exist, both as national standards and as de facto standards. This is natural, since the principles for character allocation to keys inherited from typewriters have become completely inadequate in data processing, necessitating multiple keyboard solutions.

The framework standard for keyboard layout is the multipart ISO/IEC 9995. The present edition of this standard defines two fundamental layout concepts, namely "level" and "group".

The "level" relates to the conventional shift-key function, except that the 9995 permits also a third level, i.e. another level accessible with a key (or a key combination) different from the shift key.

The "group" concept makes possible two or more alternative layouts for the same physical keyboard. For example, on a Swedish Microsoft keyboard the character @ can be considered as located in a group different from the "primary" (default) one. It shall be noticed, however, that Microsoft does not use the concept "group", instead employing the term "AltGr state".

Note: Within the ISO/IEC 9995 framework, the @ character of the Swedish keyboard could also be considered as located in a third level of the digit 2 key, the AltGr key in such a case serving as a level 3 shift key.
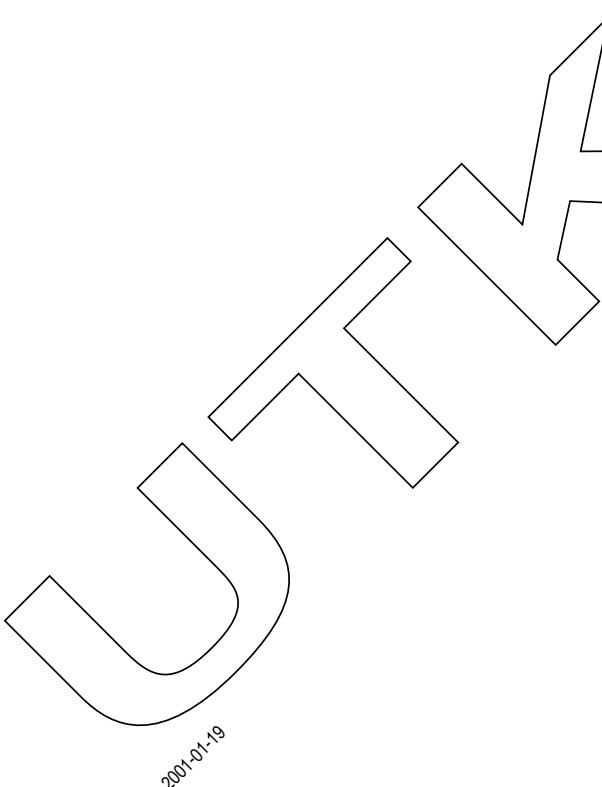
The Swedish Microsoft keyboard contains no upper-shift characters in AltGr state, but some other keyboards do. For instance, with the AltGr key depressed, the "A" key of a Canadian keyboard generates the character æ in lower shift and Æ in upper shift.

Due to the differences in terminology, in the following text the terms "group" and "AltGr state" are not used, but instead the terms "primary" and "alternative" layout. How the alternative layout is invoked will be system-dependent, and is not specified. It is however assumed that for Microsoft-type keyboards the alternative layout is invoked by the AltGr key, and for Macintosh keyboards by the "Option" key.

It shall also be noticed that in most present-day computer systems a keyboard input does not generate a character code, but only a key-press identification (called "scan code" in PCs). The translation from this key identification into a character coding is done by "driver" software.

It is therefore possible to "update" keyboards simply by introducing new driver software in the systems. It is assumed that this will be the method used for conformance with this Technical Standard in some systems.

4.3.2 Scope of keyboard requirements

As described in clause 2.1 above, different keyboard layouts are desirable for the two situations "Swedish letters in Sami text" and "Sami letters in Swedish text". In both cases, as a minimum, all the characters specified in clause 4.1 are required to be available for input from the keyboard.

This Statskontoret Technical Standard does however not specify complete keyboard layouts, but only the positioning of the alphabetic keys in the primary and the alternative layouts. The positioning of the letters Éé Üü is however not specified; these can optionally either be made available "pre-composed" or input by diacritical mark – base letter sequences.

In a purchasing situation, required locations of the non-alphabetic characters (like the digits 0–9) and possibly also pre-composed letters Éé and/or Üü should be specified in addition to the letters contained in the layouts. This can be done explicitly, or by reference to more complete layouts.

4.3.3 Layout for Sami-language input

The required allocation of alphabetic characters is specified in Annex A, and illustrated below (the

2001-01-19
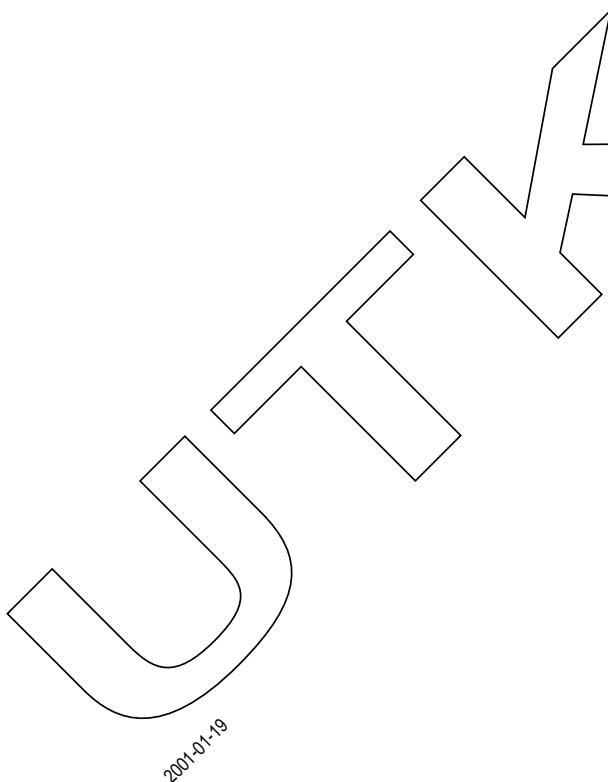
letters on the right part of the keys indicate characters available in the alternative layout):
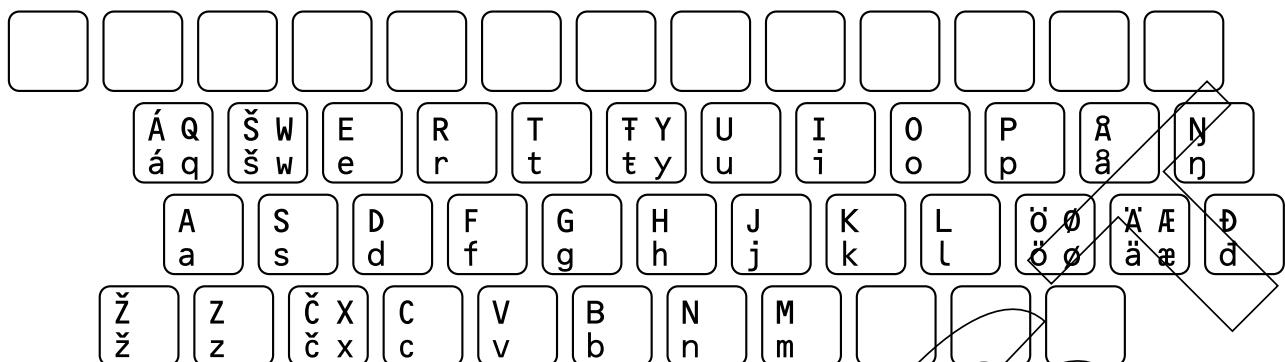
4.3.4 Layouts for Swedish-language input

Two different schemes are specified in Annex A and illustrated below.

The first scheme corresponds to present Swedish keyboards, with the Sami (and Norwegian) letters arranged in the alternative layout on the "graphically corresponding" Swedish letter keys:
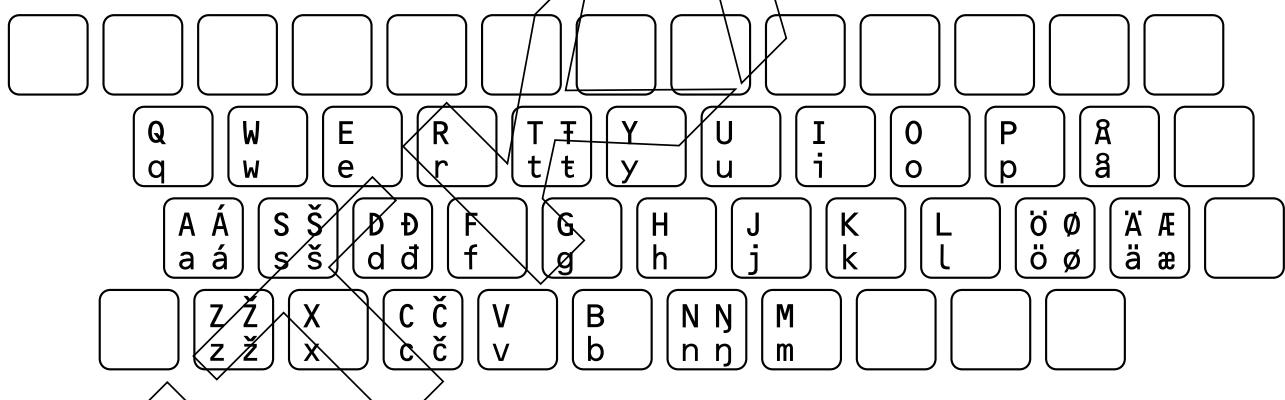
The second scheme conforms to the Swedish keyboard standard SS 66 22 41 version 2, with the Sami (and Norwegian) letters positioned according to that standard's "Group 2".
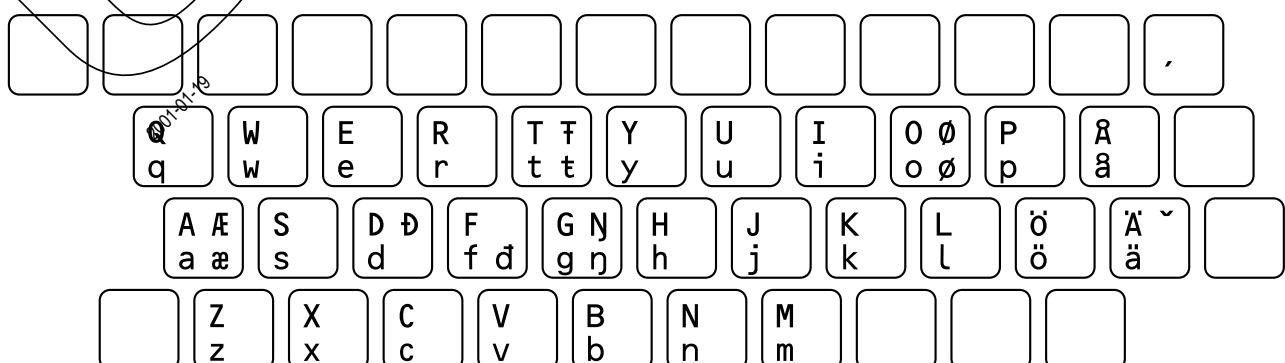
In this layout, the letters Áá Čč Šš Žž are not available directly. Instead, they must be "composed" by the use of the acute accent and the caron; located respectively in the primary layout in lower shift on the key to the right above the Å key, and in the alternative layout in upper shift on the Ä key.

4.4 Fonts

As a minimum, systems are required to provide support in three types of fonts for the complete character set given in a purchasing situation (i.e. the character set of clause 4.1 items A and B, and any additional characters specified).

The three types are fixed-spacing (e.g. "Courier"), variable-spacing non-serif (e.g. "Helvetica"), and variable-spacing serif (e.g. "Times"). In each font the variations regular, italic, bold and bold-italic are required.

Additional requirements can be specified in purchasing.

4.5 Sorting order

The operating system and any additional software being part of a specific purchasing is required to comply with the sorting order specified in Annex B.

Only the sorting order of the letters in the required set, i.e. the set according to clause 4.1 items A and B, is specified. Additional requirements will generally be made in purchasing.

2001-01-19
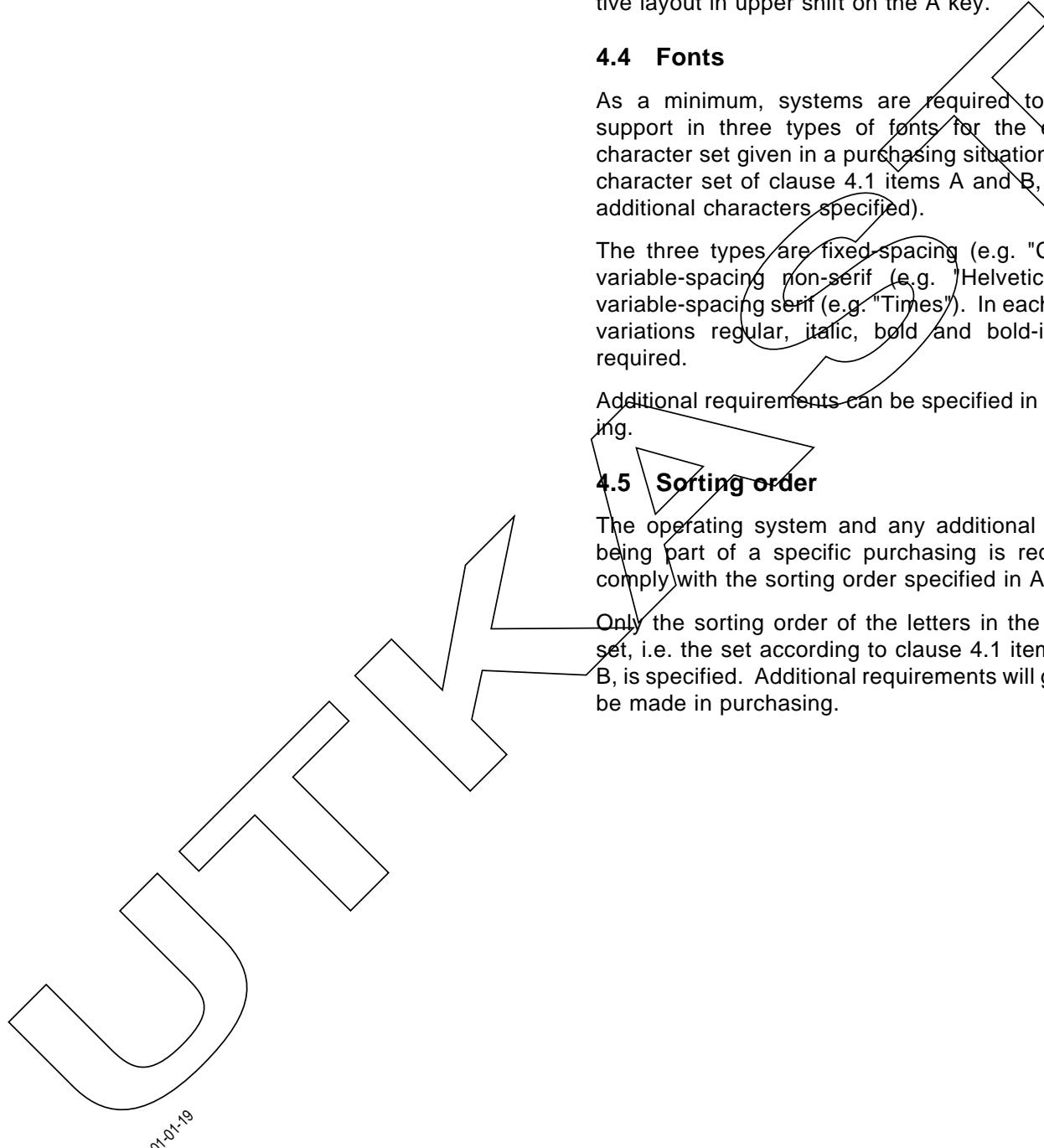
Annex A

Teckenplacering på tangentbord
Keyboard character allocation

I detta annex specificeras placeringen av tecken för de tre tangentbordsalternativ som anges i normens avsnitt 4.3.

For key identification (ID), designations according to ISO/IEC 9995-1 are used. The corresponding character identity is given according to ISO/IEC 10646-1.

In this annex is specified the allocation of characters for the three keyboard alternatives given in clause 4.3 of the standard.

För tangentidentifiering (ID) används beteckningar enligt ISO/IEC 9995-1. Motsvarande teckenidentitet anges enligt ISO/IEC 10646-1.

Samisk tangentbordslayout
Sami keyboard layout

ID	PRIMÄRT - NEDRE SKIFT PRIMARY - LOWER SHIFT	PRIMÄRT - ÖVRE SKIFT PRIMARY - UPPER SHIFT	ALTERNATIVT - NEDRE SKIFT ALTERNATIVE - LOWER SHIFT	ALTERNATIVT - ÖVRE SKIFT ALTERNATIVE - UPPER SHIFT
D01	LATIN SMALL LETTER A WITH ACUTE ACCENT (U+00E1)	LATIN CAPITAL LETTER A WITH ACUTE ACCENT (U+00C1)	LATIN SMALL LETTER Q (U+0071)	LATIN CAPITAL LETTER Q (U+0051)
D02	LATIN SMALL LETTER S WITH CARON (U+0161)	LATIN CAPITAL LETTER S WITH CARON (U+0160)	LATIN SMALL LETTER W (U+0077)	LATIN CAPITAL LETTER W (U+0057)
D03	LATIN SMALL LETTER E (U+0065)	LATIN CAPITAL LETTER E (U+0045)		
D04	LATIN SMALL LETTER R (U+0072)	LATIN CAPITAL LETTER R (U+0052)		
D05	LATIN SMALL LETTER T (U+0074)	LATIN CAPITAL LETTER T (U+0054)		
D06	LATIN SMALL LETTER T WITH STROKE (U+0167)	LATIN CAPITAL LETTER T WITH STROKE (U+0166)	LATIN SMALL LETTER Y (U+0079)	LATIN CAPITAL LETTER Y (U+0059)
D07	LATIN SMALL LETTER U (U+0075)	LATIN CAPITAL LETTER U (U+0055)		
D08	LATIN SMALL LETTER I (U+0069)	LATIN CAPITAL LETTER I (U+0059)		
D09	LATIN SMALL LETTER O (U+006F)	LATIN CAPITAL LETTER O (U+004F)		
D10	LATIN SMALL LETTER P (U+0070)	LATIN CAPITAL LETTER P (U+0050)		
D11	LATIN SMALL LETTER A WITH RING ABOVE (U+00E5)	LATIN CAPITAL LETTER A WITH RING ABOVE (U+00C5)		
D12	LATIN SMALL LETTER ENG (U+014B)	LATIN CAPITAL LETTER ENG (U+014A)		
C01	LATIN SMALL LETTER A (U+0061)	LATIN CAPITAL LETTER A (U+0041)		
C02	LATIN SMALL LETTER S (U+0073)	LATIN CAPITAL LETTER S (U+0053)		
C03	LATIN SMALL LETTER D (U+0064)	LATIN CAPITAL LETTER D (U+0044)		
C04	LATIN SMALL LETTER F (U+0066)	LATIN CAPITAL LETTER F (U+0046)		
C05	LATIN SMALL LETTER G (U+0067)	LATIN CAPITAL LETTER G (U+0047)		

ID	PRIMÄRT - NEDRE SKIFT PRIMARY - LOWER SHIFT	PRIMÄRT - ÖVRE SKIFT PRIMARY - UPPER SHIFT	ALTERNATIVT - NEDRE SKIFT ALTERNATIVE - LOWER SHIFT	ALTERNATIVT - ÖVRE SKIFT ALTERNATIVE - UPPER SHIFT
C06	LATIN SMALL LETTER H (U+0068)	LATIN CAPITAL LETTER H (U+0048)		
C07	LATIN SMALL LETTER J (U+006A)	LATIN CAPITAL LETTER J (U+004A)		
C08	LATIN SMALL LETTER K (U+006B)	LATIN CAPITAL LETTER K (U+004B)		
C09	LATIN SMALL LETTER L (U+006C)	LATIN CAPITAL LETTER L (U+004C)		
C10	LATIN SMALL LETTER O WITH DIAERESIS (U+00F6)	LATIN CAPITAL LETTER O WITH DIAERESIS (U+00D6)	LATIN SMALL LETTER O WITH STROKE (U+00F8)	LATIN CAPITAL LETTER O WITH STROKE (U+00D8)
C11	LATIN SMALL LETTER A WITH DIAERESIS (U+00E4)	LATIN CAPITAL LETTER A WITH DIAERESIS (U+00C4)	LATIN SMALL LETTER AE (U+00C6)	LATIN CAPITAL LETTER AE (U+00E6)
C12	LATIN SMALL LETTER D WITH STROKE (U+0111)	LATIN CAPITAL LETTER D WITH STROKE (U+0110)		
B00	LATIN SMALL LETTER Z WITH CARON (U+017E)	LATIN CAPITAL LETTER Z WITH CARON (U+017D)		
B01	LATIN SMALL LETTER Z (U+007A)	LATIN CAPITAL LETTER Z (U+005A)		
B02	LATIN SMALL LETTER C WITH CARON (U+010D)	LATIN CAPITAL LETTER C WITH CARON (U+010C)	LATIN SMALL LETTER X (U+0078)	LATIN CAPITAL LETTER X (U+0058)
B03	LATIN SMALL LETTER C (U+0063)	LATIN CAPITAL LETTER C (U+0043)		
B04	LATIN SMALL LETTER V (U+0076)	LATIN CAPITAL LETTER V (U+0056)		
B05	LATIN SMALL LETTER B (U+0042)	LATIN CAPITAL LETTER B (U+0062)		
B06	LATIN SMALL LETTER N (U+006E)	LATIN CAPITAL LETTER N (U+004E)		
B07	LATIN SMALL LETTER M (U+006D)	LATIN CAPITAL LETTER M (U+004D)		

Svensk tangentbordslayout 1
Swedish keyboard layout 1

ID	PRIMÄRT - NEDRE SKIFT PRIMARY - LOWER SHIFT	PRIMÄRT - ÖVRE SKIFT PRIMARY - UPPER SHIFT	ALTERNATIVT - NEDRE SKIFT ALTERNATIVE - LOWER SHIFT	ALTERNATIVT - ÖVRE SKIFT ALTERNATIVE - UPPER SHIFT
D01	LATIN SMALL LETTER Q (U+0071)	LATIN CAPITAL LETTER Q (U+0051)		
D02	LATIN SMALL LETTER W (U+0077)	LATIN CAPITAL LETTER W (U+0057)		
D03	LATIN SMALL LETTER E (U+0065)	LATIN CAPITAL LETTER E (U+0045)		
D04	LATIN SMALL LETTER R (U+0072)	LATIN CAPITAL LETTER R (U+0052)		
D05	LATIN SMALL LETTER T (U+0074)	LATIN CAPITAL LETTER T (U+0054)	LATIN SMALL LETTER T WITH STROKE (U+0167)	LATIN CAPITAL LETTER T WITH STROKE (U+0166)

D06	LATIN SMALL LETTER Y (U+0079)	LATIN CAPITAL LETTER Y (U+0059)		
D07	LATIN SMALL LETTER U (U+0075)	LATIN CAPITAL LETTER U (U+0055)		
D08	LATIN SMALL LETTER I (U+0069)	LATIN CAPITAL LETTER I (U+0059)		
D09	LATIN SMALL LETTER O (U+006F)	LATIN CAPITAL LETTER O (U+004F)		
D10	LATIN SMALL LETTER P (U+0070)	LATIN CAPITAL LETTER P (U+0050)		
D11	LATIN SMALL LETTER A WITH RING ABOVE (U+00E5)	LATIN CAPITAL LETTER A WITH RING ABOVE (U+00C5)		
C01	LATIN SMALL LETTER A (U+0061)	LATIN CAPITAL LETTER A (U+0041)	LATIN SMALL LETTER A WITH ACUTE ACCENT (U+00E1)	LATIN CAPITAL LETTER A WITH ACUTE ACCENT (U+00C1)
C02	LATIN SMALL LETTER S (U+0073)	LATIN CAPITAL LETTER S (U+0053)	LATIN SMALL LETTER S WITH CARON (U+0161)	LATIN CAPITAL LETTER S WITH CARON (U+0160)
C03	LATIN SMALL LETTER D (U+0064)	LATIN CAPITAL LETTER D (U+0044)	LATIN SMALL LETTER D WITH STROKE (U+0111)	LATIN CAPITAL LETTER D WITH STROKE (U+0110)
C04	LATIN SMALL LETTER F (U+0066)	LATIN CAPITAL LETTER F (U+0046)		
C05	LATIN SMALL LETTER G (U+0067)	LATIN CAPITAL LETTER G (U+0047)		
C06	LATIN SMALL LETTER H (U+0068)	LATIN CAPITAL LETTER H (U+0048)		
C07	LATIN SMALL LETTER J (U+006A)	LATIN CAPITAL LETTER J (U+004A)		
C08	LATIN SMALL LETTER K (U+006B)	LATIN CAPITAL LETTER K (U+004B)		
C09	LATIN SMALL LETTER L (U+006C)	LATIN CAPITAL LETTER L (U+004C)		
C10	LATIN SMALL LETTER O WITH DIAERESIS (U+00F6)	LATIN CAPITAL LETTER O WITH DIAERESIS (U+00D6)	LATIN SMALL LETTER O WITH STROKE (U+00F8)	LATIN CAPITAL LETTER O WITH STROKE (U+00D8)
C11	LATIN SMALL LETTER A WITH DIAERESIS (U+00E4)	LATIN CAPITAL LETTER A WITH DIAERESIS (U+00C4)	LATIN SMALL LETTER AE (U+00C6)	LATIN CAPITAL LETTER AE (U+00E6)
B01	LATIN SMALL LETTER Z (U+007A)	LATIN CAPITAL LETTER Z (U+005A)	LATIN SMALL LETTER Z WITH CARON (U+017E)	LATIN CAPITAL LETTER Z WITH CARON (U+017D)
B02	LATIN SMALL LETTER X (U+0078)	LATIN CAPITAL LETTER X (U+0058)		
B03	LATIN SMALL LETTER C (U+0063)	LATIN CAPITAL LETTER C (U+0043)	LATIN SMALL LETTER C WITH CARON (U+010D)	LATIN CAPITAL LETTER C WITH CARON (U+010C)
B04	LATIN SMALL LETTER V (U+0076)	LATIN CAPITAL LETTER V (U+0056)		
B05	LATIN SMALL LETTER B (U+0042)	LATIN CAPITAL LETTER B (U+0062)		
B06	LATIN SMALL LETTER N (U+006E)	LATIN CAPITAL LETTER N (U+004E)	LATIN SMALL LETTER ENG (U+014B)	LATIN CAPITAL LETTER ENG (U+014A)
B07	LATIN SMALL LETTER M (U+006D)	LATIN CAPITAL LETTER M (U+004D)		

Svensk tangentbordslayout 2
Swedish keyboard layout 2

ID	PRIMÄRT - NEDRE SKIFT PRIMARY - LOWER SHIFT	PRIMÄRT - ÖVRE SKIFT PRIMARY - UPPER SHIFT	ALTERNATIVT - NEDRE SKIFT ALTERNATIVE - LOWER SHIFT	ALTERNATIVT - ÖVRE SKIFT ALTERNATIVE - UPPER SHIFT
E12	COMBINING ACUTE ACCENT (U+0301)			
D01	LATIN SMALL LETTER Q (U+0071)	LATIN CAPITAL LETTER Q (U+0051)		
D02	LATIN SMALL LETTER W (U+0077)	LATIN CAPITAL LETTER W (U+0057)		
D03	LATIN SMALL LETTER E (U+0065)	LATIN CAPITAL LETTER E (U+0045)		
D04	LATIN SMALL LETTER R (U+0072)	LATIN CAPITAL LETTER R (U+0052)		
D05	LATIN SMALL LETTER T (U+0074)	LATIN CAPITAL LETTER T (U+0054)	LATIN SMALL LETTER T WITH STROKE (U+0167)	LATIN CAPITAL LETTER T WITH STROKE (U+0166)
D06	LATIN SMALL LETTER Y (U+0079)	LATIN CAPITAL LETTER Y (U+0059)		
D07	LATIN SMALL LETTER U (U+0075)	LATIN CAPITAL LETTER U (U+0055)		
D08	LATIN SMALL LETTER I (U+0069)	LATIN CAPITAL LETTER I (U+0059)		
D09	LATIN SMALL LETTER O (U+006F)	LATIN CAPITAL LETTER O (U+004F)	LATIN SMALL LETTER O WITH STROKE (U+00F8)	LATIN CAPITAL LETTER O WITH STROKE (U+00D8)
D10	LATIN SMALL LETTER P (U+0070)	LATIN CAPITAL LETTER P (U+0050)		
D11	LATIN SMALL LETTER A WITH RING ABOVE (U+00E5)	LATIN CAPITAL LETTER A WITH RING ABOVE (U+00C5)		
C01	LATIN SMALL LETTER A (U+0061)	LATIN CAPITAL LETTER A (U+0041)	LATIN SMALL LETTER AE (U+00C6)	LATIN CAPITAL LETTER AE (U+00E6)
C02	LATIN SMALL LETTER S (U+0073)	LATIN CAPITAL LETTER S (U+0053)		
C03	LATIN SMALL LETTER D (U+0064)	LATIN CAPITAL LETTER D (U+0044)		LATIN CAPITAL LETTER D WITH STROKE (U+0110)
C04	LATIN SMALL LETTER F (U+0066)	LATIN CAPITAL LETTER F (U+0046)	LATIN SMALL LETTER D WITH STROKE (U+0111)	
C05	LATIN SMALL LETTER G (U+0067)	LATIN CAPITAL LETTER G (U+0047)	LATIN SMALL LETTER ENG (U+014B)	LATIN CAPITAL LETTER ENG (U+014A)
C06	LATIN SMALL LETTER H (U+0068)	LATIN CAPITAL LETTER H (U+0048)		
C07	LATIN SMALL LETTER J (U+006A)	LATIN CAPITAL LETTER J (U+004A)		
C08	LATIN SMALL LETTER K (U+006B)	LATIN CAPITAL LETTER K (U+004B)		
C09	LATIN SMALL LETTER L (U+006C)	LATIN CAPITAL LETTER L (U+004C)		
C10	LATIN SMALL LETTER O WITH DIAERESIS (U+00F6)	LATIN CAPITAL LETTER O WITH DIAERESIS (U+00D6)		
C11	LATIN SMALL LETTER A WITH DIAERESIS (U+00E4)	LATIN CAPITAL LETTER A WITH DIAERESIS (U+00C4)		COMBINING CARON (U+030C)
B01	LATIN SMALL LETTER Z (U+007A)	LATIN CAPITAL LETTER Z (U+005A)		

B02	LATIN SMALL LETTER X (U+0078)	LATIN CAPITAL LETTER X (U+0058)		
B03	LATIN SMALL LETTER C (U+0063)	LATIN CAPITAL LETTER C (U+0043)		
B04	LATIN SMALL LETTER V (U+0076)	LATIN CAPITAL LETTER V (U+0056)		
B05	LATIN SMALL LETTER B (U+0042)	LATIN CAPITAL LETTER B (U+0062)		
B06	LATIN SMALL LETTER N (U+006E)	LATIN CAPITAL LETTER N (U+004E)		
B07	LATIN SMALL LETTER M (U+006D)	LATIN CAPITAL LETTER M (U+004D)		

2001-01-19

Annex B

Sorteringsordning för bokstäver
Sorting order for letters

A	LATIN CAPITAL LETTER A (U+0041)
a	LATIN SMALL LETTER A (U+0061)
Á	LATIN CAPITAL LETTER A WITH ACUTE (U+00C1)
á	LATIN SMALL LETTER A WITH ACUTE (U+00E1)
B	LATIN CAPITAL LETTER B (U+0042)
b	LATIN SMALL LETTER B (U+0062)
C	LATIN CAPITAL LETTER C (U+0043)
c	LATIN SMALL LETTER C (U+0063)
Č	LATIN CAPITAL LETTER C WITH CARON (U+010C)
č	LATIN SMALL LETTER C WITH CARON (U+010D)
D	LATIN CAPITAL LETTER D (U+0044)
d	LATIN SMALL LETTER D (U+0064)
Đ	LATIN CAPITAL LETTER D WITH STROKE (U+0110)
đ	LATIN SMALL LETTER D WITH STROKE (U+0111)
E	LATIN CAPITAL LETTER E (U+0045)
e	LATIN SMALL LETTER E (U+0065)
É	LATIN CAPITAL LETTER E WITH ACUTE (U+00C9)
é	LATIN SMALL LETTER E WITH ACUTE (U+00E9)
F	LATIN CAPITAL LETTER F (U+0046)
f	LATIN SMALL LETTER F (U+0066)
G	LATIN CAPITAL LETTER G (U+0047)
g	LATIN SMALL LETTER G (U+0067)
H	LATIN CAPITAL LETTER H (U+0048)
h	LATIN SMALL LETTER H (U+0068)
I	LATIN CAPITAL LETTER I (U+0049)
i	LATIN SMALL LETTER I (U+0069)
J	LATIN CAPITAL LETTER J (U+004A)
j	LATIN SMALL LETTER J (U+006A)
K	LATIN CAPITAL LETTER K (U+004B)
k	LATIN SMALL LETTER K (U+006B)
L	LATIN CAPITAL LETTER L (U+004C)
l	LATIN SMALL LETTER L (U+006C)
M	LATIN CAPITAL LETTER M (U+004D)
m	LATIN SMALL LETTER M (U+006D)
N	LATIN CAPITAL LETTER N (U+004E)
n	LATIN SMALL LETTER N (U+006E)
ŋ	LATIN CAPITAL LETTER ENG (U+014A)
ŋ	LATIN SMALL LETTER ENG (U+014B)
O	LATIN CAPITAL LETTER O (U+004F)
o	LATIN SMALL LETTER O (U+06F)
P	LATIN CAPITAL LETTER P (U+0050)
p	LATIN SMALL LETTER P (U+0070)
Q	LATIN CAPITAL LETTER Q (U+0051)
q	LATIN SMALL LETTER Q (U+0071)
R	LATIN CAPITAL LETTER R (U+0052)
r	LATIN SMALL LETTER R (U+0072)
S	LATIN CAPITAL LETTER S (U+0053)
s	LATIN SMALL LETTER S (U+0073)
Š	LATIN CAPITAL LETTER S WITH CARON (U+0160)
š	LATIN SMALL LETTER S WITH CARON (U+0161)
T	LATIN CAPITAL LETTER T (U+0054)
t	LATIN SMALL LETTER T (U+0074)
Ŧ	LATIN CAPITAL LETTER T WITH STROKE (U+0166)
ŧ	LATIN SMALL LETTER T WITH STROKE (U+0167)
U	LATIN CAPITAL LETTER U (U+0055)
u	LATIN SMALL LETTER U (U+0075)
V	LATIN CAPITAL LETTER V (U+0056)
v	LATIN SMALL LETTER V (U+0076)
W	LATIN CAPITAL LETTER W (U+0057)
w	LATIN SMALL LETTER W (U+0077)
X	LATIN CAPITAL LETTER X (U+0058)
x	LATIN SMALL LETTER X (U+0078)
Y	LATIN CAPITAL LETTER Y (U+0059)
y	LATIN SMALL LETTER Y (U+0079)
Ü	LATIN CAPITAL LETTER U WITH DIAERESIS (U+00DC)
ü	LATIN SMALL LETTER U WITH DIAERESIS (U+00FC)
Z	LATIN CAPITAL LETTER Z (U+005A)
z	LATIN SMALL LETTER Z (U+007A)
Ž	LATIN CAPITAL LETTER Z WITH CARON (U+017D)
ž	LATIN SMALL LETTER Z WITH CARON (U+017E)
Æ	LATIN CAPITAL LETTER AE (U+00C6)
æ	LATIN SMALL LETTER AE (U+00E6)
Ø	LATIN CAPITAL LETTER O WITH STROKE (U+00D8)
ø	LATIN SMALL LETTER O WITH STROKE (U+00F8)
Å	LATIN CAPITAL LETTER A WITH RING ABOVE (U+00C5)
å	LATIN SMALL LETTER A WITH RING ABOVE (U+00E5)
Ä	LATIN CAPITAL LETTER A WITH DIAERESIS (U+00C4)
ä	LATIN SMALL LETTER A WITH DIAERESIS (U+00E4)
Ö	LATIN CAPITAL LETTER O WITH DIAERESIS (U+00D6)
ö	LATIN SMALL LETTER O WITH DIAERESIS (U+00F6)