

ICS 35.040.10

## **Zeichen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa; mit digitalem Anhang**

Characters in Unicode for the electronic processing of names and data exchange in Europe; with digital attachment

Caractères en Unicode pour le traitement électronique des noms et l'échange de données en Europe; avec adaptateur numérique

Zur Erstellung einer DIN SPEC können verschiedene Verfahrensweisen herangezogen werden:  
Das vorliegende Dokument wurde nach den Verfahrensregeln einer PAS erstellt.

Gesamtumfang 62 Seiten

Dieses Dokument wurde durch die im Vorwort genannten Verfasser erarbeitet und verabschiedet.

# Inhalt

|                                                                                           | Seite     |
|-------------------------------------------------------------------------------------------|-----------|
| <b>Vorwort .....</b>                                                                      | <b>3</b>  |
| <b>Einleitung .....</b>                                                                   | <b>5</b>  |
| <b>1 Anwendungsbereich.....</b>                                                           | <b>9</b>  |
| <b>2 Normative Verweisungen .....</b>                                                     | <b>9</b>  |
| <b>3 Begriffe .....</b>                                                                   | <b>10</b> |
| <b>4 Normative Schriftzeichen .....</b>                                                   | <b>11</b> |
| 4.1 Konformität von IT-Verfahren.....                                                     | 11        |
| 4.2 Lateinische Buchstaben.....                                                           | 11        |
| 4.3 Kombinerende diakritische Zeichen.....                                                | 26        |
| 4.4 Nicht-Buchstaben N1.....                                                              | 26        |
| 4.5 Nicht-Buchstaben N2.....                                                              | 27        |
| 4.6 Nicht-Buchstaben N3.....                                                              | 28        |
| 4.7 Nicht-Buchstaben N4.....                                                              | 28        |
| <b>5 Normative Abbildung lateinischer Buchstaben auf Grundbuchstaben (Suchform) .....</b> | <b>29</b> |
| <b>Anhang A (informativ) Erweiterte Schriftzeichen (nicht-normativ).....</b>              | <b>50</b> |
| A.1 Allgemeines .....                                                                     | 50        |
| A.2 Griechische Buchstaben.....                                                           | 50        |
| A.3 Kyrillische Buchstaben.....                                                           | 52        |
| A.4 Nicht-Buchstaben E1 .....                                                             | 54        |
| <b>Anhang B (informativ) Technische Datentypen (nicht-normativ).....</b>                  | <b>55</b> |
| B.1 Allgemeines .....                                                                     | 55        |
| B.2 Datentyp A.....                                                                       | 57        |
| B.3 Datentyp B.....                                                                       | 57        |
| B.4 Datentyp C.....                                                                       | 57        |
| B.5 Datentyp D.....                                                                       | 58        |
| B.6 Datentyp E .....                                                                      | 58        |
| <b>Anhang C (informativ) Strukturierte und tabellarische Übersicht aller Zeichen.....</b> | <b>59</b> |
| <b>Anhang D (informativ) Abbildung auf Legacy Codepages .....</b>                         | <b>61</b> |
| <b>Literaturhinweise.....</b>                                                             | <b>62</b> |

## Vorwort

Diese DIN SPEC wurde nach dem PAS-Verfahren (Publicly Available Specification) erarbeitet. Die Erarbeitung von DIN SPEC nach dem PAS-Verfahren erfolgt in Workshops und nicht zwingend unter Einbeziehung aller interessierten Kreise.

Die Erarbeitung und Verabschiedung des Dokuments erfolgten durch die nachfolgend genannten Initiatoren und Verfasser:

- Koordinierungsstelle für IT-Standards (KoSIT)  
Frank Steinke
- ]init[ AG für digitale Kommunikation  
Rico Apitz
- Amtsgericht Essen Projektgruppe RegisSTAR/AuRegis  
Helmut Lansing
- Anstalt für Kommunale Datenverarbeitung in Bayern  
Franz-Xaver Salat
- Bayerisches Landesamt für Finanzen (LfF)  
Martina Weiß
- Bayerisches Landesamt für Steuern, Projekt ELSTER  
Franz Widholm
- Brandenburgisches Oberlandesgericht  
Susann Weingardt
- Bundesagentur für Arbeit  
Alina Lorenz
- Bundesministerium des Innern, für Bau und Heimat (BMI), Referat DG I 4, IT-Steuerung Ressort BMI, Standardisierung im Bund, Open Source  
Oliver Tings
- Bundesministerium des Innern, für Bau und Heimat (BMI), Referat V II 1 – Verwaltungsverfahrenswesen, Informationszugangsrecht, Personenstandswesen  
Rainer Bockstette
- Bundesverband der Deutschen Standesbeamtinnen und Standesbeamten e.V. (BDS), Akademie für Personenstandswesen, Akademie für Staatsangehörigkeitsrecht und Meldewesen  
Beate Tripp
- Deutsche Rentenversicherung Bund, Bereich 0500  
Werner Meckelein; Dr. Clemens Mühlberger
- Deutsche Rentenversicherung Bund, Zentrale Zulagenstelle für Altersvermögen (ZfA)  
Silke Klostermann
- Freelancer  
Bernd Kappenberg
- Landeshauptstadt München, Referat für Informations- und Telekommunikationstechnik, Hauptabteilung I - IT-Strategie und IT-Steuerung / IT-Controlling (STRAC)  
Dr. Marlen Jurisch
- Westernacher Solutions AG  
Volker Dittmar

Für dieses Thema bestehen derzeit keine Normen im Deutschen Normenwerk.

DIN SPEC sind nicht Teil des Deutschen Normenwerks.

Für diese DIN SPEC wurde ein Entwurf veröffentlicht.

Trotz großer Anstrengungen zur Sicherstellung der Korrektheit, Verlässlichkeit und Präzision technischer und nicht-technischer Beschreibungen kann der Workshop weder eine explizite noch eine implizite Gewährleistung für die Korrektheit des Dokuments übernehmen. Die Benutzung dieses Dokuments geschieht in dem Bewusstsein, dass der Workshop für Schäden oder Verluste jeglicher Art nicht haftbar gemacht werden kann. Die Anwendung der vorliegenden DIN SPEC entbindet den Nutzer nicht von der Verantwortung für eigenes Handeln und geschieht damit auf eigene Gefahr.

Es wird auf die Möglichkeit hingewiesen, dass einige Elemente dieses Dokuments Patentrechte berühren können. DIN ist nicht dafür verantwortlich, einige oder alle diesbezüglichen Patentrechte zu identifizieren.

Die kostenfreie Bereitstellung dieses Dokuments als PDF-Version über den Beuth WebShop wurde im Vorfeld finanziert.

## Einleitung

### Ausgangslage

Buchstaben und andere Schriftzeichen (Nicht-Buchstaben) können in IT-Verfahren nur verarbeitet werden, indem sie auf Zahlen abgebildet werden. Es gibt auf der Welt viele verschiedene Alphabete. Für die meisten davon gibt es wiederum viele unterschiedliche Abbildungen der enthaltenen Schriftzeichen auf Zahlen (Zeichenkodierungen oder Encodings). Aus dem Nebeneinander verschiedener Kodierungssysteme für unterschiedliche Alphabete können Interoperabilitätsprobleme resultieren. Der Unicode-Standard bzw. die Norm ISO/IEC 10646 wurde entwickelt, um diese Probleme zu lösen.

Die Forderung nach vollständiger Umsetzung des Unicode-Standards auch auf der Anwendungsebene von IT-Verfahren ist jedoch in der Regel weder sachgerecht noch wirtschaftlich, weil der kulturelle Kontext der Anwender und des intendierten Anwendungsbereichs zu berücksichtigen sind.

IT-Verfahren der öffentlichen Verwaltung Deutschlands sind regelhaft auf die Verarbeitung von Daten in deutscher Sprache und somit in lateinischer Schrift ausgerichtet, denn die Amtssprache ist Deutsch. So legt beispielsweise § 23 Verwaltungsverfahrensgesetz fest: „Werden bei einer Behörde in einer fremden Sprache Anträge gestellt oder Eingaben, Belege, Urkunden oder sonstige Dokumente vorgelegt, soll die Behörde unverzüglich die Vorlage einer Übersetzung verlangen“.

Entsprechend des Übereinkommens über die Angabe von Familiennamen und Vornamen in den Personenstandsbüchern vom 13. September 1973 (BGBl. 1976 II, Seite 1473 f.) sind diese „ohne Übersetzung soweit wie möglich durch Transliteration wiederzugeben. Sind von der Internationalen Normenorganisation (ISO) empfohlene Normen vorhanden, so sind sie anzuwenden“. Sofern die Namen in lateinischen Schriftzeichen vorliegen, „so sind diese Familiennamen und Vornamen buchstabengetreu ohne Änderung oder Übersetzung wiederzugeben“. Diakritische Zeichen sind dementsprechend unverändert wiederzugeben, selbst wenn die deutsche Sprache diese Zeichen nicht kennt.

Diese Vorgaben gelten analog auch für elektronische Register und Datenübermittlungen. Die von der ISO empfohlenen Normen für Transliterationen in die lateinische Schrift sind in den Literaturhinweisen dieser DIN SPEC aufgeführt.

### Standardisierungsbedarf

Aus der Ausgangslage ergibt sich die Notwendigkeit, für IT-Verfahren (insbesondere solche der öffentlichen Verwaltung Deutschlands)

- Einerseits festzulegen, dass alle Schriftzeichen der lateinischen Schrift, die auf hoheitlichen Dokumenten und Urkunden verwendet werden können, vollumfänglich unterstützt werden müssen;
- Andererseits die Möglichkeit zu eröffnen, andere Schriftzeichen von Datenübermittlungen auszuschließen, da diese in den meisten IT-gestützten Prozessen der deutschen Verwaltung nicht akzeptiert werden.

Basierend auf dem Standard Unicode ist somit die Menge der darin enthaltenen, für das Verwaltungshandeln notwendigen und zulässigen Schriftzeichen festzulegen. Diese Teilmenge von Unicode ist anhand der nachfolgend genannten, fachlichen Überlegungen bestimmt worden:

- a) Welche lateinischen Schriftzeichen werden bei der Erfassung und Nacherfassung von Personenstandsdaten für die elektronisch geführten Personenstandsregister festgestellt?
- b) Welche lateinischen Schriftzeichen können als Ergebnis einer Transliteration (d.h. der buchstabengetreuen Übertragung von Wörtern aus einer Schrift in eine andere) aus anderen Schriften auftreten? Dabei müssen die Transliterationsregeln der ISO zugrunde gelegt werden.
- c) Welche lateinischen Schriftzeichen werden in den europäischen Amtssprachen genutzt?
- d) Welche lateinischen Schriftzeichen werden für den Datenaustausch zwischen Sozialversicherungsträgern benötigt?
- e) Welche Schriftzeichen werden für Einträge im Handelsregister benötigt?
- f) Welche Schriftzeichen werden für den EU-weiten Datenaustausch inkl. Island, Liechtenstein, Norwegen und der Schweiz zwischen Behörden benötigt?

### Abwägung von Alternativen

Es ist keine Option, die vollständige Unterstützung aller Unicode Blöcke zu fordern, in denen unter anderem lateinische Zeichen enthalten sind, denn:

- Bei der Bestimmung von Zeichenblöcken durch das Unicode-Konsortium spielen die oben genannten, fachlichen Aspekte keine oder nur eine untergeordnete Rolle;
- Es sind auch historische Zeichen ohne Relevanz für das aktuelle Verwaltungshandeln enthalten. Es wäre weder sachgerecht noch wirtschaftlich, deren technische Umsetzung zu fordern;
- Für Transliterationen notwendige kodierte Zeichensequenzen wären nicht explizit im Standard aufgeführt;
- Die Eingabe von ungewöhnlichen Schriftzeichen ist schwieriger und fehleranfälliger, wenn mehr Schriftzeichen unterstützt werden müssen, als gebraucht werden;
- Eine abschließende Liste der zulässigen Schriftzeichen ist eine Voraussetzung für die Implementation robuster Algorithmen und Verfahren, um Personen zu identifizieren und hoheitliche Dokumente und Urkunden zu erstellen. Die in Anhang B empfohlenen Datentypen dienen der Unterstützung dieses Ziels.

Zudem entstehen Sicherheitsrisiken, wenn beliebige Kombinationen von Basiszeichen mit diakritischen Zeichen zulässig sind. Dies ermöglicht optisch identische Zeichen, die informationstechnisch unterschiedlich sind. Für eine detaillierte Darstellung wird auf den Unicode Technical Report #36 „UNICODE SECURITY CONSIDERATIONS“ sowie zur Illustration auf die Webseite [unicode.org/cldr/utility/confusables.jsp](https://unicode.org/cldr/utility/confusables.jsp) verwiesen. Durch die DIN SPEC lassen sich diese Risiken ausschließen oder zumindest erheblich reduzieren. Denn sie kann in fachlich begründeten Fällen für die Einschränkung genutzt werden, ausschließlich die darin genannten Zeichen zu verarbeiten bzw. zu übermitteln. Die Durchsetzung einer entsprechenden Einschränkung kann sowohl rechtlich als auch technisch erfolgen, beispielsweise durch Nutzung der im Anhang B genannten Datentypen.

Die genannten Gründe führen zu dem Ergebnis, dass eine abschließende Auflistung der zu unterstützenden Schriftzeichen als eine echte Teilmenge der in Unicode definierten Zeichen in der DIN SPEC notwendig ist.

## Überblick

Im Sinne des Investitionsschutzes berücksichtigt die DIN SPEC den vom IT-Planungsrat und der Innenministerkonferenz derzeit bereits für den gleichen Anwendungsbereich verbindlich vorgegebenen Standard „Lateinische Zeichen in Unicode, Version 1.1 vom 27.1.2012“ (auch „String.Latin 1.1“ genannt) angemessen.

Der Legacy-Zeichensatz ISO 8859-15 war eine Grundlage der Betrachtungen bei der Zusammenstellung der Zeichen dieser DIN SPEC. Allerdings wurden einige wenige Zeichen aufgrund einer fachlichen Bewertung nicht aufgenommen.

Die DIN SPEC legt in Abschnitt 4 diejenigen Schriftzeichen fest, die von allen zu dieser DIN SPEC konformen IT-Verfahren zumindest bei allen für Namen (im weiteren Sinne) bestimmten Datenfeldern unterstützt werden MÜSSEN.

Sie empfiehlt in Anhang A diejenigen Schriftzeichen, die darüber hinaus von allen zu dieser DIN SPEC konformen IT-Verfahren unterstützt werden SOLLTEN.

Sie legt in Abschnitt 5 fest, wie die in Abschnitt 4 enthaltenen lateinischen Buchstaben auf die lateinischen Grundbuchstaben A bis Z abgebildet werden MÜSSEN. Für diesen Zweck werden die Empfehlungen des für maschinenlesbare Reisedokumente einschlägigen Standards ICAO 9303, Part 3 übernommen und erweitert.

Sie empfiehlt in Anhang B Datentypen für Einschränkungen auf bestimmte Teilmengen der in dieser DIN SPEC genannten Schriftzeichen, die insbesondere bei Schnittstellenvereinbarungen nützlich sein können.

Sie enthält Schriftzeichen in folgenden Schriftzeichengruppen (siehe Tabelle 1):

**Tabelle 1 — Schriftzeichengruppen**

| Schriftzeichengruppe                                                     | Tabelle | Anzahl Schriftzeichen | davon Zeichensequenzen |
|--------------------------------------------------------------------------|---------|-----------------------|------------------------|
| Normative lateinische Buchstaben                                         | 3       | 646                   | 147                    |
| Normative kombinierende diakritische Zeichen                             | 4       | 19                    | 0                      |
| Normative Nicht-Buchstaben, erforderlich für alle Namen                  | 5       | 18                    | 0                      |
| Sonstige normative Nicht-Buchstaben, erforderlich für nicht alle Namen   | 6       | 60                    | 0                      |
| Normative Nicht-Buchstaben, erforderlich für Namen juristischer Personen | 7       | 6                     | 0                      |
| Normative Nicht-Buchstaben, für Namen ungeeignet                         | 8       | 4                     | 0                      |
| Griechische Buchstaben                                                   | A.1     | 69                    | 0                      |
| Kyrillische Buchstaben                                                   | A.2     | 62                    | 0                      |
| Sonstige erweiterte Nicht-Buchstaben, erforderlich für nicht alle Namen  | A.3     | 40                    | 0                      |
| <b>Summe</b>                                                             |         | <b>924</b>            | <b>147</b>             |

„Lateinische Zeichen“ im Sinne dieser DIN SPEC sind alle in den Tabellen 3 bis 8 und der Tabelle A.3 enthaltenen Schriftzeichen, also sowohl die Buchstaben als auch die Nicht-Buchstaben.

Die Schriftzeichengruppen werden in nachfolgenden Abschnitten erläutert und durch tabellarische Aufzählung der darin enthaltenen Schriftzeichen dargestellt. Innerhalb jeder Tabelle der in einer Gruppe enthaltenen Schriftzeichen sind diese nach ihrem Codepoint (bei kodierten Zeichen) bzw. ihren Codepoints (bei kodierten Zeichensequenzen) geordnet. Für jedes kodierte Zeichen wird dessen Codepoint und der vom Unicode Standard zugeordnete Name genannt. Für jede kodierte Zeichensequenz werden deren Codepoints und deren Name genannt. Der angegebene Name der Sequenz ist der vom Unicode Standard festgelegte Name, sofern es sich um eine „named sequence“ gemäß des Unicode Standard Annex #34 „UNICODE NAMED CHARACTER SEQUENCES“ handelt. Andernfalls wird der Name unter analoger Anwendung der in Abschnitt 4 des Unicode Standard Annex #34 genannten Regeln zur Bildung des Namens kodierter Zeichensequenzen festgelegt.

Für jedes darstellbare Zeichen wird eine Glyphe angegeben, die an der für die ISO/IEC 10646 gewählten Darstellung orientiert ist. Für jedes nicht darstellbare Zeichen wird ersatzweise die Glyphe „“ (REPLACEMENT CHARACTER, U+FFFD) verwendet. Die Darstellung der Zeichen durch Glyphen dient nur der Illustration. Für Implementationen der DIN SPEC muss die Angabe der Codepoints herangezogen werden.

Diese DIN SPEC wird ergänzt durch die nachfolgend in Tabelle 2 genannten Dateien, die informativ beigelegt werden:

**Tabelle 2 — Angehängte Dateien**

| <b>Datei</b>                    | <b>Beschreibung</b>                                                                            |
|---------------------------------|------------------------------------------------------------------------------------------------|
| din-91379-datatypes.xsd         | Eine technische Umsetzung der in Anhang B beschriebenen Datentypen                             |
| latinchars.xml                  | Eine XML Datei mit allen in dieser DIN SPEC genannten Schriftzeichen und Schriftzeichengruppen |
| latinchars.xsd                  | eine XML Schema 1.0 Datei mit der Beschreibung der Struktur der Datei latinchars.xml           |
| 2018-11-15_din-spec-91379.xlsx  | Eine tabellarische Übersicht über alle in dieser DIN SPEC genannten Schriftzeichen             |
| 2018-11-15_legacy-mappings.xlsx | Eine Abbildung der in dieser DIN SPEC enthaltenen Schriftzeichen auf Legacy Codepages          |

## 1 Anwendungsbereich

Diese DIN SPEC 91379 „Zeichen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa“ legt die Zeichen fest, die für Namen im weiteren Sinne benötigt werden. Namen im weiteren Sinne sind Namen natürlicher Personen gemäß der Vorgaben des Personenstandsrechts ebenso wie Namen juristischer Personen, Namen von Produkten, Orten und Straßen sowie Titel von Dokumenten oder Gesetzen. Namen im weiteren Sinne können konkrete Objekte bezeichnen, aber auch virtuelle Konstrukte wie Produktgruppen oder Musikstile.

Die DIN SPEC teilt sich auf in einen normativen Teil und einen nicht-normativen Teil.

Der normative Teil der DIN SPEC legt diejenige Teilmenge der in Unicode enthaltenen Zeichen und Zeichensequenzen fest, die für IT-Verfahren für die elektronische Verarbeitung von Namen (im weiteren Sinne) auf Basis der lateinischen Schrift erforderlich ist. Sie MUSS deshalb von allen zu dieser Spezifikation konformen IT-Verfahren zumindest bei allen für Namen (im weiteren Sinne) bestimmten Datenfeldern unterstützt werden. Im normativen Teil wird auch eine gegebenenfalls erforderliche Abbildung der normativen Buchstaben auf die Großbuchstaben A bis Z bestimmt.

Im nicht-normativen Teil der DIN SPEC werden darüber hinaus ergänzende Zeichen aus Unicode festgelegt, die eventuell zur Bildung von Namen juristischer Personen oder von Produkten benötigt werden. Im Kontext des EU-weiten Datenaustauschs ergibt sich ein erweiterter Bedarf an griechischen und kyrillischen Zeichen.

Durch den normativen Teil wird eine Mindestmenge von Schriftzeichen bestimmt. Zur DIN SPEC konforme IT-Verfahren dürfen darüber hinausgehende Schriftzeichen unterstützen. In Schnittstellenvereinbarungen ist es jedoch erforderlich, den jeweiligen Zeichenvorrat abschließend festzulegen. Die DIN SPEC stellt in Anhang B Datentypen für fachlich bestimmte Anwendungsfälle zur Verfügung. Dazu werden bestimmte Schriftzeichengruppen zu einer Menge zulässiger Schriftzeichen zusammengestellt. Die informell beigefügte Datei `din-91379-datatypes.xsd` enthält technische Umsetzungen dieser Datentypen, welche ausschließlich die jeweils zugrunde liegenden Schriftzeichen akzeptieren.

Diese DIN SPEC basiert auf dem Unicode-Standard und wendet sich vorrangig an Behörden und Organisationen, die IT-Verfahren betreiben, welche dem behördenübergreifenden Datenaustausch oder dem Datenaustausch mit Bürgern und Wirtschaft dienen. Sie deckt in ihrer Gesamtheit die EU-Amtssprachen und die Amtssprachen Islands, Liechtensteins, Norwegens und der Schweiz sowie die deutschen Minderheitensprachen ab.

Alle in dieser DIN SPEC genannten Zeichen und Zeichensequenzen liegen in der Normalization Form C (NFC) des Unicode Standard Annex #15 „Unicode Normalization Forms“ vor.

Die DIN SPEC trifft weder Aussagen zum Umgang mit historischen Zeichen noch zum Umgang mit Fließtext. Die DIN SPEC regelt nicht die Darstellung von Zeichen (Glyphen).

## 2 Normative Verweisungen

Die folgenden Dokumente werden im Text in solcher Weise in Bezug genommen, dass einige Teile davon oder ihr gesamter Inhalt Anforderungen des vorliegenden Dokuments darstellen. Bei datierten Verweisungen gilt nur die in Bezug genommene Ausgabe. Bei undatierten Verweisungen gilt die letzte Ausgabe des in Bezug genommenen Dokuments (einschließlich aller Änderungen).

ISO/IEC 10646, *Information technology — Universal Coded Character Set (UCS)*

### 3 Begriffe

Die für diese Spezifikation relevanten Definitionen des Unicode-Standards werden unverändert übernommen. Für die Anwendung dieses Dokuments gelten die folgenden Begriffe.

DIN und DKE stellen terminologische Datenbanken für die Verwendung in der Normung unter den folgenden Adressen bereit:

- DIN-TERMinologieportal: unter <https://www.din.de/go/din-term>
- DKE-IEV: unter <http://www.dke.de/DKE-IEV>

#### 3.1

##### **abstraktes Zeichen**

(en: abstract character)

Informationseinheit, die für die Organisation, Kontrolle oder Darstellung von Textdaten verwendet wird

[QUELLE: The Unicode Standard, Version 9.0, 3.4, D.7]

#### 3.2

##### **Codepoint**

Wert innerhalb des Wertebereichs der ganzen Zahlen von 0 bis  $10FFFF_{16}$

Anmerkung 1 zum Begriff: Der Wert repräsentiert die Position eines Zeichens im Unicode-Standard.

Anmerkung 2 zum Begriff: Für diese DIN SPEC kommen nur Zeichen aus der „Plane 0“ des Unicode-Standards zum Einsatz, sodass der Wert eines Codepoints  $FFFF_{16}$  nicht überschreiten kann.

[QUELLE: The Unicode Standard, Version 9.0, 3.4, D.10 in Verbindung mit D.9]

#### 3.3

##### **kodiertes Zeichen**

(en: encoded character, coded character)

##### **Zeichen**

(en: character)

Verbindung (oder Abbildung) zwischen einem abstrakten Zeichen und einem Codepoint

Anmerkung 1 zum Begriff: Der Begriff „Zeichen“ ist allein stehend als Abkürzung für den Begriff „kodiertes Zeichen“ zu verstehen. Siehe hierfür die Erläuterung in Unicode 9.0, 3.4, zwischen den Definitionen D.12 und D.13.

[QUELLE: The Unicode Standard, Version 9.0, 3.4, D.11]

#### 3.4

##### **kodierte Zeichensequenz**

(en: coded character sequence)

##### **Zeichensequenz**

(en: character sequence)

geordnete Sequenz eines oder mehrerer Codepoints

Anmerkung 1 zum Begriff: Der Begriff „Zeichensequenz“ steht als Abkürzung für den Begriff „kodierte Zeichensequenz“. Siehe hierfür die Erläuterung in Unicode 9.0, 3.4, zwischen den Definitionen D.12 und D.13.

[QUELLE: The Unicode Standard, Version 9.0, 3.4, D.12]

### 3.5

#### Schriftzeichen

kleinste im Schreibfluss aufeinanderfolgende Einheiten einer Schrift, z. B. des lateinischen oder kyrillischen Alphabets, die sprachunabhängig sind, da sie für verschiedene Sprachen genutzt werden können, z. B. das lateinische Alphabet für Deutsch und Französisch

Anmerkung 1 zum Begriff: Zu den Schriftzeichen gehören sowohl kodierte Zeichen als auch kodierte Zeichensequenzen, jedoch keine einzelnen kombinierenden diakritischen Zeichen außerhalb von Zeichensequenzen.

Anmerkung 2 zum Begriff: Im Rahmen dieser DIN SPEC werden die Schriftzeichen unterschieden in Buchstaben und Nicht-Buchstaben (z. B. Interpunktion, Ziffern und Leerzeichen).

Anmerkung 3 zum Begriff: Die Menge der Schriftzeichen, deren Unterstützung durch IT-Verfahren für Zwecke der elektronischen Verarbeitung von Namen und den Datenaustausch in Europa vorgeschrieben bzw. empfohlen wird, ist eine echte Teilmenge der im Unicode-Standard insgesamt enthaltenen Schriftzeichen.

Anmerkung 4 zum Begriff: In dieser DIN SPEC wird der Begriff „Schriftzeichen“ weitgehend synonym zum Begriff „Graphem“ verwendet. Es werden aber auch „Plurigraphen“ mit eingeschlossen.

### 3.6

#### Schriftzeichengruppe

Zusammenfassung von Schriftzeichen unter fachlichen Aspekten

Anmerkung 1 zum Begriff: Jedes der in dieser Spezifikation genannten Schriftzeichen ist genau einer Schriftzeichengruppe zugeordnet. Jeder Schriftzeichengruppe ist mindestens ein Schriftzeichen zugeordnet. Jede Schriftzeichengruppe hat einen innerhalb dieser Spezifikation eindeutigen Namen, z. B. „Lateinische Buchstaben“ oder „Nicht-Buchstaben N1“.

Anmerkung 2 zum Begriff: In Tabelle 1 sind alle Schriftzeichengruppen genannt.

## 4 Normative Schriftzeichen

### 4.1 Konformität von IT-Verfahren

Jedes konforme IT-Verfahren MUSS die in Tabelle 3 bis Tabelle 8 genannten Schriftzeichen bei allen für Namen (im weiteren Sinne) bestimmten Datenfeldern in vollem Umfang verarbeiten können. Dies umfasst insbesondere die Erfassung, Speicherung, Übermittlung, Anzeige und den Ausdruck aller Schriftzeichen.

### 4.2 Lateinische Buchstaben

Die in Tabelle 3 genannten Schriftzeichen für lateinische Buchstaben sind für die Darstellung von Namen im weiteren Sinne zwingend erforderlich.

**Tabelle 3 — Normative lateinische Buchstaben**

| Codepoints      | Name                                                      | Glyphe |
|-----------------|-----------------------------------------------------------|--------|
| U+0041          | LATIN CAPITAL LETTER A                                    | A      |
| <U+0041,U+030B> | LATIN CAPITAL LETTER A WITH COMBINING DOUBLE ACUTE ACCENT | Ā      |
| U+0042          | LATIN CAPITAL LETTER B                                    | B      |
| U+0043          | LATIN CAPITAL LETTER C                                    | C      |
| <U+0043,U+0300> | LATIN CAPITAL LETTER C WITH COMBINING GRAVE ACCENT        | Ĉ      |
| <U+0043,U+0304> | LATIN CAPITAL LETTER C WITH COMBINING MACRON              | Ċ      |
| <U+0043,U+0306> | LATIN CAPITAL LETTER C WITH COMBINING BREVE               | Č      |
| <U+0043,U+0308> | LATIN CAPITAL LETTER C WITH COMBINING DIAERESIS           | Ĉ      |
| <U+0043,U+0315> | LATIN CAPITAL LETTER C WITH COMBINING COMMA ABOVE RIGHT   | Ć      |

| Codepoints             | Name                                                                                 | Glyphe |
|------------------------|--------------------------------------------------------------------------------------|--------|
| <U+0043,U+0323>        | LATIN CAPITAL LETTER C WITH COMBINING DOT BELOW                                      | Ç      |
| <U+0043,U+0326>        | LATIN CAPITAL LETTER C WITH COMBINING COMMA BELOW                                    | Ç      |
| <U+0043,U+0328,U+0306> | LATIN CAPITAL LETTER C WITH COMBINING OGONEK AND COMBINING BREVE                     | Ć      |
| U+0044                 | LATIN CAPITAL LETTER D                                                               | D      |
| <U+0044,U+0302>        | LATIN CAPITAL LETTER D WITH COMBINING CIRCUMFLEX ACCENT                              | Ď      |
| U+0045                 | LATIN CAPITAL LETTER E                                                               | E      |
| U+0046                 | LATIN CAPITAL LETTER F                                                               | F      |
| <U+0046,U+0300>        | LATIN CAPITAL LETTER F WITH COMBINING GRAVE ACCENT                                   | Ě      |
| <U+0046,U+0304>        | LATIN CAPITAL LETTER F WITH COMBINING MACRON                                         | Ě      |
| U+0047                 | LATIN CAPITAL LETTER G                                                               | G      |
| <U+0047,U+0300>        | LATIN CAPITAL LETTER G WITH COMBINING GRAVE ACCENT                                   | Ĝ      |
| U+0048                 | LATIN CAPITAL LETTER H                                                               | H      |
| <U+0048,U+0304>        | LATIN CAPITAL LETTER H WITH COMBINING MACRON                                         | Ĥ      |
| <U+0048,U+0326>        | LATIN CAPITAL LETTER H WITH COMBINING COMMA BELOW                                    | Ĥ      |
| <U+0048,U+0331>        | LATIN CAPITAL LETTER H WITH COMBINING MACRON BELOW                                   | Ĥ      |
| U+0049                 | LATIN CAPITAL LETTER I                                                               | I      |
| U+004A                 | LATIN CAPITAL LETTER J                                                               | J      |
| <U+004A,U+0301>        | LATIN CAPITAL LETTER J WITH COMBINING ACUTE ACCENT                                   | Ĵ      |
| <U+004A,U+030C>        | LATIN CAPITAL LETTER J WITH COMBINING CARON                                          | Ĵ      |
| U+004B                 | LATIN CAPITAL LETTER K                                                               | K      |
| <U+004B,U+0300>        | LATIN CAPITAL LETTER K WITH COMBINING GRAVE ACCENT                                   | Ķ      |
| <U+004B,U+0302>        | LATIN CAPITAL LETTER K WITH COMBINING CIRCUMFLEX ACCENT                              | Ķ      |
| <U+004B,U+0304>        | LATIN CAPITAL LETTER K WITH COMBINING MACRON                                         | Ķ      |
| <U+004B,U+0307>        | LATIN CAPITAL LETTER K WITH COMBINING DOT ABOVE                                      | Ķ      |
| <U+004B,U+0315>        | LATIN CAPITAL LETTER K WITH COMBINING COMMA ABOVE RIGHT                              | Ķ      |
| <U+004B,U+031B>        | LATIN CAPITAL LETTER K WITH COMBINING HORN                                           | Ķ      |
| <U+004B,U+0326>        | LATIN CAPITAL LETTER K WITH COMBINING COMMA BELOW                                    | K      |
| <U+004B,U+035F,U+0048> | LATIN CAPITAL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN CAPITAL LETTER H | ĶH     |
| <U+004B,U+035F,U+0068> | LATIN CAPITAL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN SMALL LETTER H   | Ķh     |
| U+004C                 | LATIN CAPITAL LETTER L                                                               | L      |
| <U+004C,U+0302>        | LATIN CAPITAL LETTER L WITH COMBINING CIRCUMFLEX ACCENT                              | L      |
| <U+004C,U+0325>        | LATIN CAPITAL LETTER L WITH COMBINING RING BELOW                                     | Ł      |
| <U+004C,U+0325,U+0304> | LATIN CAPITAL LETTER L WITH COMBINING RING BELOW AND COMBINING MACRON                | Ł      |
| <U+004C,U+0326>        | LATIN CAPITAL LETTER L WITH COMBINING COMMA BELOW                                    | L      |
| U+004D                 | LATIN CAPITAL LETTER M                                                               | M      |
| <U+004D,U+0300>        | LATIN CAPITAL LETTER M WITH COMBINING GRAVE ACCENT                                   | Ĺ      |
| <U+004D,U+0302>        | LATIN CAPITAL LETTER M WITH COMBINING CIRCUMFLEX ACCENT                              | M      |
| <U+004D,U+0306>        | LATIN CAPITAL LETTER M WITH COMBINING BREVE                                          | Ĺ      |
| <U+004D,U+0310>        | LATIN CAPITAL LETTER M WITH COMBINING CANDRABINDU                                    | Ṁ      |
| U+004E                 | LATIN CAPITAL LETTER N                                                               | N      |
| <U+004E,U+0302>        | LATIN CAPITAL LETTER N WITH COMBINING CIRCUMFLEX ACCENT                              | N      |
| <U+004E,U+0304>        | LATIN CAPITAL LETTER N WITH COMBINING MACRON                                         | Ñ      |
| <U+004E,U+0306>        | LATIN CAPITAL LETTER N WITH COMBINING BREVE                                          | Ñ      |
| <U+004E,U+0326>        | LATIN CAPITAL LETTER N WITH COMBINING COMMA BELOW                                    | N      |
| U+004F                 | LATIN CAPITAL LETTER O                                                               | O      |
| U+0050                 | LATIN CAPITAL LETTER P                                                               | P      |

| Codepoints             | Name                                                                  | Glyphe |
|------------------------|-----------------------------------------------------------------------|--------|
| <U+0050,U+0300>        | LATIN CAPITAL LETTER P WITH COMBINING GRAVE ACCENT                    | Ṗ      |
| <U+0050,U+0304>        | LATIN CAPITAL LETTER P WITH COMBINING MACRON                          | Ṕ      |
| <U+0050,U+0315>        | LATIN CAPITAL LETTER P WITH COMBINING COMMA ABOVE RIGHT               | Ṗ́     |
| <U+0050,U+0323>        | LATIN CAPITAL LETTER P WITH COMBINING DOT BELOW                       | Ṗ̣     |
| U+0051                 | LATIN CAPITAL LETTER Q                                                | Q      |
| U+0052                 | LATIN CAPITAL LETTER R                                                | R      |
| <U+0052,U+0306>        | LATIN CAPITAL LETTER R WITH COMBINING BREVE                           | Ṛ      |
| <U+0052,U+0325>        | LATIN CAPITAL LETTER R WITH COMBINING RING BELOW                      | Ṛ      |
| <U+0052,U+0325,U+0304> | LATIN CAPITAL LETTER R WITH COMBINING RING BELOW AND COMBINING MACRON | Ṝ     |
| U+0053                 | LATIN CAPITAL LETTER S                                                | S      |
| <U+0053,U+0300>        | LATIN CAPITAL LETTER S WITH COMBINING GRAVE ACCENT                    | Ṣ      |
| <U+0053,U+0304>        | LATIN CAPITAL LETTER S WITH COMBINING MACRON                          | Ṣ̄     |
| <U+0053,U+031B,U+0304> | LATIN CAPITAL LETTER S WITH COMBINING HORN AND COMBINING MACRON       | Ṣ̄́    |
| <U+0053,U+0331>        | LATIN CAPITAL LETTER S WITH COMBINING MACRON BELOW                    | Ṣ̵     |
| U+0054                 | LATIN CAPITAL LETTER T                                                | T      |
| <U+0054,U+0300>        | LATIN CAPITAL LETTER T WITH COMBINING GRAVE ACCENT                    | Ṭ      |
| <U+0054,U+0304>        | LATIN CAPITAL LETTER T WITH COMBINING MACRON                          | Ṭ̄     |
| <U+0054,U+0308>        | LATIN CAPITAL LETTER T WITH COMBINING DIAERESIS                       | Ṭ̈     |
| <U+0054,U+0315>        | LATIN CAPITAL LETTER T WITH COMBINING COMMA ABOVE RIGHT               | Ṭ́     |
| <U+0054,U+031B>        | LATIN CAPITAL LETTER T WITH COMBINING HORN                            | Ṭ́     |
| U+0055                 | LATIN CAPITAL LETTER U                                                | U      |
| <U+0055,U+0307>        | LATIN CAPITAL LETTER U WITH COMBINING DOT ABOVE                       | Ṫ      |
| U+0056                 | LATIN CAPITAL LETTER V                                                | V      |
| U+0057                 | LATIN CAPITAL LETTER W                                                | W      |
| U+0058                 | LATIN CAPITAL LETTER X                                                | X      |
| U+0059                 | LATIN CAPITAL LETTER Y                                                | Y      |
| U+005A                 | LATIN CAPITAL LETTER Z                                                | Z      |
| <U+005A,U+0300>        | LATIN CAPITAL LETTER Z WITH COMBINING GRAVE ACCENT                    | Ṛ      |
| <U+005A,U+0304>        | LATIN CAPITAL LETTER Z WITH COMBINING MACRON                          | Ṝ     |
| <U+005A,U+0306>        | LATIN CAPITAL LETTER Z WITH COMBINING BREVE                           | Ṛ̆     |
| <U+005A,U+0308>        | LATIN CAPITAL LETTER Z WITH COMBINING DIAERESIS                       | Ṛ̈     |
| <U+005A,U+0327>        | LATIN CAPITAL LETTER Z WITH COMBINING CEDILLA                         | Ṛ̸     |
| U+0061                 | LATIN SMALL LETTER A                                                  | a      |
| <U+0061,U+030B>        | LATIN SMALL LETTER A WITH COMBINING DOUBLE ACUTE ACCENT               | ă      |
| U+0062                 | LATIN SMALL LETTER B                                                  | b      |
| U+0063                 | LATIN SMALL LETTER C                                                  | c      |
| <U+0063,U+0300>        | LATIN SMALL LETTER C WITH COMBINING GRAVE ACCENT                      | ç      |
| <U+0063,U+0304>        | LATIN SMALL LETTER C WITH COMBINING MACRON                            | ċ      |
| <U+0063,U+0306>        | LATIN SMALL LETTER C WITH COMBINING BREVE                             | ċ̆     |
| <U+0063,U+0308>        | LATIN SMALL LETTER C WITH COMBINING DIAERESIS                         | ċ̈     |
| <U+0063,U+0315>        | LATIN SMALL LETTER C WITH COMBINING COMMA ABOVE RIGHT                 | ċ́     |
| <U+0063,U+0323>        | LATIN SMALL LETTER C WITH COMBINING DOT BELOW                         | ċ̣     |
| <U+0063,U+0326>        | LATIN SMALL LETTER C WITH COMBINING COMMA BELOW                       | ċ̶     |
| <U+0063,U+0328,U+0306> | LATIN SMALL LETTER C WITH COMBINING OGONEK AND COMBINING BREVE        | ċ̆̇    |
| U+0064                 | LATIN SMALL LETTER D                                                  | d      |
| <U+0064,U+0302>        | LATIN SMALL LETTER D WITH COMBINING CIRCUMFLEX ACCENT                 | đ      |

| Codepoints             | Name                                                                             | Glyphe |
|------------------------|----------------------------------------------------------------------------------|--------|
| U+0065                 | LATIN SMALL LETTER E                                                             | e      |
| U+0066                 | LATIN SMALL LETTER F                                                             | f      |
| <U+0066,U+0300>        | LATIN SMALL LETTER F WITH COMBINING GRAVE ACCENT                                 | ƒ      |
| <U+0066,U+0304>        | LATIN SMALL LETTER F WITH COMBINING MACRON                                       | ƒ̄     |
| U+0067                 | LATIN SMALL LETTER G                                                             | g      |
| <U+0067,U+0300>        | LATIN SMALL LETTER G WITH COMBINING GRAVE ACCENT                                 | ġ      |
| U+0068                 | LATIN SMALL LETTER H                                                             | h      |
| <U+0068,U+0304>        | LATIN SMALL LETTER H WITH COMBINING MACRON                                       | ĥ      |
| <U+0068,U+0326>        | LATIN SMALL LETTER H WITH COMBINING COMMA BELOW                                  | h̸     |
| U+0069                 | LATIN SMALL LETTER I                                                             | i      |
| U+006A                 | LATIN SMALL LETTER J                                                             | j      |
| <U+006A,U+0301>        | LATIN SMALL LETTER J WITH COMBINING ACUTE ACCENT                                 | ĵ      |
| U+006B                 | LATIN SMALL LETTER K                                                             | k      |
| <U+006B,U+0300>        | LATIN SMALL LETTER K WITH COMBINING GRAVE ACCENT                                 | ķ      |
| <U+006B,U+0302>        | LATIN SMALL LETTER K WITH COMBINING CIRCUMFLEX ACCENT                            | ķ̂     |
| <U+006B,U+0304>        | LATIN SMALL LETTER K WITH COMBINING MACRON                                       | ķ̄     |
| <U+006B,U+0307>        | LATIN SMALL LETTER K WITH COMBINING DOT ABOVE                                    | ķ̣     |
| <U+006B,U+0315>        | LATIN SMALL LETTER K WITH COMBINING COMMA ABOVE RIGHT                            | ķ̸ʹ    |
| <U+006B,U+031B>        | LATIN SMALL LETTER K WITH COMBINING HORN                                         | ķ̴     |
| <U+006B,U+0326>        | LATIN SMALL LETTER K WITH COMBINING COMMA BELOW                                  | k̸     |
| <U+006B,U+035F,U+0068> | LATIN SMALL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN SMALL LETTER H | kh     |
| U+006C                 | LATIN SMALL LETTER L                                                             | l      |
| <U+006C,U+0302>        | LATIN SMALL LETTER L WITH COMBINING CIRCUMFLEX ACCENT                            | Ľ      |
| <U+006C,U+0325>        | LATIN SMALL LETTER L WITH COMBINING RING BELOW                                   | ł      |
| <U+006C,U+0325,U+0304> | LATIN SMALL LETTER L WITH COMBINING RING BELOW AND COMBINING MACRON              | Ł      |
| <U+006C,U+0326>        | LATIN SMALL LETTER L WITH COMBINING COMMA BELOW                                  | l̸     |
| U+006D                 | LATIN SMALL LETTER M                                                             | m      |
| <U+006D,U+0300>        | LATIN SMALL LETTER M WITH COMBINING GRAVE ACCENT                                 | ṁ      |
| <U+006D,U+0302>        | LATIN SMALL LETTER M WITH COMBINING CIRCUMFLEX ACCENT                            | Ṃ      |
| <U+006D,U+0306>        | LATIN SMALL LETTER M WITH COMBINING BREVE                                        | Ṣ      |
| <U+006D,U+0310>        | LATIN SMALL LETTER M WITH COMBINING CANDRABINDU                                  | Ṡ      |
| U+006E                 | LATIN SMALL LETTER N                                                             | n      |
| <U+006E,U+0302>        | LATIN SMALL LETTER N WITH COMBINING CIRCUMFLEX ACCENT                            | ņ      |
| <U+006E,U+0304>        | LATIN SMALL LETTER N WITH COMBINING MACRON                                       | ñ      |
| <U+006E,U+0306>        | LATIN SMALL LETTER N WITH COMBINING BREVE                                        | ṇ      |
| <U+006E,U+0326>        | LATIN SMALL LETTER N WITH COMBINING COMMA BELOW                                  | n̸     |
| U+006F                 | LATIN SMALL LETTER O                                                             | o      |
| U+0070                 | LATIN SMALL LETTER P                                                             | p      |
| <U+0070,U+0300>        | LATIN SMALL LETTER P WITH COMBINING GRAVE ACCENT                                 | ṑ      |
| <U+0070,U+0304>        | LATIN SMALL LETTER P WITH COMBINING MACRON                                       | ṑ̄     |
| <U+0070,U+0315>        | LATIN SMALL LETTER P WITH COMBINING COMMA ABOVE RIGHT                            | p̸ʹ    |
| <U+0070,U+0323>        | LATIN SMALL LETTER P WITH COMBINING DOT BELOW                                    | p̣     |
| U+0071                 | LATIN SMALL LETTER Q                                                             | q      |
| U+0072                 | LATIN SMALL LETTER R                                                             | r      |
| <U+0072,U+0306>        | LATIN SMALL LETTER R WITH COMBINING BREVE                                        | ṛ      |
| <U+0072,U+0325>        | LATIN SMALL LETTER R WITH COMBINING RING BELOW                                   | ŕ      |

| Codepoints             | Name                                                                | Glyphe |
|------------------------|---------------------------------------------------------------------|--------|
| <U+0072,U+0325,U+0304> | LATIN SMALL LETTER R WITH COMBINING RING BELOW AND COMBINING MACRON | ṛ      |
| U+0073                 | LATIN SMALL LETTER S                                                | s      |
| <U+0073,U+0300>        | LATIN SMALL LETTER S WITH COMBINING GRAVE ACCENT                    | š      |
| <U+0073,U+0304>        | LATIN SMALL LETTER S WITH COMBINING MACRON                          | s̄     |
| <U+0073,U+031B,U+0304> | LATIN SMALL LETTER S WITH COMBINING HORN AND COMBINING MACRON       | ṡ      |
| <U+0073,U+0331>        | LATIN SMALL LETTER S WITH COMBINING MACRON BELOW                    | s̕     |
| U+0074                 | LATIN SMALL LETTER T                                                | t      |
| <U+0074,U+0300>        | LATIN SMALL LETTER T WITH COMBINING GRAVE ACCENT                    | ț      |
| <U+0074,U+0304>        | LATIN SMALL LETTER T WITH COMBINING MACRON                          | t̄     |
| <U+0074,U+0315>        | LATIN SMALL LETTER T WITH COMBINING COMMA ABOVE RIGHT               | ṫ      |
| <U+0074,U+031B>        | LATIN SMALL LETTER T WITH COMBINING HORN                            | ṭ      |
| U+0075                 | LATIN SMALL LETTER U                                                | u      |
| <U+0075,U+0307>        | LATIN SMALL LETTER U WITH COMBINING DOT ABOVE                       | û      |
| U+0076                 | LATIN SMALL LETTER V                                                | v      |
| U+0077                 | LATIN SMALL LETTER W                                                | w      |
| U+0078                 | LATIN SMALL LETTER X                                                | x      |
| U+0079                 | LATIN SMALL LETTER Y                                                | y      |
| U+007A                 | LATIN SMALL LETTER Z                                                | z      |
| <U+007A,U+0300>        | LATIN SMALL LETTER Z WITH COMBINING GRAVE ACCENT                    | ž      |
| <U+007A,U+0304>        | LATIN SMALL LETTER Z WITH COMBINING MACRON                          | z̄     |
| <U+007A,U+0306>        | LATIN SMALL LETTER Z WITH COMBINING BREVE                           | z̆     |
| <U+007A,U+0308>        | LATIN SMALL LETTER Z WITH COMBINING DIAERESIS                       | z̈     |
| <U+007A,U+0327>        | LATIN SMALL LETTER Z WITH COMBINING CEDILLA                         | ẓ      |
| U+00C0                 | LATIN CAPITAL LETTER A WITH GRAVE                                   | À      |
| U+00C1                 | LATIN CAPITAL LETTER A WITH ACUTE                                   | Á      |
| U+00C2                 | LATIN CAPITAL LETTER A WITH CIRCUMFLEX                              | Â      |
| U+00C3                 | LATIN CAPITAL LETTER A WITH TILDE                                   | Ã      |
| U+00C4                 | LATIN CAPITAL LETTER A WITH DIAERESIS                               | Ä      |
| U+00C5                 | LATIN CAPITAL LETTER A WITH RING ABOVE                              | Å      |
| U+00C6                 | LATIN CAPITAL LETTER AE                                             | Æ      |
| U+00C7                 | LATIN CAPITAL LETTER C WITH CEDILLA                                 | Ç      |
| <U+00C7,U+0306>        | LATIN CAPITAL LETTER C WITH CEDILLA WITH COMBINING BREVE            | Č      |
| U+00C8                 | LATIN CAPITAL LETTER E WITH GRAVE                                   | È      |
| U+00C9                 | LATIN CAPITAL LETTER E WITH ACUTE                                   | É      |
| U+00CA                 | LATIN CAPITAL LETTER E WITH CIRCUMFLEX                              | Ê      |
| U+00CB                 | LATIN CAPITAL LETTER E WITH DIAERESIS                               | Ë      |
| U+00CC                 | LATIN CAPITAL LETTER I WITH GRAVE                                   | Ì      |
| U+00CD                 | LATIN CAPITAL LETTER I WITH ACUTE                                   | Í      |
| U+00CE                 | LATIN CAPITAL LETTER I WITH CIRCUMFLEX                              | Î      |
| U+00CF                 | LATIN CAPITAL LETTER I WITH DIAERESIS                               | Ï      |
| U+00D0                 | LATIN CAPITAL LETTER ETH                                            | Ð      |
| U+00D1                 | LATIN CAPITAL LETTER N WITH TILDE                                   | Ñ      |
| U+00D2                 | LATIN CAPITAL LETTER O WITH GRAVE                                   | Ò      |
| U+00D3                 | LATIN CAPITAL LETTER O WITH ACUTE                                   | Ó      |
| U+00D4                 | LATIN CAPITAL LETTER O WITH CIRCUMFLEX                              | Ô      |
| U+00D5                 | LATIN CAPITAL LETTER O WITH TILDE                                   | Õ      |
| U+00D6                 | LATIN CAPITAL LETTER O WITH DIAERESIS                               | Ö      |

| Codepoints      | Name                                                            | Glyphe |
|-----------------|-----------------------------------------------------------------|--------|
| U+00D8          | LATIN CAPITAL LETTER O WITH STROKE                              | Ø      |
| U+00D9          | LATIN CAPITAL LETTER U WITH GRAVE                               | Ù      |
| U+00DA          | LATIN CAPITAL LETTER U WITH ACUTE                               | Ú      |
| U+00DB          | LATIN CAPITAL LETTER U WITH CIRCUMFLEX                          | Û      |
| <U+00DB,U+0304> | LATIN CAPITAL LETTER U WITH CIRCUMFLEX WITH COMBINING MACRON    | Ū      |
| U+00DC          | LATIN CAPITAL LETTER U WITH DIAERESIS                           | Ü      |
| U+00DD          | LATIN CAPITAL LETTER Y WITH ACUTE                               | Ý      |
| U+00DE          | LATIN CAPITAL LETTER THORN                                      | þ      |
| U+00DF          | LATIN SMALL LETTER SHARP S                                      | ß      |
| U+00E0          | LATIN SMALL LETTER A WITH GRAVE                                 | à      |
| U+00E1          | LATIN SMALL LETTER A WITH ACUTE                                 | á      |
| U+00E2          | LATIN SMALL LETTER A WITH CIRCUMFLEX                            | â      |
| U+00E3          | LATIN SMALL LETTER A WITH TILDE                                 | ã      |
| U+00E4          | LATIN SMALL LETTER A WITH DIAERESIS                             | ä      |
| U+00E5          | LATIN SMALL LETTER A WITH RING ABOVE                            | å      |
| U+00E6          | LATIN SMALL LETTER AE                                           | æ      |
| U+00E7          | LATIN SMALL LETTER C WITH CEDILLA                               | ç      |
| <U+00E7,U+0306> | LATIN SMALL LETTER C WITH CEDILLA WITH COMBINING BREVE          | ç      |
| U+00E8          | LATIN SMALL LETTER E WITH GRAVE                                 | è      |
| U+00E9          | LATIN SMALL LETTER E WITH ACUTE                                 | é      |
| U+00EA          | LATIN SMALL LETTER E WITH CIRCUMFLEX                            | ê      |
| U+00EB          | LATIN SMALL LETTER E WITH DIAERESIS                             | ë      |
| U+00EC          | LATIN SMALL LETTER I WITH GRAVE                                 | ì      |
| U+00ED          | LATIN SMALL LETTER I WITH ACUTE                                 | í      |
| U+00EE          | LATIN SMALL LETTER I WITH CIRCUMFLEX                            | î      |
| U+00EF          | LATIN SMALL LETTER I WITH DIAERESIS                             | ï      |
| U+00F0          | LATIN SMALL LETTER ETH                                          | ð      |
| U+00F1          | LATIN SMALL LETTER N WITH TILDE                                 | ñ      |
| U+00F2          | LATIN SMALL LETTER O WITH GRAVE                                 | ò      |
| U+00F3          | LATIN SMALL LETTER O WITH ACUTE                                 | ó      |
| U+00F4          | LATIN SMALL LETTER O WITH CIRCUMFLEX                            | ô      |
| U+00F5          | LATIN SMALL LETTER O WITH TILDE                                 | õ      |
| U+00F6          | LATIN SMALL LETTER O WITH DIAERESIS                             | ö      |
| U+00F8          | LATIN SMALL LETTER O WITH STROKE                                | ø      |
| U+00F9          | LATIN SMALL LETTER U WITH GRAVE                                 | ù      |
| U+00FA          | LATIN SMALL LETTER U WITH ACUTE                                 | ú      |
| U+00FB          | LATIN SMALL LETTER U WITH CIRCUMFLEX                            | û      |
| <U+00FB,U+0304> | LATIN SMALL LETTER U WITH CIRCUMFLEX WITH COMBINING MACRON      | û      |
| U+00FC          | LATIN SMALL LETTER U WITH DIAERESIS                             | ü      |
| U+00FD          | LATIN SMALL LETTER Y WITH ACUTE                                 | ý      |
| U+00FE          | LATIN SMALL LETTER THORN                                        | þ      |
| U+00FF          | LATIN SMALL LETTER Y WITH DIAERESIS                             | ÿ      |
| <U+00FF,U+0301> | LATIN SMALL LETTER Y WITH DIAERESIS WITH COMBINING ACUTE ACCENT | ÿ      |
| U+0100          | LATIN CAPITAL LETTER A WITH MACRON                              | Ā      |
| U+0101          | LATIN SMALL LETTER A WITH MACRON                                | ā      |
| U+0102          | LATIN CAPITAL LETTER A WITH BREVE                               | Ă      |

| Codepoints      | Name                                                               | Glyphe |
|-----------------|--------------------------------------------------------------------|--------|
| U+0103          | LATIN SMALL LETTER A WITH BREVE                                    | ă      |
| U+0104          | LATIN CAPITAL LETTER A WITH OGONEK                                 | Ą      |
| U+0105          | LATIN SMALL LETTER A WITH OGONEK                                   | ą      |
| U+0106          | LATIN CAPITAL LETTER C WITH ACUTE                                  | Ć      |
| U+0107          | LATIN SMALL LETTER C WITH ACUTE                                    | ć      |
| U+0108          | LATIN CAPITAL LETTER C WITH CIRCUMFLEX                             | Ĉ      |
| U+0109          | LATIN SMALL LETTER C WITH CIRCUMFLEX                               | ĉ      |
| U+010A          | LATIN CAPITAL LETTER C WITH DOT ABOVE                              | Ċ      |
| U+010B          | LATIN SMALL LETTER C WITH DOT ABOVE                                | ċ      |
| U+010C          | LATIN CAPITAL LETTER C WITH CARON                                  | Č      |
| <U+010C,U+0315> | LATIN CAPITAL LETTER C WITH CARON WITH COMBINING COMMA ABOVE RIGHT | Č́     |
| <U+010C,U+0323> | LATIN CAPITAL LETTER C WITH CARON WITH COMBINING DOT BELOW         | Č̣     |
| U+010D          | LATIN SMALL LETTER C WITH CARON                                    | č      |
| <U+010D,U+0315> | LATIN SMALL LETTER C WITH CARON WITH COMBINING COMMA ABOVE RIGHT   | č́     |
| <U+010D,U+0323> | LATIN SMALL LETTER C WITH CARON WITH COMBINING DOT BELOW           | č̣     |
| U+010E          | LATIN CAPITAL LETTER D WITH CARON                                  | Ď      |
| U+010F          | LATIN SMALL LETTER D WITH CARON                                    | ď      |
| U+0110          | LATIN CAPITAL LETTER D WITH STROKE                                 | Đ      |
| U+0111          | LATIN SMALL LETTER D WITH STROKE                                   | đ      |
| U+0112          | LATIN CAPITAL LETTER E WITH MACRON                                 | Ē      |
| U+0113          | LATIN SMALL LETTER E WITH MACRON                                   | ē      |
| U+0114          | LATIN CAPITAL LETTER E WITH BREVE                                  | Ĕ      |
| U+0115          | LATIN SMALL LETTER E WITH BREVE                                    | ĕ      |
| U+0116          | LATIN CAPITAL LETTER E WITH DOT ABOVE                              | Ė      |
| U+0117          | LATIN SMALL LETTER E WITH DOT ABOVE                                | ė      |
| U+0118          | LATIN CAPITAL LETTER E WITH OGONEK                                 | Ę      |
| U+0119          | LATIN SMALL LETTER E WITH OGONEK                                   | ę      |
| U+011A          | LATIN CAPITAL LETTER E WITH CARON                                  | Ě      |
| U+011B          | LATIN SMALL LETTER E WITH CARON                                    | ě      |
| U+011C          | LATIN CAPITAL LETTER G WITH CIRCUMFLEX                             | Ĝ      |
| U+011D          | LATIN SMALL LETTER G WITH CIRCUMFLEX                               | ĝ      |
| U+011E          | LATIN CAPITAL LETTER G WITH BREVE                                  | Ğ      |
| U+011F          | LATIN SMALL LETTER G WITH BREVE                                    | ğ      |
| U+0120          | LATIN CAPITAL LETTER G WITH DOT ABOVE                              | Ġ      |
| U+0121          | LATIN SMALL LETTER G WITH DOT ABOVE                                | ġ      |
| U+0122          | LATIN CAPITAL LETTER G WITH CEDILLA                                | Ģ      |
| U+0123          | LATIN SMALL LETTER G WITH CEDILLA                                  | ģ      |
| U+0124          | LATIN CAPITAL LETTER H WITH CIRCUMFLEX                             | Ĥ      |
| U+0125          | LATIN SMALL LETTER H WITH CIRCUMFLEX                               | ĥ      |
| U+0126          | LATIN CAPITAL LETTER H WITH STROKE                                 | Ħ      |
| U+0127          | LATIN SMALL LETTER H WITH STROKE                                   | ħ      |
| U+0128          | LATIN CAPITAL LETTER I WITH TILDE                                  | İ      |
| U+0129          | LATIN SMALL LETTER I WITH TILDE                                    | ı      |
| U+012A          | LATIN CAPITAL LETTER I WITH MACRON                                 | Ī      |
| <U+012A,U+0301> | LATIN CAPITAL LETTER I WITH MACRON WITH COMBINING ACUTE ACCENT     | Ī́     |

| Codepoints      | Name                                                         | Glyphe |
|-----------------|--------------------------------------------------------------|--------|
| U+012B          | LATIN SMALL LETTER I WITH MACRON                             | ī      |
| <U+012B,U+0301> | LATIN SMALL LETTER I WITH MACRON WITH COMBINING ACUTE ACCENT | í      |
| U+012C          | LATIN CAPITAL LETTER I WITH BREVE                            | Ĭ      |
| U+012D          | LATIN SMALL LETTER I WITH BREVE                              | ĭ      |
| U+012E          | LATIN CAPITAL LETTER I WITH OGONEK                           | Į      |
| U+012F          | LATIN SMALL LETTER I WITH OGONEK                             | į      |
| U+0130          | LATIN CAPITAL LETTER I WITH DOT ABOVE                        | İ      |
| U+0131          | LATIN SMALL LETTER DOTLESS I                                 | ı      |
| U+0132          | LATIN CAPITAL LIGATURE IJ                                    | IJ     |
| U+0133          | LATIN SMALL LIGATURE IJ                                      | ij     |
| U+0134          | LATIN CAPITAL LETTER J WITH CIRCUMFLEX                       | Ĵ      |
| U+0135          | LATIN SMALL LETTER J WITH CIRCUMFLEX                         | ĵ      |
| U+0136          | LATIN CAPITAL LETTER K WITH CEDILLA                          | Ƙ      |
| U+0137          | LATIN SMALL LETTER K WITH CEDILLA                            | ƙ      |
| U+0138          | LATIN SMALL LETTER KRA                                       | κ      |
| U+0139          | LATIN CAPITAL LETTER L WITH ACUTE                            | Ĺ      |
| U+013A          | LATIN SMALL LETTER L WITH ACUTE                              | ĺ      |
| U+013B          | LATIN CAPITAL LETTER L WITH CEDILLA                          | ƚ      |
| U+013C          | LATIN SMALL LETTER L WITH CEDILLA                            | Ɵ      |
| U+013D          | LATIN CAPITAL LETTER L WITH CARON                            | Ł      |
| U+013E          | LATIN SMALL LETTER L WITH CARON                              | ł      |
| U+013F          | LATIN CAPITAL LETTER L WITH MIDDLE DOT                       | Ł̣     |
| U+0140          | LATIN SMALL LETTER L WITH MIDDLE DOT                         | ł̣     |
| U+0141          | LATIN CAPITAL LETTER L WITH STROKE                           | Ł̈     |
| U+0142          | LATIN SMALL LETTER L WITH STROKE                             | ł̈     |
| U+0143          | LATIN CAPITAL LETTER N WITH ACUTE                            | Ń      |
| U+0144          | LATIN SMALL LETTER N WITH ACUTE                              | ń      |
| U+0145          | LATIN CAPITAL LETTER N WITH CEDILLA                          | Ɲ      |
| U+0146          | LATIN SMALL LETTER N WITH CEDILLA                            | ƞ      |
| U+0147          | LATIN CAPITAL LETTER N WITH CARON                            | Ň      |
| U+0148          | LATIN SMALL LETTER N WITH CARON                              | ň      |
| U+0149          | LATIN SMALL LETTER N PRECEDED BY APOSTROPHE                  | ’n     |
| U+014A          | LATIN CAPITAL LETTER ENG                                     | Ŋ      |
| U+014B          | LATIN SMALL LETTER ENG                                       | ŋ      |
| U+014C          | LATIN CAPITAL LETTER O WITH MACRON                           | Ō      |
| U+014D          | LATIN SMALL LETTER O WITH MACRON                             | ō      |
| U+014E          | LATIN CAPITAL LETTER O WITH BREVE                            | Ŏ      |
| U+014F          | LATIN SMALL LETTER O WITH BREVE                              | ö      |
| U+0150          | LATIN CAPITAL LETTER O WITH DOUBLE ACUTE                     | Ő      |
| U+0151          | LATIN SMALL LETTER O WITH DOUBLE ACUTE                       | ó      |
| U+0152          | LATIN CAPITAL LIGATURE OE                                    | Œ      |
| U+0153          | LATIN SMALL LIGATURE OE                                      | œ      |
| U+0154          | LATIN CAPITAL LETTER R WITH ACUTE                            | Ŕ      |
| U+0155          | LATIN SMALL LETTER R WITH ACUTE                              | ŗ      |
| U+0156          | LATIN CAPITAL LETTER R WITH CEDILLA                          | ƚ      |
| U+0157          | LATIN SMALL LETTER R WITH CEDILLA                            | Ɵ      |
| U+0158          | LATIN CAPITAL LETTER R WITH CARON                            | Ř      |
| U+0159          | LATIN SMALL LETTER R WITH CARON                              | ř      |

| Codepoints      | Name                                                         | Glyphe |
|-----------------|--------------------------------------------------------------|--------|
| U+015A          | LATIN CAPITAL LETTER S WITH ACUTE                            | Ś      |
| U+015B          | LATIN SMALL LETTER S WITH ACUTE                              | ś      |
| U+015C          | LATIN CAPITAL LETTER S WITH CIRCUMFLEX                       | Ŝ      |
| U+015D          | LATIN SMALL LETTER S WITH CIRCUMFLEX                         | ŝ      |
| U+015E          | LATIN CAPITAL LETTER S WITH CEDILLA                          | Ş      |
| U+015F          | LATIN SMALL LETTER S WITH CEDILLA                            | ş      |
| U+0160          | LATIN CAPITAL LETTER S WITH CARON                            | Š      |
| U+0161          | LATIN SMALL LETTER S WITH CARON                              | š      |
| U+0162          | LATIN CAPITAL LETTER T WITH CEDILLA                          | Ť      |
| U+0163          | LATIN SMALL LETTER T WITH CEDILLA                            | ť      |
| U+0164          | LATIN CAPITAL LETTER T WITH CARON                            | Ț      |
| U+0165          | LATIN SMALL LETTER T WITH CARON                              | ț      |
| U+0166          | LATIN CAPITAL LETTER T WITH STROKE                           | Ƨ      |
| U+0167          | LATIN SMALL LETTER T WITH STROKE                             | Ƨ      |
| U+0168          | LATIN CAPITAL LETTER U WITH TILDE                            | Ŭ      |
| U+0169          | LATIN SMALL LETTER U WITH TILDE                              | ŭ      |
| U+016A          | LATIN CAPITAL LETTER U WITH MACRON                           | Ū      |
| U+016B          | LATIN SMALL LETTER U WITH MACRON                             | ū      |
| U+016C          | LATIN CAPITAL LETTER U WITH BREVE                            | Ů      |
| U+016D          | LATIN SMALL LETTER U WITH BREVE                              | ů      |
| U+016E          | LATIN CAPITAL LETTER U WITH RING ABOVE                       | Ű      |
| U+016F          | LATIN SMALL LETTER U WITH RING ABOVE                         | ű      |
| U+0170          | LATIN CAPITAL LETTER U WITH DOUBLE ACUTE                     | Ú      |
| U+0171          | LATIN SMALL LETTER U WITH DOUBLE ACUTE                       | ú      |
| U+0172          | LATIN CAPITAL LETTER U WITH OGONEK                           | Ų      |
| U+0173          | LATIN SMALL LETTER U WITH OGONEK                             | ų      |
| U+0174          | LATIN CAPITAL LETTER W WITH CIRCUMFLEX                       | Ŵ      |
| U+0175          | LATIN SMALL LETTER W WITH CIRCUMFLEX                         | ŵ      |
| U+0176          | LATIN CAPITAL LETTER Y WITH CIRCUMFLEX                       | Ŷ      |
| U+0177          | LATIN SMALL LETTER Y WITH CIRCUMFLEX                         | ŷ      |
| U+0178          | LATIN CAPITAL LETTER Y WITH DIAERESIS                        | Ÿ      |
| U+0179          | LATIN CAPITAL LETTER Z WITH ACUTE                            | Ź      |
| U+017A          | LATIN SMALL LETTER Z WITH ACUTE                              | ź      |
| U+017B          | LATIN CAPITAL LETTER Z WITH DOT ABOVE                        | Ž      |
| U+017C          | LATIN SMALL LETTER Z WITH DOT ABOVE                          | ž      |
| U+017D          | LATIN CAPITAL LETTER Z WITH CARON                            | Ž      |
| <U+017D,U+0326> | LATIN CAPITAL LETTER Z WITH CARON WITH COMBINING COMMA BELOW | Ẓ̌     |
| <U+017D,U+0327> | LATIN CAPITAL LETTER Z WITH CARON WITH COMBINING CEDILLA     | Ẓ̌̃    |
| U+017E          | LATIN SMALL LETTER Z WITH CARON                              | ž      |
| <U+017E,U+0326> | LATIN SMALL LETTER Z WITH CARON WITH COMBINING COMMA BELOW   | ẓ̌     |
| <U+017E,U+0327> | LATIN SMALL LETTER Z WITH CARON WITH COMBINING CEDILLA       | ẓ̌̃    |
| U+0187          | LATIN CAPITAL LETTER C WITH HOOK                             | Č      |
| U+0188          | LATIN SMALL LETTER C WITH HOOK                               | č      |
| U+018F          | LATIN CAPITAL LETTER SCHWA                                   | Ə      |
| U+0197          | LATIN CAPITAL LETTER I WITH STROKE                           | Ǝ      |