

DB65

新疆维吾尔自治区地方标准

DB 65/T 3690—2015

现行维吾尔文与拉丁维吾尔文 编码字符转换规则

Coded Character Conversion Rules between Uyghur and Latin Uyghur

2015 - 01 - 13 发布

2015 - 03 - 13 实施

新疆维吾尔自治区质量技术监督局 发布

前 言

本标准依照GB/T 1.1-2009《标准化工作导则 第1部分:标准的结构和编写》、GB 13000-2010《信息技术 通用多八位编码字符集(UCS)》、GB 21669-2008《信息技术 维吾尔文、哈萨克文、柯尔克孜文编码字符集》的规定编写。

本标准由新疆维吾尔自治区民族语言文字工作委员会提出。

本标准由新疆维吾尔自治区经济和信息化委员会归口。

本标准起草单位:新疆维吾尔自治区标准化研究院、新疆西北星信息技术有限公司、中科院新疆理化技术研究所、新疆维吾尔自治区民族语言文字工作委员会。

本标准主要起草人:董军、吾守尔·斯拉木、石常海、刘金龙、亚森·伊明、柯俊帆、王磊、阿里木·哈萨尼、申文英、王煜煜、吐尔根·伊布拉音、杨雅婷、塔依尔·阿布都外力。

引 言

现行维吾尔文与拉丁维吾尔文都是用于表达维吾尔语言的拼音文字，但两种文字表达维吾尔语言字母的方式有许多显著区别，主要包括以下几个方面：

- 现行维吾尔文是从右向左书写的文字，拉丁维吾尔文是从左向右书写的文字。
- 现行维吾尔文不区分大小写，拉丁维吾尔文区分大小写。
- 维吾尔文有 8 个元音字母。现行维吾尔文用字符 **ۋ**（该字符的编码为 0x0626）与另外 8 个字符结合构成 8 个双字符字母来表示这 8 个维吾尔文元音字母，且根据上下文环境的不同，字符 **ۋ** 有时需要出现，有时不需要出现；拉丁维吾尔文用 8 个字符分别表示这 8 个维吾尔文元音字母。
- 拉丁维吾尔文有 5 个双字符字母 **Ch, Zh, Sh, Gh, Ng**，其中涉及的 6 个字符 **C, Z, S, G, N, H** 也都分别表示 1 个维吾尔文字母，而在现行维吾尔文中这 11 个维吾尔文字母各用 1 个不同字符表示。
- 缩略语在拉丁维吾尔文中用每个单词首字母的大写形式表示，而在现行维吾尔文中则用每个单词的首字母加空格表示。

为了保证现行维吾尔文与拉丁维吾尔文编码字符转换过程中信息的完整性，应对这两种文字的书写方式进行必要的规范。因此，本标准提出了与编码字符转换有关的两种文字的书写规则，并两种文字编码字符的转换规则。

现行维吾尔文与拉丁维吾尔文编码字符转换规则

1 范围

本标准规定了现行维吾尔文与拉丁维吾尔文编码字符转换的术语和定义、书写规则及转换规则等。本标准适用于现行维吾尔文与拉丁维吾尔文编码字符的转换工具和各类信息系统。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅所注日期的版本适用于本文件。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB 13000 信息技术 通用多八位编码字符集(UCS)

3 术语和定义

下列术语和定义适用于本文件。

3.1

现行维吾尔文 uyghur

指阿拉伯字母为基础的维吾尔文。

3.2

拉丁维吾尔文 latin uyghur

用拉丁字母转写的维吾尔文。

3.3

字母 letter

注音及拼读的符号。

3.4

字符 character

供组织、控制或表示数据用的元素集合中的一个元素。

3.5

编码字符 coded character

字符及其编码表示。

3.6

单字符字母 mono letter

由一个字符表示的字母。

3.7

双字符字母 dual letter

由两个相邻字符按固定的先后顺序构成字符序列表示的字母。

3.8

阿拉伯字符 arabic character

用于表示阿拉伯字母和符号的编码字符，在GB 13000中，有5个阿拉伯字符的编码区域，分别是基本区（编码从0x0600到0x06FF）、补充区（编码从0x0750到0x077F）、扩展区（编码从0x08A0到0x08FF）、变形显现区A（编码从0xFB50到0xFDEE）、变形显现区B（编码从0xFE70到0xFEFF）。

3.9

拉丁字符 latin character

用于表示拉丁字母和符号的编码字符，在GB 13000中，有9个拉丁字符的编码区域，分别是基本区（编码从0x0000到0x007F）、补充区（编码从0x0080到0x00FF）、扩展区A（编码从0x0100到0x017F）、扩展区B（编码从0x0180到0x024F）、扩展区C（编码从0x2C60到0x2C7F）、扩展区D（编码从0xA720到0xA7FF）、附加扩展区（编码从0x1E00到0x1EFF）、连字区（编码从0xFB00到0xFB06）、全宽字母区（编码从0xFF00到0xFF5E）。

3.10

分隔符 delimiter

字符 '，主要用于音节分隔或区分双字符字母和单字符字母，该字符的编码为0x0027。

3.11

连接符 tatweel

字符 ـ，主要用于调整现行维吾尔文的编辑排版效果，该字符的编码为0x0640。

3.12

镜像字符 mirrored character

在GB 13000中，具有Bidi Mirrored属性的字符，例如各种括号。

3.13

布局和格式控制字符 layout and format control character

用于控制字符的连接行为、双向顺序行为和选择显示格式行为的字符，例如：编码为0x200C的零宽不连接字符，编码为0x200F的从右向左标记字符等。

3.14

零宽空格 zero width space

编码为0x200B的字符，宽度为零，主要用于指示单词的边界。

3.15

零宽不中断空格 zero width no-break space

编码为0xFEFF的字符，宽度为零，主要用于指示字符的前后即不分词也不断行。

3.16

一对一的编码字符转换 conversion of one coded character to one coded character

一个编码字符转换为一个编码字符。

3.17

一对多的编码字符转换 conversion of one coded character to several coded character

一个编码字符转换为多个编码字符。

3.18

多对多的编码字符转换 conversion of several coded character to several coded character

多个编码字符转换为多个编码字符。

3.19

多对一的编码字符转换 conversion of several coded character to one coded character

多个编码字符转换为一个编码字符。

4 与编码字符转换相关的现行维吾尔文书写规则

缩略语的每个字母之间仅有一个空格（见示例1）。

示例1: ش ئۆ ئا ر (شەنجاڭ ئۇيغۇر ئاپتونوم رايونى، 新疆维吾尔自治区)

5 与编码字符转换相关的拉丁维吾尔文书写规则

5.1 句首字母应大写（见示例2）。

5.2 双引号内的首字母应大写（见示例3）。

5.3 专有名词(人名、地名、机构名称)的首字母应大写（见示例4）。

5.4 双字符字母的大写形式只大写首字符（见示例5）。

5.5 一个单词中，前一个音节是闭音节（以辅音字母结束的音节），并且后一个音节是元音字母开头时，前音节与后音节中间必须标志分隔符 '（见示例6）。

5.6 为了准确区分单字符字母和双字符字母，在一个单词中，如果相邻的两个字符 C 和 H，Z 和 H，S 和 H，G 和 H，N 和 G 分别代表一个维吾尔文字母，则必须在它们之间标志分隔符 '（见示例7）。

5.7 缩略语中的字母必须大写（见示例8）。

示例2: Men mektepte oquymen. (我在学校读书)

示例3: Alim mundaq d ädi: "Men burun bundaq ägiz ademni k ör üp baqmighan." (阿里木说: “我以前没有见过这么高的人。”))

示例4: Alim (阿里木), Turpan (吐鲁番)

示例5: Shinjang Uyghur Aptonom Rayoni (新疆维吾尔自治区)

示例6: chet'el (外国)

示例7: Ez'her (埃孜哈尔)

示例8: ShUAR (新疆维吾尔自治区)

6 现行维吾尔文编码字符转换为拉丁维吾尔文编码字符的规则

6.1 总体要求

6.1.1 扩展区编码字符的转换

阿拉伯字符变形显现区A和变形显现区B的编码字符表达的现行维吾尔文，应首先转换为阿拉伯字符基本区编码字符表达的现行维吾尔文，然后再转换为对应的拉丁维吾尔文编码字符。

6.1.2 现行维吾尔文应首先转换为拉丁维吾尔文的小写形式

如无特别说明，现行维吾尔文编码字符应先转换为对应的拉丁维吾尔文编码字符的小写形式，然后再根据拉丁维吾尔文的书写规则进行必要的大小写转换。

6.1.3 镜像字符的处理

如无特别说明，镜像字符保留不变。信息系统不能依据上下文环境选择镜像字符的正确字形用于显示，需将一个镜像字符的编码字符转换为与之镜像的另一个编码字符。表1列出了部分镜像字符的镜像关系，完整的镜像字符列表可以参照GB 13000。

表1 部分镜像字符

| 文字 | 字符 | | | | | | | | | |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 现行维吾尔文 | (0028 |) 0029 | [005B |] 005D | { 007B | } 007D | « 00AB | » 00BB | ‹ 2039 | › 203A |
| 拉丁维吾尔文 |) 0029 | (0028 |] 005D | [005B | } 007D | { 007B | » 00BB | « 00AB | › 203A | ‹ 2039 |

6.2 一对一的编码字符转换

6.2.1 现行维吾尔文 19 个辅音字母，其编码字符应直接转换为对应的拉丁维吾尔文辅音字母的编码字符具体见表 2（示例见 9 至 12）。

表2 现行维吾尔文辅音字母的转换

| 文字 | 字符 | | | | | | | | | |
|--------|----|---|---|---|---|---|---|---|---|---|
| 现行维吾尔文 | ب | پ | ت | ج | خ | د | ر | ز | س | ف |
| 拉丁维吾尔文 | b | p | t | j | x | d | r | z | s | f |
| 现行维吾尔文 | ق | ك | گ | ل | م | ن | ه | و | ى | |
| 拉丁维吾尔文 | q | k | g | l | m | n | h | w | y | |

示例9: قول (手) 应转换为 qol

示例10: باش (头) 应转换为 bash

示例11: پۇت (足) 应转换为 put

示例12: كۆز (眼) 应转换为 köz

6.2.2 现行维吾尔文 8 个元音字母是双字符字母，当出现在非词首位置时，根据上下文环境字符 ۋ 有时需要出现，有时不需要出现。如表 3 所示，当字符 ۋ 不出现时，则这 8 个元音字母直接转换为对应的拉丁维吾尔文元音字母的编码字符（示例见 13 至 17）。

表3 现行维吾尔文元音字母在非词首位置的转换

| 文字 | 字符 | | | | | | | |
|--------|----|---|---|---|---|---|---|---|
| 现行维吾尔文 | ا | ە | و | ۇ | ۆ | ۈ | ي | ى |
| 拉丁维吾尔文 | a | e | o | u | ö | ü | ë | i |

- 示例13: كىتاب (书) 应转换为 kitab
- 示例14: ۋەتەن (祖国) 应转换为 weten
- 示例15: تومۇر (经脉) 应转换为 tomur
- 示例16: كۆمۈر (煤) 应转换为 kǎmür
- 示例17: ئېلېكتىر (电) 应转换为 ðǎktir

6.3 一对多的编码字符转换

现行维吾尔文5个字母，其编码字符应直接转换为对应的拉丁维吾尔文双字符字母的编码字符，具体见表3（示例见18至21）。

表4 现行维吾尔文字母چ、ژ、ش、غ、ڭ的转换

| 文字 | 字符 | | | | |
|--------|----|----|----|----|----|
| 现行维吾尔文 | چ | ژ | ش | غ | ڭ |
| 拉丁维吾尔文 | ch | zh | sh | gh | ng |

- 示例18: چايدان (暖瓶) 应转换为 chaydan
- 示例19: ژۇرنال (杂志) 应转换为 zhurnal
- 示例20: شىنجاڭ (新疆) 应转换为 Shinxjang
- 示例21: غەلبە (胜利) 应转换为 ghelibe

6.4 多对一的编码字符转换

现行维吾尔文8个元音字母是双字符字母，当出现在词首位置时，应忽略词首的字符 ۋ,直接转换为对应的拉丁维吾尔文元音字母的编码字符，具体见表4（示例见22至29）。

表5 现行维吾尔文元音字母在词首位置的转换

| 文字 | 字符 | | | | | | | |
|--------|----|----|----|----|----|----|----|----|
| 现行维吾尔文 | ئا | ئە | ئو | ئۇ | ئۆ | ئۈ | ئى | ئې |
| 拉丁维吾尔文 | a | e | o | u | ö | ü | ë | i |

- 示例22: ئانار (石榴) 应转换为 anar
- 示例23: ئەنجۈر (无花果) 应转换为 enjür
- 示例24: ئوردا (皇宫) 应转换为 orda
- 示例25: ئۇرۇش (战争) 应转换为 urush
- 示例26: ئۆردەك (鸭子) 应转换为 ördék
- 示例27: ئۈزۈم (葡萄) 应转换为 üzüm
- 示例28: ئېلان (广告) 应转换为 ðlan
- 示例29: ئىنكاس (反映) 应转换为 inkas

6.5 多对多的编码字符转换

6.5.1 现行维吾尔文8个元音字母是双字符字母，当出现在非词首位置时，根据上下文环境字符 ۋ 有时需要出现，有时不需要出现。如表5所示，当字符 ۋ 出现时，则字符 ۋ 转换为分隔符 ' 的字符编码，双字符字母的另一个字符转换为对应的拉丁维吾尔文元音字母的编码字符（示例见30至37）。

表6 现行维吾尔文元音字母在非词首位置的转换

| 文字 | 字符 | | | | | | | |
|--------|----|----|----|----|----|----|----|----|
| 现行维吾尔文 | ئا | ئە | ئو | ئۇ | ئۆ | ئۈ | ئې | ئى |
| 拉丁维吾尔文 | 'a | 'e | 'o | 'u | 'ö | 'ü | 'ë | 'i |

示例30: ئىنىكئانا (奶妈) 应转换为 inik'ana

示例31: ئەسئەت (艾斯艾提) 应转换为 Es'et

示例32: رادىئو (收音机) 应转换为 radi'o

示例33: مەسئۇل (负责) 应转换为 mes'ul

示例34: قارىئۆرۈك (乌梅) 应转换为 qari'örük

示例35: ئائىمىد (失望) 应转换为 na'ümid

示例36: ئىتئېيىق (狗熊) 应转换为 it'ëyiq

示例37: جەمئىي (共计) 应转换为 jem'iy

6.5.2 单词中有如表7所示的现行维吾尔文字符序列，字符序列中的每个编码字符转换为对应的拉丁维吾尔文字母的编码字符，同时应在两个拉丁维吾尔文字母之间标志分隔符'的编码字符（示例见38至40）。

表7 现行维吾尔文字符序列的转换

| 文字 | 字符 | | | |
|--------|-----|-----|-----|-----|
| 现行维吾尔文 | زھ | سھ | گھ | نگ |
| 拉丁维吾尔文 | z'h | s'h | g'h | n'g |

示例38: ئەزھەر (埃孜哈尔) 应转换为 Ez'her

示例39: ئىسھاق (伊萨克) 应转换为 Is'haq

示例40: نەمەنگان (纳曼干) 应转换为 Nemen'gan

6.6 缩略语的编码字符转换

相邻的两个或两个以上现行维吾尔文左右及之间有1个空格，应忽略字母之间的空格，并将现行维吾尔文字母转换为对应的拉丁维吾尔文的大写形式（示例见41至43）。

示例41: ج ك پ (中国共产党) 应转换为 JKP

示例42: ج خ (公安) 应转换为 JX

示例43: ش ئۇ ئار (新疆维吾尔自治区) 应转换为 ShUAR

6.7 标点符号的编码字符转换

现行维吾尔文标点符号的编码字符应转换为对应的拉丁维吾尔文标点符号的编码字符，具体见表7（示例见44至46）。

表8 现行维吾尔文标点符号的转换

| 文字 | 字符 | | |
|--------|------|------|------|
| 现行维吾尔文 | ‘ | ؟ | ؛ |
| | 060C | 061F | 061B |
| 拉丁维吾尔文 | ’ | ? | ; |
| | 002C | 003F | 003B |

示例44: بۇ نېمە؟ (这是什么?)应转换为 bu n'äme?

示例45: مەن، سەن (我，你)应转换为 men,sen

示例46: ئادەم؛ ھايۋان (人；动物)应转换为 adem; haywan

6.8 非现行维吾尔文编码字符的转换

6.8.1 现行维吾尔文中拉丁字符的处理

现行维吾尔文中的拉丁字符应保留不变,同时在拉丁字符的前后分别增加零宽空格和零宽不中断空格。

6.8.2 零宽空格和零宽不中断空格的处理

忽略现行维吾尔文中的零宽空格和零宽不中断空格,同时零宽空格和零宽不中断空格之间的字符保留不变。

6.8.3 布局和格式控制字符的处理

现行维吾尔文中的布局和格式控制字符保留不变。

6.8.4 连接符的处理

当连接符出现在现行维吾尔文单词中,转换为拉丁维吾尔文时应忽略连接符。

6.8.5 其它非现行维吾尔文编码字符的处理

现行维吾尔文中的其它非现行维吾尔文编码字符保留不变。

6.8.6 现行维吾尔文与拉丁维吾尔文编码字符对照见附录 A。

7 拉丁维吾尔文编码字符转换为现行维吾尔文编码字符的规则

7.1 总体要求

7.1.1 扩展区编码字符的转换

如无特别说明,拉丁维吾尔文编码字符转换为对应现行维吾尔文文字母名义形式的编码字符。些信息系统不能依据现行维吾尔文文字母名义形式的编码字符,选择正确的变形显现形式编码字符用于显示的,需将现行维吾尔文文字母名义形式的编码字符依据现行维吾尔文的书写习惯进一步转换为正确的变形显现形式编码字符。

7.1.2 拉丁维吾尔文文字母大小写形式的处理

如无特别说明,在拉丁维吾尔文编码字符转换为现行维吾尔文编码字符的过程中,对同一个拉丁维吾尔文文字母的大写形式和小写形式采用同样的方式处理。

7.1.3 镜像字符的处理

如无特别说明,镜像字符保留不变。信息系统不能依据上下文环境选择镜像字符的正确字形用于显示的,需将一个镜像字符的编码字符转换为与之镜像的另一个编码字符。表8列出了部分镜像字符的镜像关系,完整的镜像字符列表可以参照GB 13000。

表9 部分镜像字符的转换

| 文字 | 字符 | | | | | | | | | |
|--------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 拉丁维吾尔文 | (0028 |) 0029 | [005B |] 005D | { 007B | } 007D | « 00AB | » 00BB | ‹ 2039 | › 203A |
| 现行维吾尔文 |) 0029 | (0028 |] 005D | [005B | } 007D | { 007B | » 00BB | « 00AB | › 203A | ‹ 2039 |

7.2 一对一的编码字符转换

7.2.1 拉丁维吾尔文 19 个辅音字母，其编码字符应直接转换为对应的现行维吾尔文元音字母的编码字符，具体见表 9（示例见 47 至 50）。

表10 拉丁维吾尔文辅音字母的转换

| 文字 | 字符 | | | | | | | | | |
|--------|----|---|---|---|---|---|---|---|---|---|
| 拉丁维吾尔文 | b | p | t | j | x | d | r | z | s | f |
| 现行维吾尔文 | ب | پ | ت | ج | خ | د | ر | ز | س | ف |
| 拉丁维吾尔文 | q | k | h | l | m | n | h | w | y | |
| 现行维吾尔文 | ق | ك | گ | ل | م | ن | ھ | ۋ | ي | |

示例47: qol（手）应转换为 قول

示例48: bash（头）应转换为 باش

示例49: put（足）应转换为 پۇت

示例50: köz（眼）应转换为 كۆز

7.2.2 拉丁维吾尔文 8 个元音字母出现在非词首时，转换为对应现行维吾尔文元音字母不带字符 ۋ 的形式的编码字符，具体见表 10（示例见 51 至 55）。

表11 拉丁维吾尔文元音字母在非词首位置的转换

| 文字 | 字符 | | | | | | | |
|--------|----|---|---|---|---|---|---|---|
| 拉丁维吾尔文 | a | e | o | u | ö | ü | ë | i |
| 现行维吾尔文 | ا | ە | و | ۇ | ۆ | ۈ | ى | ي |

示例51: kitab（书）应转换为 كىتاب

示例52: weten（祖国）应转换为 ۋەتەن

示例53: tomur（经脉）应转换为 تومۇر

示例54: kömür（煤）应转换为 كۆمۈر

示例55: dāktir（电）应转换为 ئېلېكتىر

7.3 一对多的编码字符转换

7.3.1 拉丁维吾尔文 8 个元音字母出现在词首位置时，转换为对应现行维吾尔文元音字母的编码字符，具体见表 11（示例见 56 至 63）。

表12 拉丁维吾尔文元音字母在词首位置的转换

| 文字 | 字符 | | | | | | | |
|--------|----|----|----|----|----|----|----|----|
| 拉丁维吾尔文 | a | e | o | u | ö | ü | ë | i |
| 现行维吾尔文 | ئا | ئە | ئو | ئۇ | ئۆ | ئۈ | ئى | ئى |

示例56: anar（石榴）应转换为 ئانار

示例57: enjür（无花果）应转换为 ئەنجۈر

示例58: orda（皇宫）应转换为 ئوردا

- 示例59: urush (战争) 应转换为 ئۇرۇش
- 示例60: ǎrdek (鸭子) 应转换为 ئۆردەك
- 示例61: ǖzǖm (葡萄) 应转换为 ئۈزۈم
- 示例62: ǎn (广告) 应转换为 ئېلان
- 示例63: inkas (反映) 应转换为 ئىنكاس

7.4 多对一的编码字符转换

7.4.1 拉丁维吾尔文 5 个双字符字母的编码字符直接转换为对应的现行维吾尔文 5 个字母的编码字符，具体见表 12（示例见 64 至 67）。

表13 拉丁维吾尔文双字符字母的转换

| 文字 | 字符 | | | | |
|--------|----|----|----|----|----|
| 拉丁维吾尔文 | ch | zh | sh | gh | ng |
| 现行维吾尔文 | چ | ژ | ش | غ | ڭ |

- 示例64: chaydan (暖瓶) 应转换为 چايدان
- 示例65: zhurnal (杂志) 应转换为 ژۇرنال
- 示例66: Shinjang (新疆) 应转换为 شىنجاڭ
- 示例67: ghelibe (胜利) 应转换为 غەلبە

7.5 分隔符编码字符的转换

7.5.1 分隔符 '在拉丁维吾尔文 字母 z 和 h, s 和 h, g 和 h, n 和 g 之间，转换应忽略分隔符；分隔符 '在拉丁维吾尔文元音字母之前，转换为字符 ۋ 的编码字符；分隔符 '在其它位置保留不变（示例见 68 至 70）。

- 示例68: Ez'her (埃孜哈尔) 应转换为 ئەزھەر
- 示例69: Is'haq (伊萨克) 应转换为 ئىسھاق
- 示例70: Nemen'gan (纳曼干) 应转换为 نەمەنگان

7.6 缩略语编码字符的转换

7.6.1 相邻的多个大写拉丁维吾尔文 字母的编码字符转换为现行维吾尔文 编码字符，并在现行维吾尔文 编码字符之间加 1 个空格编码字符（示例见 71 至 73）。

- 示例71: JKP (中国共产党) 应转换为 ج ك پ
- 示例72: JX (公安) 应转换为 ج خ
- 示例73: ShUAR (新疆维吾尔自治区) 应转换为 ش ئۇ ئا ر

7.7 标点符号的转换

7.7.1 表 14 所示的拉丁维吾尔文 标点符号的编码字符应转换为对应的现行维吾尔文 标点符号的编码字符（示例见 74 至 76）。

表 14 拉丁维吾尔文标点符号的转换

| 文字 | 字符 | | |
|--------|------|------|------|
| 拉丁维吾尔文 | , | ? | ; |
| | 002C | 003F | 003B |
| 现行维吾尔文 | ، | ؟ | ؛ |
| | 060C | 061F | 061B |

- 示例74: bu n äne? (这是什么?) 应转换为 بۇ نېمە؟

示例75: men, sen (我, 你)应转换为 مەن، سەن

示例76: adem; haywan (人; 动物)应转换为 ئادەم؛ ھايۋان

7.8 非拉丁维吾尔文编码字符的转换

7.8.1 拉丁维吾尔文中阿拉伯字符的处理

拉丁维吾尔文中的阿拉伯字符应保留不变,同时在阿拉伯字符的前后分别增加零宽空格和零宽不中断空格。

7.8.2 零宽空格和零宽不中断空格的处理

忽略拉丁维吾尔文中的零宽空格和零宽不中断空格,同时零宽空格和零宽不中断空格之间的字符保留不变。

7.8.3 布局和格式控制字符的处理

拉丁维吾尔文中的布局和格式控制字符保留不变。

7.8.4 其它非拉丁维吾尔文编码字符的处理

拉丁维吾尔文中的其它非拉丁维吾尔文编码字符保留不变。

7.8.5 有特殊要求需要保留不变的字符或字符串

有特殊要求需要保留的字符或字符串保留不变,例如罗马数字(I、II、III、IV、X)、字母名词(WTO、VCD、DVD)等。

7.9 拉丁维吾尔文与现行维吾尔文编码字符对照见附录 B。

附 录 A
(规范性附录)
现行维吾尔文与拉丁维吾尔文编码字符对照

A.1 现行维吾尔文与拉丁维吾尔文编码字符对照

现行维吾尔文与拉丁维吾尔文编码字符对照见表A.1。

表 A.1 现行维吾尔文与拉丁维吾尔文编码字符对照表

| 序号 | 现行维吾尔文 | | | | | 拉丁维吾尔文 | |
|----|-----------------|-------------------|-----------|-----------|-------------------|-----------------|-----------------|
| | 名义形式 | 独立字符形式 | 首写字符形式 | 中字符形式 | 尾字符形式 | 大写形式 | 小写形式 |
| 1 | ئا 0626+0627 | ئا ا FBEA FE8D | | | ئا ل FBEB FE8E | A 0041 | a 0061 |
| | | | | | | 'A 0027+0041 | 'a 0027+0061 |
| 2 | ئە 0626+06D5 | ئە ە FBEC FEE9 | | | ئە ە FBED FEEA | E 0045 | e 0065 |
| | | | | | | 'E 0027+0045 | 'e 0027+0065 |
| 3 | ب 0628 | ب FE8F | ب FE91 | ب FE92 | ب FE90 | B 0042 | b 0062 |
| 4 | پ 067E | پ FB56 | پ FB58 | پ FB59 | پ FB57 | P 0050 | p 0070 |
| 5 | ت 062A | ت FE95 | ت FE97 | ت FE98 | ت FE96 | T 0054 | t 0074 |
| 6 | ج 062C | ج FE9D | ج FE9F | ج FEA0 | ج FE9E | J 004A | j 006A |
| 7 | چ 0686 | چ FB7A | چ FB7C | چ FB7B | چ FB7D | Ch 0043+0068 | ch 0063+0068 |
| 8 | خ 062E | خ FEA5 | خ FEA7 | خ FEA8 | خ FEA6 | X 0058 | x 0078 |
| 9 | د 062F | د FEA9 | | | د FEAA | D 0044 | d 0064 |
| 10 | ر 0631 | ر FEAD | | | ر FEAE | R 0052 | r 0072 |
| 11 | ز 0632 | ز FEAF | | | ز FEB0 | Z 005A | z 007A |

表 A.1 (续)

| 序号 | 现行维吾尔文 | | | | | 拉丁维吾尔文 | |
|----|-----------------|-------------------|-----------|-----------|-------------------|-----------------|-----------------|
| | 名义形式 | 独立字符形式 | 首写字符形式 | 中字符形式 | 尾字符形式 | 大写形式 | 小写形式 |
| 12 | ژ 0698 | ژ FB8A | | | ژ FB8B | Zh 005A+0068 | zh 007A+0068 |
| 13 | س 0633 | س FEb1 | س FEb3 | س FEB4 | س FEB2 | S 0053 | s 0073 |
| 14 | ش 0634 | ش FEB5 | ش FEB7 | ش FEB8 | ش FEB6 | Sh 0053+0068 | sh 0073+0068 |
| 15 | غ 063A | غ FECD | غ FECF | غ FED0 | غ FECE | Gh 0047+0068 | gh 0067+0068 |
| 16 | ف 0641 | ف FED1 | ف FED3 | ف FED4 | ف FED2 | F 0046 | f 0066 |
| 17 | ق 0642 | ق FED5 | ق FED7 | ق FED8 | ق FED6 | Q 0051 | q 0071 |
| 18 | ك 0643 | ك FED9 | ك FEDB | ك FEDC | ك FEDA | K 004B | k 006B |
| 19 | گ 06AF | گ FB92 | گ FB94 | گ FB95 | گ FB93 | G 0047 | g 0067 |
| 20 | ڭ 06AD | ڭ FBD3 | ڭ FBD5 | ڭ FBD6 | ڭ FBD4 | Ng 004E+0067 | ng 006E+0067 |
| 21 | ل 0644 | ل FEDD | ل FEDF | ل FEE0 | ل FEDE | L 004C | l 006C |
| 22 | م 0645 | م FEE1 | م FEE3 | م FEE4 | م FEE2 | M 004D | m 006D |
| 23 | ن 0646 | ن FEE5 | ن FEE7 | ن FEE8 | ن FEE6 | N 004E | n 006E |
| 24 | ھ 06BE | ھ FBAA | ھ FBAC | ھ FBAD | ھ FBAB | H 0048 | h 0068 |
| 25 | ئو 0626+0648 | و ئو FBEE FEED | | | و ئو FBEF FEEE | O 004F | o 006F |
| | | | | | | 'O 0027+004F | 'o 0027+006F |
| 26 | ئو 0626+06C7 | ۇ ئو FBF0 FBD7 | | | ۇ ئو FBF1 FBD8 | U 0055 | u 0075 |
| | | | | | | 'U 0027+0055 | 'u 0027+0075 |

表 A.1 (续)

| 序号 | 现行维吾尔文 | | | | | 拉丁维吾尔文 | |
|----|------------------|-------------------|-------------------|------------------|-------------------|-----------------------|-----------------------|
| | 名义形式 | 独立字符形式 | 首写字符形式 | 中字符形式 | 尾字符形式 | 大写形式 | 小写形式 |
| 27 | ئۆ 0626+06C6 | ۆ ئۆ FBF2 FBD9 | | | ۆ ئۆ FBF3 FBDA | Ö 00D6 | ö 00F6 |
| | | | | | | 'Ö 0027+00D6 | 'ö 0027+00F6 |
| 28 | ئو 0626+06C8 | و ئو FBF4 FBDB | | | و ئو FBF5 FBDC | Ü 00DC | ü 00FC |
| | | | | | | 'Ü 0027+00DC | 'ü 0027+00FC |
| 29 | ۇ 06CB | ۇ FBDE | | | ۇ FBDF | W 0057 | w 0077 |
| 30 | ئې 0626+06D0 | ې ئې FBF6 FBE4 | ې ئې FBF8 FBE6 | ې FBE7 | ې ئې FBF7 FBE5 | Ë 00CB | ë 00EB |
| | | | | | | 'Ë 0027+00CB | 'ë 0027+00EB |
| 31 | ئى 0626+0649 | ى ئى FBF9 FEEF | د ئى FBFB FBE8 | ى FBE9 | ى ئى FBFA FEF0 | I 0049 | i 0069 |
| | | | | | | İ 0027+0049 | ı 0027+0069 |
| 32 | ي 064A | ي FEF1 | ي FEF3 | ي FEF4 | ي FEF2 | Y 0059 | y 0079 |
| 33 | ئى 0626 | ئى FE89 | ئى FE8B | ئى FE8C | ئى FE8A | ‘ 0027 | ‘ 0027 |
| | | | | | | | |
| 34 | زھ 0632+06BE | زھ FEAF+FBAA | زھ FEAF+FBAC | زھ FEB0+FBAC | زھ FEB0+FBAA | Z'H 005A+0027+0048 | z'h 007A+0027+0068 |
| 35 | سېپ 0633+06BE | سېپ FEB3+FBAB | سېپ FEB3+FBAC | سېپ FEB4+FBAD | سېپ FEB4+FBAB | S'H 0053+0027+0048 | s'h 0073+0027+0068 |
| 36 | گېپ 06AF+06BE | گېپ FB94+FBAB | گېپ FB94+FBAD | گېپ FB95+FBAD | گېپ FB93+FBAB | G'H 0047+0027+0048 | g'h 0067+0027+0068 |
| 37 | نگ 0646+06AF | نگ FEE5+FB92 | نگ FEE7+FB95 | نگ FEE8+FB95 | نگ FEE8+FB93 | H'G 0048+0027+0047 | n'g 006E+0027+0067 |

表 A.1（续）

| 序号 | 现行维吾尔文 | | | | | 拉丁维吾尔文 | |
|----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | 名义形式 | 独立字符形式 | 首写字符形式 | 中字符形式 | 尾字符形式 | 大写形式 | 小写形式 |
| 38 | ‘ 060C | ‘ 060C | ‘ 060C | ‘ 060C | ‘ 060C | ‘ 002C | ‘ 002C |
| 39 | ؟ 061F | ؟ 061F | ؟ 061F | ؟ 061F | ؟ 061F | ؟ 003F | ؟ 003F |
| 40 | ؛ 061B | ؛ 061B | ؛ 061B | ؛ 061B | ؛ 061B | ؛ 003B | ؛ 003B |

附 录 B
(规范性附录)
拉丁维吾尔文与现行维吾尔文编码字符对照

B.1 拉丁维吾尔文与现行维吾尔文编码字符对照

拉丁维吾尔文与现行维吾尔文编码字符对照见表B.1。

表 B.1 拉丁维吾尔文与现行维吾尔文编码字符对照表

| 序号 | 拉丁维吾尔文 | | 现行维吾尔文 | | | | |
|----|-----------|-----------|-----------------|-------------------|-------------------|-----------|-----------------|
| | 大写形式 | 小写形式 | 名义形式 | 独立字符形式 | 首写字符形式 | 中字符形式 | 尾字符形式 |
| 1 | A 0041 | a 0061 | ئا 0626+0627 | ا ئا FBEA FE8D | | | ئا FBEB FE8E |
| 2 | B 0042 | b 0062 | ب 0628 | ب FE8F | ب FE91 | ب FE92 | ب FE90 |
| 3 | D 0044 | d 0064 | د 062F | د FEA9 | | | د FEAA |
| 4 | E 0045 | e 0065 | ئە 0626+06D5 | ە ئە FBEC FEE9 | | | ئە FBED FEEA |
| 5 | F 0046 | f 0066 | ف 0641 | ف FED1 | ف FED3 | ف FED4 | ف FED2 |
| 6 | G 0047 | g 0067 | گ 06AF | گ FB92 | گ FB94 | گ FB95 | گ FB93 |
| 7 | H 0048 | h 0068 | ھ 06BE | ھ FBAA | ھ FBAC | ھ FBAD | ھ FBAB |
| 8 | I 0049 | i 0069 | ئى 0626+0649 | ى ئى FBF9 FEEF | د ئى FBFB FBE8 | د FBE9 | ئى FBFA FEF0 |
| 9 | J 004A | j 006A | ج 062C | ج FE9D | ج FE9F | ج FEA0 | ج FE9E |
| 10 | K 004B | k 006B | ك 0643 | ك FED9 | ك FEDB | ك FEDC | ك FEDA |
| 11 | L 004C | l 006C | ل 0644 | ل FEDD | ل FEDF | ل FEE0 | ل FEDE |
| 12 | M 004D | m 006D | م 0645 | م FEE1 | م FEE3 | م FEE4 | م FEE2 |
| 13 | N 004E | n 006E | ن 0646 | ن FEE5 | ن FEE7 | ن FEE8 | ن FEE6 |

表 B. 1 (续)

| 序号 | 拉丁维吾尔文 | | 现行维吾尔文 | | | | |
|----|-----------------|-----------------|-----------------|-------------------|-------------------|-----------|-------------------|
| | 大写形式 | 小写形式 | 名义形式 | 独立字符形式 | 首写字符形式 | 中字符形式 | 尾字符形式 |
| 14 | O 004F | o 006F | ئو 0626+0648 | و ئو FBEE FEED | | | و ئو FBEF FEEE |
| 15 | P 0050 | p 0070 | پ 067E | پ FB56 | پ FB58 | پ FB59 | پ FB57 |
| 16 | Q 0051 | q 0071 | ق 0642 | ق FED5 | ق FED7 | ق FED8 | ق FED6 |
| 17 | R 0052 | r 0072 | ر 0631 | ر FEAD | | | ر FEAE |
| 18 | S 0053 | s 0073 | س 0633 | س FEb1 | س FEb3 | س FEB4 | س FEB2 |
| 19 | T 0054 | t 0074 | ت 062A | ت FE95 | ت FE97 | ت FE98 | ت FE96 |
| 20 | U 0055 | u 0075 | ئو 0626+06C7 | و ئو FBF0 FBD7 | | | و ئو FBF1 FBD8 |
| 21 | W 0057 | w 0077 | ۇ 06CB | ۇ FBDE | | | ۇ FBDF |
| 22 | X 0058 | x 0078 | خ 062E | خ FEA5 | خ FEA7 | خ FEA8 | خ FEA6 |
| 23 | Y 0059 | y 0079 | ي 064A | ي FEF1 | ي FEF3 | ي FEF4 | ي FEF2 |
| 24 | Z 005A | z 007A | ز 0632 | ز FEAF | | | ز FEB0 |
| 25 | Ë 00CB | ë 00EB | ئې 0626+06D0 | ي ئې FBF6 FBE4 | ب ئې FBF8 FBE6 | ب FBE7 | ي ئې FBF7 FBE5 |
| 26 | Ö 00D6 | ö 00F6 | ئۆ 0626+06C6 | و ئۆ FBF2 FBD9 | | | و ئۆ FBF3 FBDA |
| 27 | Ü 00DC | ü 00FC | ئو 0626+06C8 | و ئو FBF4 FBDB | | | و ئو FBF5 FBDC |
| 28 | Ch 0043+0068 | ch 0063+0068 | چ 0686 | چ FB7A | چ FB7C | چ FB7B | چ FB7D |
| 29 | Gh 0047+0068 | gh 0067+0068 | غ 063A | غ FECD | غ FECF | غ FED0 | غ FECE |
| 30 | Ng 004E+0067 | ng 006E+0067 | ڭ 06AD | ڭ FBD3 | ڭ FBD5 | ڭ FBD6 | ڭ FBD4 |
| 31 | Sh 0053+0068 | sh 0073+0068 | ش 0634 | ش FEB5 | ش FEB7 | ش FEB8 | ش FEB6 |

表 B. 1 (续)

| 序号 | 拉丁维吾尔文 | | 现行维吾尔文 | | | | |
|----|---------------------------|---------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | 大写形式 | 小写形式 | 名义形式 | 独立字符形式 | 首写字符形式 | 中字符形式 | 尾字符形式 |
| 32 | Zh 005A+0068 | zh 007A+0068 | ژ 0698 | ژ FB8A | | | ژ FB8B |
| 33 | G'H 0047+0027 +0048 | g'h 0067+0027 +0068 | گھ 06AF+06BE | گھ FB94+FBAB | گھ FB94+FBAD | گھ FB95+FBAD | گھ FB93+FBAB |
| 34 | N'G 004E+0027 +0047 | n'g 006E+0027 +0067 | نگ 0646+06AF | نگ FEE5+FB92 | نگ FEE7+FB95 | نگ FEE8+FB95 | نگ FEE8+FB93 |
| 35 | S'H 0053+0027 +0048 | s'h 0073+0027 +0068 | سھ 0633+06BE | سھ FEB3+FBAB | سھ FEB3+FBAC | سھ FEB4+FBAD | سھ FEB4+FBAB |
| 36 | Z'H 005A+0027+ 0048 | z'h 007A+0027 +0068 | زھ 0632+06BE | زھ FEAF+FBAA | زھ FEAF+FBAC | زھ FEB0+FBAC | زھ FEB0+FBAA |
| 37 | ، 0027 | ، 0027 | ، 0027 | ، 0027 | ، 0027 | ، 0027 | ، 0027 |
| | | | غ 0626 | غ FE89 | غ FE8B | غ FE8C | غ FE8A |
| | | | | | | | |
| 38 | ، 002C | ، 002C | ، 060C | ، 060C | ، 060C | ، 060C | ، 060C |
| 39 | ؟ 003F | ؟ 003F | ؟ 061F | ؟ 061F | ؟ 061F | ؟ 061F | ؟ 061F |
| 40 | ؛ 003B | ؛ 003B | ؛ 061B | ؛ 061B | ؛ 061B | ؛ 061B | ؛ 061B |