

DIN 91379



ICS 35.040.10

Ersatz für
DIN SPEC 91379:2019-03**Zeichen und definierte Zeichensequenzen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa, mit CD-ROM**

Characters and defined character sequences in Unicode for the electronic processing of names and data exchange in Europe, with CD-ROM

Caractères et séquences de caractères définis en Unicode pour le traitement électronique des noms et l'échange de données en Europe, avec CD-ROM

Gesamtumfang 80 Seiten

DIN-Normenausschuss Informationstechnik und Anwendungen (NIA)

Inhalt

	Seite
Vorwort	4
Einleitung	6
1 Anwendungsbereich	10
2 Normative Verweisungen	10
3 Begriffe	11
4 Konformität von IT-Verfahren	12
5 Normative Schriftzeichen	13
5.1 Lateinische Buchstaben	13
5.2 Nicht-Buchstaben N1	31
5.3 Nicht-Buchstaben N2	32
5.4 Nicht-Buchstaben N3	34
5.5 Nicht-Buchstaben N4	34
5.6 Veraltete Schriftzeichen	35
6 Normative Abbildung lateinischer Buchstaben auf Grundbuchstaben (Suchform)	36
Anhang A (informativ) Erweiterte Schriftzeichen	61
A.1 Allgemeines	61
A.2 Griechische Buchstaben	61
A.3 Kyrillische Buchstaben	63
A.4 Nicht-Buchstaben E1	65
Anhang B (informativ) Basiszeichen und diakritische Zeichen für lateinische Buchstaben	67
Anhang C (informativ) Technische Datentypen	71
C.1 Allgemeines	71
C.2 Datentyp A	73
C.3 Datentyp B	73
C.4 Datentyp C	73
C.5 Datentyp D	73
C.6 Datentyp E	73
Anhang D (informativ) Strukturierte und tabellarische Übersicht aller Schriftzeichen	75
Anhang E (informativ) Abbildung auf Legacy Codepages	77
Anhang F (informativ) Abbildung von Legacy Codepages	78
Literaturhinweise	79

Bilder

Bild C.1 — Datentypen der Norm DIN 91379	72
Bild D.1 — Struktur der Datenbasis aller Schriftzeichen dieses Dokuments	75

Tabellen

Tabelle 1 — Schriftzeichengruppen	8
Tabelle 2 — Angehängte Dateien	9
Tabelle 3 — Lateinische Buchstaben	13
Tabelle 4 — Nicht-Buchstaben N1	32
Tabelle 5 — Nicht-Buchstaben N2	32
Tabelle 6 — Nicht-Buchstaben N3	34
Tabelle 7 — Nicht-Buchstaben N4	35
Tabelle 8 — Schriftzeichen, von deren Verwendung abgeraten wird	35

Tabelle 9 — Abbildung lateinischer Buchstaben auf Grundbuchstaben analog ICAO	37
Tabelle A.1 — Griechische Buchstaben	61
Tabelle A.2 — Kyrillische Buchstaben	63
Tabelle A.3 — Nicht-Buchstaben E1	65
Tabelle B.1 — Übersicht über Basiszeichen in der Schriftzeichengruppe lateinischer Zeichen . .	67
Tabelle B.2 — Kombinierende diakritische Zeichen	70
Tabelle C.1 — Übersicht über die empfohlenen Datentypen und zugehörigen Schriftzeichengruppen	72
Tabelle F.1 — Abbildung von Legacy-Zeichen außerhalb dieses Dokuments in ihren Zeichenvorrat	78

Vorwort

Dieses Dokument wurde vom Arbeitsausschuss NA 043-03-03-04 AK „Elektronische Verarbeitung von Namen im DIN Normenausschuss Informationstechnik und Anwendungen (NIA) erarbeitet.

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.

Aktuelle Informationen zu diesem Dokument können über die Internetseiten von DIN (www.din.de) durch eine Suche nach der Dokumentennummer aufgerufen werden.

Änderungen

Gegenüber DIN SPEC 91379:2019-03 wurden folgende Änderungen vorgenommen:

- a) Bezug zum Unicode-Standard Version 13 anstatt Version 9;
- b) Begriff „Schriftzeichen“ im Titel geändert in „Zeichen und definierte Zeichensequenzen“;
- c) neue Absätze mit Zwischenüberschrift im Abschnitt Einleitung hinzugefügt: „Unicode Normalform C in diesem Dokument“;
- d) Abschnitt 2 „Normative Verweisungen“ aktualisiert;
- e) neue Begriffe in Abschnitt 3 „Begriffe“ aufgenommen;
- f) Begriffsdefinition für „Kombinierendes diakritisches Zeichen“ entsprechend Aussagen aus dem Unicode-Standard erweitert;
- g) normative Aussagen zur Konformität von IT-Verfahren in den neuen Abschnitt 4 verschoben und ergänzt um Aussagen zu Normalform C (NFC) und UTF-8;
- h) drei normative Zeichen im Abschnitt 5.1 „Lateinische Buchstaben“ (für alle Namen) neu aufgenommen:
 - LATIN SMALL LETTER E WITH MACRON WITH COMBINING VERTICAL LINE ABOVE (U+0113,U+030D);
 - LATIN SMALL LETTER O WITH MACRON WITH COMBINING VERTICAL LINE ABOVE (U+014D,U+030D);
 - LATIN SMALL LETTER E WITH MACRON AND ACUTE (U+1E17);
- i) Abschnitt 5.6 (normativ) „Veraltete Schriftzeichen“ neu aufgenommen;
- j) neue Kennzeichnung der Buchstaben in Tabelle 9 „Abbildung lateinischer Buchstaben auf Grundbuchstaben analog ICAO“, zu denen ICAO keine Aussage trifft, siehe Abschnitt 6 (normativ);
- k) zwei informative Zeichen im Anhang A.4 „Nicht-Buchstaben E1“ (vor allem für Namen juristischer Personen und für Produktnamen) neu aufgenommen:
 - PRIME (U+2032);
 - DOUBLE PRIME (U+2033);

- l) Anhang B (informativ) „Basiszeichen und diakritische Zeichen für lateinische Buchstaben“ neu aufgenommen;
- m) die meisten digitalen Anhänge umbenannt – Versionsangaben befinden sich nun innerhalb der Datei, nicht mehr im Dateinamen;
- n) Änderungen am digitalen Anhang D mit der tabellarischen Übersicht aller Zeichen:
 - im Dateiformat .ods ausgeliefert anstatt .xlsx;
 - deutschsprachige Spaltenüberschriften verwendet;
 - neue Spalten „NFD-Basis“ und „NFD-Diakritika“;
- o) Änderungen am digitalen Anhang E mit der Abbildung auf Legacy Codepages:
 - im Dateiformat .ods ausgeliefert anstatt .xlsx;
 - erweitert um die Legacy Codepage String.Latin 1.1.1;
 - das Mapping von vier Zeichen, die auf dem Buchstaben Ezh basieren, entsprechend der normativen Abbildung aus Abschnitt 6 von „Y“ zu „Z“ korrigiert;
 - Mappings für griechische und kyrillische Buchstaben entfernt, da sie mittels aktuell gültiger ISO-Transliterationen in die lateinische Schrift überführt werden müssen;
- p) Anhang F (informativ) „Abbildung von Legacy Codepages“ neu aufgenommen;
- q) Literaturhinweise aktualisiert;
- r) redaktionelle Anpassungen vorgenommen.

Frühere Ausgaben

DIN SPEC 91379:2019-03

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 Unterstützung 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 dieses Dokuments 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 sollen.

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 C empfohlenen Datentypen dienen der Unterstützung dieses Ziels.

Zudem entstehen Sicherheitsrisiken, wenn beliebige Kombinationen von Basiszeichen mit diakritischen Zeichen zulässig sind. Dies ermöglicht optisch identische Schriftzeichen, die informationstechnisch unterschiedlich sind. Für eine detaillierte Darstellung wird auf den Unicode Technical Report #36 „UNICODE SECURITY CONSIDERATIONS“ [10] verwiesen. Durch dieses Dokument lassen sich diese Risiken erheblich reduzieren. Denn sie kann in fachlich begründeten Fällen für die Einschränkung genutzt werden, ausschließlich die darin genannten Schriftzeichen zu verarbeiten bzw. zu übermitteln. Die Durchsetzung einer entsprechenden Einschränkung kann sowohl rechtlich als auch technisch erfolgen, beispielsweise durch Nutzung der im Anhang C genannten Datentypen.

Die genannten Gründe führen zu dem Ergebnis, dass eine abschließende Auflistung der zu unterstützenden Schriftzeichen als eine echte Teilmenge der in Unicode enthaltenen Zeichen und Zeichensequenzen in diesem Dokument notwendig ist.

Überblick

Im Sinne des Investitionsschutzes berücksichtigt dieses Dokument angemessen die vom IT-Planungsrat und der Innenministerkonferenz derzeit bereits für den gleichen Anwendungsbereich verbindlich vorgegebene DIN SPEC 91379:2019-03 „Zeichen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa“ (auch „String.Latin+ 1.2“ genannt).

Die Legacy-Zeichensätze ISO/IEC 8859-1 [3], ISO/IEC 8859-15 [4] und Windows 1252 waren weitere Grundlagen der Betrachtungen bei der Zusammenstellung der Schriftzeichen dieses Dokuments. Allerdings wurden einige wenige Zeichen aufgrund einer fachlichen Bewertung nicht aufgenommen.

Dieses Dokument beschreibt in Abschnitt 4 die Anforderungen an zu diesem Dokument konforme Implementationen.

Es legt in Abschnitt 5 diejenigen Schriftzeichen fest, die von allen zu diesem Dokument konformen IT-Verfahren zumindest bei allen für Namen (im weiteren Sinne – siehe Abschnitt 1 „Anwendungsbereich“) bestimmten Datenfeldern unterstützt werden müssen.

Es empfiehlt in Anhang A diejenigen Schriftzeichen, die darüber hinaus von allen zu diesem Dokument konformen IT-Verfahren unterstützt werden sollten.

Es legt in Abschnitt 6 fest, wie die in Abschnitt 5 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.

In dem informativen Anhang B sind die von Unicode festgelegten und von diesem Dokument verwendeten *Basiszeichen für lateinische Buchstaben* aufgeführt, die zusammen mit den ebenfalls in diesem Anhang genannten *kombinierenden diakritischen Zeichen* für die lateinischen Buchstaben der Tabelle 3 erforderlich sind. Es handelt sich um deutlich mehr Basiszeichen als a bis z und ß aus der deutschen Sprache.

Es empfiehlt in Anhang C Datentypen für Einschränkungen auf bestimmte Teilmengen der in diesem Dokument 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
Lateinische Buchstaben	3	649	149
Nicht-Buchstaben N1	4	18	0
Nicht-Buchstaben N2	5	60	0
Nicht-Buchstaben N3	6	6	0
Nicht-Buchstaben N4	7	4	0
Griechische Buchstaben	A.1	69	0
Kyrillische Buchstaben	A.2	62	0
Nicht-Buchstaben E1	A.3	42	0
Kombinierende diakritische Zeichen	B.2	20	0
Summe		930	149

„Lateinische Schriftzeichen“ im Sinne dieses Dokuments sind alle in Tabelle 3 bis Tabelle 7 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 werden 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 character sequence“ nach Unicode Standard Annex #34 „UNICODE NAMED CHARACTER SEQUENCES“ [8] 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 Schriftzeichen wird eine Glyphen angegeben, die an der für die ISO/IEC 10646 gewählten Darstellung orientiert ist. Für jedes nicht darstellbare Schriftzeichen wird ersatzweise die Glyphen „“ (REPLA-

CEMENT CHARACTER, U+FFFD) verwendet. Die Darstellung der Schriftzeichen durch Glyphen dient nur der Illustration. Für Implementationen dieses Dokuments muss die Angabe der Codepoints herangezogen werden.

Unicode Normalform C in diesem Dokument

Viele abstrakte Zeichen des Unicode-Standards können auf unterschiedliche Arten repräsentiert werden. Um gleichwohl die Äquivalenz unterschiedlicher Repräsentationen feststellen zu können, sind im Unicode Standard Annex #15 [7] vier Normalformen festgelegt. In diesem Dokument und den informativ beigefügten Materialien wird für die Repräsentation von Schriftzeichen stets die *Unicode Normalform C (NFC)* verwendet.

- Alle in den Tabellen enthaltenen Schriftzeichen werden stets in NFC repräsentiert.
- In der informativ beigefügten Datei chars.xml, die alle in diesem Dokument enthaltenen Schriftzeichen in strukturierter Form enthält, sind alle Schriftzeichen in NFC repräsentiert.
- Die in Anhang C beschriebenen Datentypen dienen der Prüfung, ob ein Text ausschließlich aus Schriftzeichen besteht, die vom jeweiligen Datentypen umfasst sind. Die informell beigefügte technische Umsetzung dieser Datentypen in Form eines XML Schema in der Datei din-91379-datatypes.xsd setzt voraus, dass der zu prüfende Text in NFC vorliegt.

Darüber hinaus ist die Verwendung von NFC bei Datenübermittlungen zwischen IT-Verfahren Gegenstand der Konformitätsanforderungen in Abschnitt 4.

Informativ beigefügte Dateien

Dieses Dokument wird ergänzt durch die nachfolgend in Tabelle 2 genannten Dateien, die informativ beigefügt werden:

Tabelle 2 — Angehängte Dateien

Datei	Beschreibung
din-91379-datatypes.xsd	eine technische Umsetzung der in Anhang B beschriebenen Datentypen
chars.xml	eine XML-Datei mit allen in diesem Dokument genannten Schriftzeichen und Schriftzeichengruppen
chars.xsd	ein XML Schema 1.1 mit der Festlegung der Struktur von chars.xml
din-91379.ods	eine tabellarische Übersicht über alle in diesem Dokument genannten Schriftzeichen
legacy-mappings.ods	eine Abbildung der in diesem Dokument enthaltenen Schriftzeichen auf Legacy Codepages

1 Anwendungsbereich

Dieses Dokument DIN 91379 „Zeichen und definierte Zeichensequenzen in Unicode für die elektronische Verarbeitung von Namen und den Datenaustausch in Europa“ legt die Schriftzeichen (siehe 3.5) fest, die für Namen im weiteren Sinne benötigt werden. Namen im weiteren Sinne sind bürgerliche und geführte Namen natürlicher Personen nach den 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.

Dieses Dokument teilt sich auf in einen normativen Teil und einen informativen Teil.

Der normative Teil dieses Dokuments legt diejenige Teilmenge der in Unicode enthaltenen Zeichen (siehe 3.3) und Zeichensequenzen (siehe 3.4) 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 diesem Dokument 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 informativen Teil dieses Dokuments werden darüber hinaus ergänzende Schriftzeichen aus Unicode festgelegt, die gegebenenfalls 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. Zu diesem Dokument konforme IT-Verfahren dürfen darüber hinausgehende Schriftzeichen unterstützen. In Schnittstellenvereinbarungen ist es jedoch erforderlich, den jeweiligen Zeichenvorrat abschließend festzulegen. Dieses Dokuments stellt in Anhang C Datentypen für fachlich bestimmte Anwendungsfälle zur Verfügung. Dazu werden bestimmte Schriftzeichengruppen (siehe 3.6) 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.

Eine strukturierte und tabellarische Übersicht aller Schriftzeichen findet sich in Anhang D.

Eine geeignete Abbildung von in diesem Dokument dargestellten Schriftzeichen auf gegebenenfalls noch verwendete Legacy Codepages ist in Anhang E dargestellt.

Eine geeignete Abbildung von nicht in diesem Dokument enthaltenen Schriftzeichen aus 8-Bit-Zeichensätzen auf den Zeichenvorrat dieses Dokuments ist in Anhang F dargestellt.

Dieses Dokument 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.

Dieses Dokument trifft weder Aussagen zum Umgang mit historischen Schriftzeichen noch zum Umgang mit Fließtext. Dieses Dokument regelt nicht die Darstellung von Schriftzeichen (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)*

International Civil Aviation Organization (ICAO): *Doc 9303 — Machine Readable Travel Documents, Part 3: Specifications common to all MRTDs, 7th Edition 2015*

3 Begriffe

Die für dieses Dokument 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: verfügbar unter <https://www.din.de/go/din-term>
- DKE-IEV: verfügbar 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 13.0.0, 3.4, D7]

3.2

Codepoint

(en: **code point**)

jeder 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-Norm 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 13.0.0, 3.4, D10 in Verbindung mit D9, modifiziert – Anmerkung 1 und Anmerkung 2 hinzugefügt.]

3.3

kodiertes Zeichen

Zeichen

(en: **encoded character, coded character**
character)

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

Anmerkung 1 zum Begriff: Der Begriff „Zeichen“ ist alleinstehend als Abkürzung für den Begriff „kodiertes Zeichen“ zu verstehen. Siehe hierfür die Erläuterung in Unicode 13.0.0, 3.4, zwischen den Definitionen D12 und D13.

[QUELLE: The Unicode Standard, Version 13.0.0, 3.4, D11, modifiziert – Anmerkung 1 hinzugefügt.]

3.4

kodierte Zeichensequenz

Zeichensequenz

(en: **coded character sequence**
character sequence)

geordnete Sequenz eines oder mehrerer Codepoints

Anmerkung 1 zum Begriff: Der Begriff „Zeichensequenz“ ist alleinstehend als Abkürzung für den Begriff „kodierte Zeichensequenz“ zu verstehen. Siehe hierfür die Erläuterung in Unicode 13.0.0, 3.4, zwischen den Definitionen D12 und D13.

[QUELLE: The Unicode Standard, Version 13.0.0, 3.4, D12, modifiziert – Anmerkung 1 hinzugefügt.]

3.5

Schriftzeichen

kleinste im Schreibfluss aufeinanderfolgende Einheit einer Schrift, z. B. des lateinischen oder kyrillischen Alphabets, die sprachunabhängig ist, da sie für verschiedene Sprachen genutzt werden kann, 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 dieses Dokuments 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 Zeichen und Zeichensequenzen.

Anmerkung 4 zum Begriff: In diesem Dokument wird der Begriff „Schriftzeichen“ weitgehend synonym zum Begriff „Graphem“ verwendet. Es werden aber auch „Plurigraphie“ miteingeschlossen.

3.6

Schriftzeichengruppe

Zusammenfassung von Schriftzeichen unter fachlichen Aspekten

Anmerkung 1 zum Begriff: Jedes der in DIN 91379 genannten Schriftzeichen ist genau einer Schriftzeichengruppe zugeordnet. Jeder Schriftzeichengruppe ist mindestens ein Schriftzeichen zugeordnet. Jede Schriftzeichengruppe hat einen innerhalb von DIN 91379 eindeutigen Namen, z. B. „Lateinische Buchstaben“ oder „Nicht-Buchstaben N1“.

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

3.7

kombinierendes diakritisches Zeichen

(en: **combining diacritical mark**)

Zeichen, welches dem Unicode Block „Combining Diacritical Marks“ zugeordnet ist

Anmerkung 1 zum Begriff: Kombinierende Zeichen sind eine spezielle Klasse von Zeichen im Unicode-Standard, die dazu bestimmt sind, mit einem vorangehenden Zeichen, ihrem Basiszeichen, kombiniert zu werden. Sie haben eine formale syntaktische Beziehung – oder Abhängigkeit – zu ihrem Basiszeichen. Diese Beziehung ist relevant für die Definition von kombinierenden Zeichensequenzen, die kanonische Umordnung und den Unicode-Normalisierungsalgorithmus.

Anmerkung 2 zum Begriff: Kombinierende diakritische Zeichen werden auf Buchstaben der lateinischen Schrift (oder ähnlichen Schriften) angewendet. Sie zeigen in der Regel eine Veränderung in der Aussprache oder einen bestimmten Ton oder eine Betonung an. Sie können auch verwendet werden, um neue Buchstaben abzuleiten.

Anmerkung 3 zum Begriff: Tabelle B.2 nennt alle diakritischen Zeichen, die erforderlich sind, um in Kombination mit den in Tabelle B.1 genannten Basiszeichen die normativ vorgegebenen Schriftzeichen der Schriftzeichengruppe „Lateinische Buchstaben“ bilden zu können.

4 Konformität von IT-Verfahren

Jedes konforme IT-Verfahren muss die in Tabelle 3 bis Tabelle 7 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. Für deren Übermittlung muss die Normalform C (NFC) verwendet werden.

Jedes konforme IT-Verfahren muss für Übermittlungen die Kodierung UTF-8 unterstützen, die unter anderem identisch in ISO/IEC 10646 definiert ist.

5 Normative Schriftzeichen

5.1 Lateinische Buchstaben

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

Tabelle 3 — 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	C'
<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	Ĉ
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	Ĥ
<U+0046,U+0304>	LATIN CAPITAL LETTER F WITH COMBINING MACRON	FŌ
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	HŌ
<U+0048,U+0326>	LATIN CAPITAL LETTER H WITH COMBINING COMMA BELOW	Ḥ
<U+0048,U+0331>	LATIN CAPITAL LETTER H WITH COMBINING MACRON BELOW	H̄
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

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
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	Ĳ
<U+004B,U+0304>	LATIN CAPITAL LETTER K WITH COMBINING MACRON	KŌ
<U+004B,U+0307>	LATIN CAPITAL LETTER K WITH COMBINING DOT ABOVE	Ķ̣
<U+004B,U+0315>	LATIN CAPITAL LETTER K WITH COMBINING COMMA ABOVE RIGHT	Kʼ
<U+004B,U+031B>	LATIN CAPITAL LETTER K WITH COMBINING HORN	Ɔ
<U+004B,U+0326>	LATIN CAPITAL LETTER K WITH COMBINING COMMA BELOW	Ɔ̣
<U+004B, U+035F, U+0048>	LATIN CAPITAL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN CAPITAL LETTER H	Ɔ̈́
<U+004B, U+035F, U+0068>	LATIN CAPITAL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN SMALL LETTER H	Ɔ̈́
U+004C	LATIN CAPITAL LETTER L	L
<U+004C,U+0302>	LATIN CAPITAL LETTER L WITH COMBINING CIRCUMFLEX ACCENT	Ĺ
<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	Ł̄
<U+004C,U+0326>	LATIN CAPITAL LETTER L WITH COMBINING COMMA BELOW	Ł̣
U+004D	LATIN CAPITAL LETTER M	M
<U+004D,U+0300>	LATIN CAPITAL LETTER M WITH COMBINING GRAVE ACCENT	Ẁ
<U+004D,U+0302>	LATIN CAPITAL LETTER M WITH COMBINING CIRCUMFLEX ACCENT	Ẃ
<U+004D,U+0306>	LATIN CAPITAL LETTER M WITH COMBINING BREVE	Ṁ
<U+004D,U+0310>	LATIN CAPITAL LETTER M WITH COMBINING CANDRABINDU	Ṃ
U+004E	LATIN CAPITAL LETTER N	N
<U+004E,U+0302>	LATIN CAPITAL LETTER N WITH COMBINING CIRCUMFLEX ACCENT	Ñ
<U+004E,U+0304>	LATIN CAPITAL LETTER N WITH COMBINING MACRON	NŌ
<U+004E,U+0306>	LATIN CAPITAL LETTER N WITH COMBINING BREVE	Ṇ
<U+004E,U+0326>	LATIN CAPITAL LETTER N WITH COMBINING COMMA BELOW	Ṇ
U+004F	LATIN CAPITAL LETTER O	O
U+0050	LATIN CAPITAL LETTER P	P
<U+0050,U+0300>	LATIN CAPITAL LETTER P WITH COMBINING GRAVE ACCENT	Ṗ
<U+0050,U+0304>	LATIN CAPITAL LETTER P WITH COMBINING MACRON	PŌ
<U+0050,U+0315>	LATIN CAPITAL LETTER P WITH COMBINING COMMA ABOVE RIGHT	Pʼ
<U+0050,U+0323>	LATIN CAPITAL LETTER P WITH COMBINING DOT BELOW	Ṕ
U+0051	LATIN CAPITAL LETTER Q	Q

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
U+0052	LATIN CAPITAL LETTER R	R
<U+0052,U+0306>	LATIN CAPITAL LETTER R WITH COMBINING BREVE	Ř
<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	Ŝ
<U+0053,U+0304>	LATIN CAPITAL LETTER S WITH COMBINING MACRON	SŌ
<U+0053, U+031B, U+0304>	LATIN CAPITAL LETTER S WITH COMBINING HORN AND COMBINING MACRON	Š̄
<U+0053,U+0331>	LATIN CAPITAL LETTER S WITH COMBINING MACRON BELOW	Ṣ
U+0054	LATIN CAPITAL LETTER T	T
<U+0054,U+0300>	LATIN CAPITAL LETTER T WITH COMBINING GRAVE ACCENT	Ť
<U+0054,U+0304>	LATIN CAPITAL LETTER T WITH COMBINING MACRON	TŌ
<U+0054,U+0308>	LATIN CAPITAL LETTER T WITH COMBINING DIAERESIS	Ț
<U+0054,U+0315>	LATIN CAPITAL LETTER T WITH COMBINING COMMA ABOVE RIGHT	Tʼ
<U+0054,U+031B>	LATIN CAPITAL LETTER T WITH COMBINING HORN	Ṯ
U+0055	LATIN CAPITAL LETTER U	U
<U+0055,U+0307>	LATIN CAPITAL LETTER U WITH COMBINING DOT ABOVE	Ů
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	Ẑ
<U+005A,U+0304>	LATIN CAPITAL LETTER Z WITH COMBINING MACRON	ZŌ
<U+005A,U+0306>	LATIN CAPITAL LETTER Z WITH COMBINING BREVE	Ž
<U+005A,U+0308>	LATIN CAPITAL LETTER Z WITH COMBINING DIAERESIS	Ẓ
<U+005A,U+0327>	LATIN CAPITAL LETTER Z WITH COMBINING CEDILLA	Ẓ̣
U+0061	LATIN SMALL LETTER A	a
<U+0061,U+030B>	LATIN SMALL LETTER A WITH COMBINING DOUBLE ACUTE ACCENT	Ǻ
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	cŌ
<U+0063,U+0306>	LATIN SMALL LETTER C WITH COMBINING BREVE	č
<U+0063,U+0308>	LATIN SMALL LETTER C WITH COMBINING DIAERESIS	č̈

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
<U+0063,U+0315>	LATIN SMALL LETTER C WITH COMBINING COMMA ABOVE RIGHT	c'
<U+0063,U+0323>	LATIN SMALL LETTER C WITH COMBINING DOT BELOW	ç
<U+0063,U+0326>	LATIN SMALL LETTER C WITH COMBINING COMMA BELOW	ç
<U+0063, U+0328, U+0306>	LATIN SMALL LETTER C WITH COMBINING OGONEK AND COMBINING BREVE	č
U+0064	LATIN SMALL LETTER D	d
<U+0064,U+0302>	LATIN SMALL LETTER D WITH COMBINING CIRCUMFLEX ACCENT	ď
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	f̄
U+0067	LATIN SMALL LETTER G	g
<U+0067,U+0300>	LATIN SMALL LETTER G WITH COMBINING GRAVE ACCENT	ǵ
U+0068	LATIN SMALL LETTER H	h
<U+0068,U+0304>	LATIN SMALL LETTER H WITH COMBINING MACRON	h̄
<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	í
U+006B	LATIN SMALL LETTER K	k
<U+006B,U+0300>	LATIN SMALL LETTER K WITH COMBINING GRAVE ACCENT	ƀ
<U+006B,U+0302>	LATIN SMALL LETTER K WITH COMBINING CIRCUMFLEX ACCENT	ķ
<U+006B,U+0304>	LATIN SMALL LETTER K WITH COMBINING MACRON	k̄
<U+006B,U+0307>	LATIN SMALL LETTER K WITH COMBINING DOT ABOVE	ķ
<U+006B,U+0315>	LATIN SMALL LETTER K WITH COMBINING COMMA ABOVE RIGHT	k'
<U+006B,U+031B>	LATIN SMALL LETTER K WITH COMBINING HORN	ƙ
<U+006B,U+0326>	LATIN SMALL LETTER K WITH COMBINING COMMA BELOW	ķ
<U+006B, U+035F, U+0068>	LATIN SMALL LETTER K WITH COMBINING DOUBLE MACRON BELOW AND LATIN SMALL LETTER H	kh̄
U+006C	LATIN SMALL LETTER L	l
<U+006C,U+0302>	LATIN SMALL LETTER L WITH COMBINING CIRCUMFLEX ACCENT	ĺ
<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	ł̸
U+006D	LATIN SMALL LETTER M	m

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
<U+006D,U+0300>	LATIN SMALL LETTER M WITH COMBINING GRAVE ACCENT	m̀
<U+006D,U+0302>	LATIN SMALL LETTER M WITH COMBINING CIRCUMFLEX ACCENT	m̂
<U+006D,U+0306>	LATIN SMALL LETTER M WITH COMBINING BREVE	m̃
<U+006D,U+0310>	LATIN SMALL LETTER M WITH COMBINING CANDRABINDU	m̐
U+006E	LATIN SMALL LETTER N	n
<U+006E,U+0302>	LATIN SMALL LETTER N WITH COMBINING CIRCUMFLEX ACCENT	n̂
<U+006E,U+0304>	LATIN SMALL LETTER N WITH COMBINING MACRON	n̄
<U+006E,U+0306>	LATIN SMALL LETTER N WITH COMBINING BREVE	ñ
<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	p̀
<U+0070,U+0304>	LATIN SMALL LETTER P WITH COMBINING MACRON	p̄
<U+0070,U+0315>	LATIN SMALL LETTER P WITH COMBINING COMMA ABOVE RIGHT	p’
<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	r̃
<U+0072,U+0325>	LATIN SMALL LETTER R WITH COMBINING RING BELOW	r̥
<U+0072, U+0325, U+0304>	LATIN SMALL LETTER R WITH COMBINING RING BELOW AND COMBINING MACRON	r̥̄
U+0073	LATIN SMALL LETTER S	s
<U+0073,U+0300>	LATIN SMALL LETTER S WITH COMBINING GRAVE ACCENT	s̀
<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	s̄̐
<U+0073,U+0331>	LATIN SMALL LETTER S WITH COMBINING MACRON BELOW	s̸̄
U+0074	LATIN SMALL LETTER T	t
<U+0074,U+0300>	LATIN SMALL LETTER T WITH COMBINING GRAVE ACCENT	t̀
<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	t’
<U+0074,U+031B>	LATIN SMALL LETTER T WITH COMBINING HORN	t̐
U+0075	LATIN SMALL LETTER U	u
<U+0075,U+0307>	LATIN SMALL LETTER U WITH COMBINING DOT ABOVE	ụ
U+0076	LATIN SMALL LETTER V	v
U+0077	LATIN SMALL LETTER W	w

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
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	ẓ
<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	z̧
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	Ç̆
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
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

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
<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	ŭ
U+00FC	LATIN SMALL LETTER U WITH DIAERESIS	ü
U+00FD	LATIN SMALL LETTER Y WITH ACUTE	ý

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
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	ÿ̇
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
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	Č
U+010B	LATIN SMALL LETTER C WITH DOT ABOVE	č
U+010C	LATIN CAPITAL LETTER C WITH CARON	Č
<U+010C,U+0315>	LATIN CAPITAL LETTER C WITH CARON WITH COMBINING COMMA ABOVE RIGHT	Č̸
<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	č̸
<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	d'
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+0113,U+030D>	LATIN SMALL LETTER E WITH MACRON WITH COMBINING VERTICAL LINE ABOVE	ē̅
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	Ê
U+0117	LATIN SMALL LETTER E WITH DOT ABOVE	ê
U+0118	LATIN CAPITAL LETTER E WITH OGONEK	Ę
U+0119	LATIN SMALL LETTER E WITH OGONEK	ę

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
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	Ĩ
U+0129	LATIN SMALL LETTER I WITH TILDE	ĩ
U+012A	LATIN CAPITAL LETTER I WITH MACRON	Ī
<U+012A,U+0301>	LATIN CAPITAL LETTER I WITH MACRON WITH COMBINING ACUTE ACCENT	Ī
U+012B	LATIN SMALL LETTER I WITH MACRON	ī
<U+012B,U+0301>	LATIN SMALL LETTER I WITH MACRON WITH COMBINING ACUTE ACCENT	ī
U+012C	LATIN CAPITAL LETTER I WITH BREVE	İ
U+012D	LATIN SMALL LETTER I WITH BREVE	ı
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	Ķ
U+0137	LATIN SMALL LETTER K WITH CEDILLA	ķ
U+0138	LATIN SMALL LETTER KRA	κ
U+0139	LATIN CAPITAL LETTER L WITH ACUTE	Ł
U+013A	LATIN SMALL LETTER L WITH ACUTE	ł
U+013B	LATIN CAPITAL LETTER L WITH CEDILLA	Ľ

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
U+013C	LATIN SMALL LETTER L WITH CEDILLA	Ĺ
U+013D	LATIN CAPITAL LETTER L WITH CARON	Ľ
U+013E	LATIN SMALL LETTER L WITH CARON	ĺ
U+013F	LATIN CAPITAL LETTER L WITH MIDDLE DOT	Ḷ́
U+0140	LATIN SMALL LETTER L WITH MIDDLE DOT	ḷ́
U+0141	LATIN CAPITAL LETTER L WITH STROKE	Ł
U+0142	LATIN SMALL LETTER L WITH STROKE	ł
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	Ñ
U+0146	LATIN SMALL LETTER N WITH CEDILLA	ñ
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+014D,U+030D>	LATIN SMALL LETTER O WITH MACRON WITH COMBINING VERTICAL LINE ABOVE	ō̅
U+014E	LATIN CAPITAL LETTER O WITH BREVE	Ŏ
U+014F	LATIN SMALL LETTER O WITH BREVE	ŏ
U+0150	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE	Ő
U+0151	LATIN SMALL LETTER O WITH DOUBLE ACUTE	ő
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	ŕ
U+0156	LATIN CAPITAL LETTER R WITH CEDILLA	Ŗ
U+0157	LATIN SMALL LETTER R WITH CEDILLA	ŗ
U+0158	LATIN CAPITAL LETTER R WITH CARON	Ř
U+0159	LATIN SMALL LETTER R WITH CARON	ř
U+015A	LATIN CAPITAL LETTER S WITH ACUTE	Ŝ
U+015B	LATIN SMALL LETTER S WITH ACUTE	ŝ
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

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
U+015F	LATIN SMALL LETTER S WITH CEDILLA	ſ
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	Ř
U+0165	LATIN SMALL LETTER T WITH CARON	ř
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	Ũ
U+0169	LATIN SMALL LETTER U WITH TILDE	ũ
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	Ů
U+016F	LATIN SMALL LETTER U WITH RING ABOVE	ů
U+0170	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE	Ű
U+0171	LATIN SMALL LETTER U WITH DOUBLE ACUTE	ű
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	Ẓ̌
U+017E	LATIN SMALL LETTER Z WITH CARON	ž
<U+017E,U+0326>	LATIN SMALL LETTER Z WITH CARON WITH COMBINING COMMA BELOW	ẓ̌

Tabelle 3 (fortgesetzt)

Codepoints	Name	Glyphe
<U+017E,U+0327>	LATIN SMALL LETTER Z WITH CARON WITH COMBINING CEDILLA	ž
U+0187	LATIN CAPITAL LETTER C WITH HOOK	Ć
U+0188	LATIN SMALL LETTER C WITH HOOK	ć
U+018F	LATIN CAPITAL LETTER SCHWA	Ə
U+0197	LATIN CAPITAL LETTER I WITH STROKE	Ʀ
U+01A0	LATIN CAPITAL LETTER O WITH HORN	Ɔ
U+01A1	LATIN SMALL LETTER O WITH HORN	ɔ
U+01AF	LATIN CAPITAL LETTER U WITH HORN	Ʊ
U+01B0	LATIN SMALL LETTER U WITH HORN	ʊ
U+01B7	LATIN CAPITAL LETTER EZH	Ʒ
U+01CD	LATIN CAPITAL LETTER A WITH CARON	Ǻ
U+01CE	LATIN SMALL LETTER A WITH CARON	ǻ
U+01CF	LATIN CAPITAL LETTER I WITH CARON	Ǫ
U+01D0	LATIN SMALL LETTER I WITH CARON	ǫ
U+01D1	LATIN CAPITAL LETTER O WITH CARON	Ǭ
U+01D2	LATIN SMALL LETTER O WITH CARON	ǭ
U+01D3	LATIN CAPITAL LETTER U WITH CARON	Ǯ
U+01D4	LATIN SMALL LETTER U WITH CARON	ǯ
U+01D5	LATIN CAPITAL LETTER U WITH DIAERESIS AND MACRON	ǰ
U+01D6	LATIN SMALL LETTER U WITH DIAERESIS AND MACRON	Ǳ
U+01D7	LATIN CAPITAL LETTER U WITH DIAERESIS AND ACUTE	ǲ
U+01D8	LATIN SMALL LETTER U WITH DIAERESIS AND ACUTE	ǳ
U+01D9	LATIN CAPITAL LETTER U WITH DIAERESIS AND CARON	Ǵ
U+01DA	LATIN SMALL LETTER U WITH DIAERESIS AND CARON	ǵ
U+01DB	LATIN CAPITAL LETTER U WITH DIAERESIS AND GRAVE	Ƕ
U+01DC	LATIN SMALL LETTER U WITH DIAERESIS AND GRAVE	Ƿ
U+01DE	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON	Ǻ̃
U+01DF	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON	ǻ̃
U+01E2	LATIN CAPITAL LETTER AE WITH MACRON	Ǽ̃
U+01E3	LATIN SMALL LETTER AE WITH MACRON	ǽ̃
U+01E4	LATIN CAPITAL LETTER G WITH STROKE	Ɠ
U+01E5	LATIN SMALL LETTER G WITH STROKE	ɠ
U+01E6	LATIN CAPITAL LETTER G WITH CARON	Ǧ
U+01E7	LATIN SMALL LETTER G WITH CARON	ǧ
U+01E8	LATIN CAPITAL LETTER K WITH CARON	Ǭ̃
U+01E9	LATIN SMALL LETTER K WITH CARON	ǭ̃