

KOORDINIERUNGSSTELLE FÜR IT-STANDARDS (KoSIT)
BREMEN

LATEINISCHE ZEICHEN IN UNICODE

Version 1.1.1 vom 27. 01. 2012

Druckdatum:.....27. Januar 2012

Einleitung	1
1 Der Zeichensatz	4
1.1 Normalform und benannte Sequenzen	4
1.2 Encoding	5
1.3 Die Liste der Zeichen	5
2 Der Datentyp String.Latin	23
2.1 Datentyp in XML Schema	23
2.2 Datentyp als XÖV-UML-Modell	23
A Die maschinenlesbare Zeichensatzdatei	24
A.1 Normativer Teil	24
Element characterSet	24
Element characterDescription	25
A.2 Nicht-normativer Teil	26
Element mappings (in characterSet)	26
Element mapping (in characterDescription)	26

Einleitung

Alle Computer arbeiten letztendlich stets nur mit Zahlen. Buchstaben und anderen Zeichen können nur dann verarbeitet werden, wenn jedem einzelnen Zeichen eine Zahl zugewiesen wird. Es gibt auf der Welt viele verschiedene [Alphabete](#). Für die meisten davon gibt es wiederum viele unterschiedliche Zuordnungen der im Alphabet enthaltenen Zeichen zu Zahlen (so genannten Zeichencodierungen). Aus dem Nebeneinander vieler hundert verschiedener Codierungssysteme für diverse Alphabete resultieren Interoperabilitätsprobleme auf unterschiedlichen Ebenen.

- **Nicht angemessene bzw. unterschiedliche Alphabete:** Zeichen außerhalb des Zeichenvorrats können grundsätzlich nicht verarbeitet werden. Häufig werden Ersatzdarstellungen für fehlende Zeichen gewählt, was dann in der Regel zu Folgeproblemen bei der weiteren Verarbeitung der Daten führt. Außerdem führt dies zu Interoperabilitätsproblemen im Rahmen von Datenübermittlungen. Wenn Zeichen übermittelt werden, die der Empfänger nicht verarbeiten kann weil sie in dessen Zeichenvorrat nicht vorgesehen sind, so muss dies offensichtlich zu Fehlern führen.

Dieses Problem kommt häufiger vor als es zunächst den Anschein haben mag. Häufig werden Alphabete implizit durch technische Festlegungen bestimmt, Wenn beispielsweise für elektronisch geführte Register das Codierungssystem [ISO/IEC 8859-1](#) (*auch bekannt als ISO-Latin 1*) festgelegt wird, dann ist damit auch automatisch die Menge der möglichen Zeichen, also das zu Grunde liegende Alphabet bestimmt. Buchstaben außerhalb dieses Alphabets können nicht verarbeitet und nicht übermittelt werden.

- **Unterschiedliche Zeichencodierungen für das gleiche Alphabet:** Der Zeichenvorrat ist identisch, aber gleichen Zeichen werden unterschiedliche Nummern zugeordnet. In diesem Falle werden die vom Sender übermittelten Zeichen beim Empfänger fehlerhaft interpretiert, sofern nicht bei einer der beteiligten Stellen eine Umschlüsselung vorgenommen wird.

Im Zuge der Globalisierung hat sich die IT-Industrie dieser Problematik angenommen und eine Lösung entwickelt. Der [Unicode Standard](#) umfasst alle gebräuchlichen Alphabete und weist jedem Zeichen eine eindeutige Zahl (seinen *Codepoint*) zu. Darüber hinaus gibt es standardisierte Verfahren zur Abbildung der Codepoints auf Bytefolgen ([Unicode Transformation Format, UTF](#)).

Dies alles ist unabhängig von Produkten, Programmiersprachen und Sprachen. Unicode (bzw. die inhaltlich identische [ISO-Norm 10646](#)) ist der international anerkannte Standard für die Zeichencodierung und geeignet, die oben genannten Interoperabilitätsprobleme zu lösen. Unicode wird von allen wichtigen Betriebssystemen und systemnahen Anwendungen wie z. B. Datenbanken unterstützt.

Gleichwohl gibt es den Bedarf, den Unicode Standard für die Anwendung in der öffentlichen Verwaltung Deutschlands gemäß der bestehenden Anforderungen und Möglichkeiten einzuschränken. Die Vorgabe des Zeichensatzes im vollem Umfang für die IT-Verfahren der öffentlichen Verwaltung würde diese überfordern. Kulturelle Gründe, intellektuelle Fähigkeiten und technische Rahmenbedingungen erfordern eine Reduktion auf einen praxistauglichen Zeichenumfang. Der Unicode Standard umfasst unter anderem europäische, asiatische, kyrillische, afrikanische, arabische und indianische Alphabete. Kaum ein Mensch wird jeden der im Unicode-Standard enthaltenen Buchstaben¹ als solchen identifizieren und korrekt interpretieren können. Daher besteht der Bedarf, den Umfang von Unicode auf ein für die öffentliche Verwaltung Deutschlands geeignetes Maß einzuschränken. Das *"geeignete Maß"* ist das [Lateinische Alphabet](#), denn dieses wird im Verwaltungshandeln und der Registerführung in Deutschland regelhaft zu Grunde gelegt.

Das [Übereinkommen über die Angabe von Familiennamen und Vornamen in den Personenstandsbüchern](#) (BGBl. 1976 II, Seite 1473) macht klare Vorgaben zum Umgang mit den unterschiedlichen Alphabeten und Schriftzeichen der Welt. Durch das Übereinkommen soll erreicht werden, dass in allen Vertragsstaaten Namen von Personen ohne Rücksicht auf ihre Staatsangehörigkeit einheitlich in die Personenstandsbücher eingetragen werden. In dem Übereinkommen heißt es sinngemäß:

1. In der aktuellen Fassung des Unicode-Standard sind ca. 99.000 Zeichen enthalten, die einer der [Unicode Kategorien](#) *Li*, *Lm*, *Lo*, *Lt* angehören und insofern als *Letter (Buchstabe)* zu bezeichnen sind. Davon gehört der größte Teil (ca. 97.000 Zeichen) der Kategorie *Lo: other letters, including syllables and ideographs* an.

- **Transliteration bei unterschiedlichen Alphabeten:** Sofern eine Abschrift aus einer Urkunde vorzunehmen ist, die Familiennamen oder Vornamen mit Schriftzeichen wiedergibt, die nicht zu der Sprache zugehörig sind, in der die Eintragung vorgenommen werden soll, so sind diese Familiennamen und Vornamen ohne Übersetzung soweit wie möglich durch **Transliteration** wiederzugeben. Sind von der Internationalen Normenorganisation (ISO) empfohlene Normen vorhanden, so sind sie anzuwenden.
- **Buchstabengetreue Übernahme innerhalb eines Alphabets:** Sofern eine Abschrift aus einer Urkunde vorzunehmen ist, die Familiennamen oder Vornamen mit den selben Schriftzeichen wiedergibt, die der Sprache zugehörig sind, in der die Eintragung vorgenommen werden soll, so sind diese Familiennamen und Vornamen buchstabengetreu ohne Änderung oder Übersetzung wiederzugeben. Die in diesen Familiennamen und Vornamen enthaltenen **diakritischen Zeichen** sind ebenfalls wiederzugeben, *selbst wenn die Sprache, in der die Eintragung vorgenommen werden soll, solche Zeichen nicht kennt.*

Zusammengefasst ergibt sich aus den oben stehenden Ausführungen folgender Grundsatz für die Registerführung der öffentlichen Verwaltung Deutschlands:

Daten sind in lateinischer Schrift zu erfassen, diakritische Zeichen sind unverändert wiederzugeben. Dabei ist der Zeichensatz nach ISO-/IEC 10646 (der Unicode Standard) zu Grunde zu legen.

Der vorliegende Standard "Lateinische Zeichen in Unicode" unterstützt bei der Umsetzung dieses Grundsatzes, indem er die Teilmenge der Lateinischen Zeichen des Unicode Standard abschließend festlegt. In dem **Abschnitt 1 auf Seite 4** ist die Tabelle aller in Unicode enthaltenen lateinischen Zeichen mit ihrem Namen und ihrem Unicode Codepoint angegeben. Neben Buchstaben enthält der Zeichensatz auch Zeichen anderer Kategorien, wie z. B. Ziffern und Interpunktionszeichen. Nur dieser Abschnitt des Dokuments ist normativ.

Sie finden dieses Dokument im Internet unter http://xoev.de/latinchars/1_1/latinchars.pdf.

Zur Unterstützung der technischen Umsetzung des Standard werden darüber hinaus folgende Dateien zur Verfügung gestellt und in diesem Dokument beschrieben:

latinchars.xml Der Zeichensatz in maschinenlesbarer Form. Die Struktur der Datei wird in **Abschnitt A auf Seite 24** genauer beschrieben.

latinchars.xsd Diese Schema-Datei beinhaltet den Datentyp String.Latin, welcher den durch diesen Standard beschriebenen Zeichenumfang abbildet (wie in **Abschnitt 2 auf Seite 23** beschrieben).

latinchars.mdxml und latinchars.uml Das UML-Modell des Datentyps String.Latin zur Verwendung in XÖV-Standards (MagicDraw 16.5 Modell, XMI 2.0).

Umstellung auf lateinische Zeichen in Unicode - Vorgaben für Identifikationsverfahren

Empfehlungen der Arbeitsgruppe XInneres zu Identifikationsverfahren inklusive der Abbildung auf andere gebräuchliche Zeichensätze.

Veröffentlichungshistorie

In diesem Abschnitt werden die Änderungen an diesem Dokument zu den jeweiligen Veröffentlichungen aufgeführt.

Release 1.0.0 (02.03.2010)

Erste Veröffentlichung der *Lateinischen Zeichen in Unicode* als Bestandteil des XÖV-Handbuchs.

Release 1.1.0 (30.09.2011)

- Die *Lateinischen Zeichen in Unicode* wurden aus dem XÖV-Handbuch entfernt und als eigenständiger Standard veröffentlicht.
- Aufnahme zusätzlicher Zeichen.

Release 1.1.1 (25.01.2012)

Die Basisklasse des Datentypen `String.Latin` wurde von `xs:normalizedString` auf `xs:string` geändert. Der W3C-Datentyp `xs:normalizedString` erlaubt es Schemavalidatoren auch nach `String.Latin` gültige Zeichen durch ein Leerzeichen zu ersetzen. Dies ist bei `xs:string` nicht der Fall.

1. DER ZEICHENSATZ

Die Vorgabe der lateinischen Zeichen innerhalb von Unicode, die der Registerführung und Datenübermittlung in Deutschland zu Grunde gelegt werden soll, wurde wie folgt erstellt:

- a. Bei den Buchstaben (*Unicode-Kategorie LETTER*) erfolgte die Auswahl auf der Basis des "paneuropäischen Zeichensatzes" des Berichtes "Zeichen setzen für Europa"¹ von B. Kappenberg. *Ligaturen* wurden jedoch regelhaft nicht aufgenommen.
- b. Mit Ausnahme der meisten Zeichen der UNICODE-Kategorie *Control* wurden alle Zeichen des Unicode-Blockes **BASIC LATIN** übernommen. Dies betrifft Ziffern, Interpunktionszeichen, Klammerzeichen und so weiter. Die einzigen Zeichen der Kategorie *Control*, die übernommen wurden, sind: *CHARACTER TABULATION (0x09)*, *LINE FEED (0x0A)* und *CARRIAGE RETURN (0x0D)*.
- c. Dieser Vorschlag wurde im Rahmen des *Deutschland Online Vorhabens "Standardisierung"* in der öffentlichen Verwaltung Deutschlands durch das Bundesministerium des Innern mit Unterstützung der Koordinierungsstelle für IT-Standards (vormals: der OSCI-Leitstelle) abgestimmt.

1.1 Normalform und benannte Sequenzen

Der Unicode-Standard definiert einen numerischen Wert (Codepoint) und einen Namen für jedes seiner Zeichen. Der Zeichenvorrat von Unicode umfasst dabei sowohl Grundbuchstaben als auch diakritische Zeichen. Für Buchstaben mit diakritischen Zeichen (z. B. ü, Å) gibt es grundsätzlich zwei Möglichkeiten der Darstellung in Unicode:

- *decomposed* – Der Grundbuchstabe und das diakritische Zeichen werden durch aufeinanderfolgende Codepoints dargestellt. In diesem Fall werden für ein geschriebenes Zeichen u. U. mehrere Codepoints benötigt.
Beispiel: 'Å' wird in dieser Weise als Sequenz der Unicode-Zeichen 004A (LATIN CAPITAL LETTER A) und 030A5 (COMBINING RING ABOVE) dargestellt.
- *composed* – Die Zusammenfassung aus Grundzeichen und diakritischem Zeichen liegt als sogenanntes kombiniertes Zeichen mit einem eigenen Codepoint vor.
Beispiel: 'Å' wird in dieser Weise durch das Unicode-Zeichen 00C5 (LATIN CAPITAL LETTER A WITH RING ABOVE) dargestellt.

Um zu eindeutigen Darstellungen im Sinne der Codepoints zu kommen, beschreibt der Unicode-Standard im Standard Annex #15² verschiedene *Normalformen*. Für den Standard Lateinische Zeichen in Unicode gilt, in Übereinstimmung mit der gängigen Praxis und den Empfehlungen der W3C, die Normalform *NFC (Normalization Form Canonical Composition)*. Damit ist für alle kombinierten Zeichen, für die ein einzelner Codepoint vorliegt, ausschließlich diese Darstellung zulässig.

Für alle anderen kombinierten Zeichen in diesem Standard sind die Codepoints der entsprechenden *benannten Sequenz* gemäß Unicode Standard Annex #34³ zu verwenden.

Unicode Normalform und Codepoints

Beispiel: Zur Darstellung des Zeichens 'Å' ist ausschließlich die Verwendung des Codepoints 00C5 (LATIN CAPITAL LETTER A WITH RING ABOVE) zulässig.

Beispiel: Für das Zeichen 'Ð' existiert kein einzelner Codepoint. Es ist die benannte Sequenz der Codepoints 0044 und 0302 zu verwenden (LATIN CAPITAL LETTER D WITH COMBINING CIRCUMFLEX ACCENT).

1.<http://www.mediensprache.net/networx/networx-49.pdf>

2.<http://unicode.org/reports/tr15/>

3.<http://unicode.org/reports/tr34/>

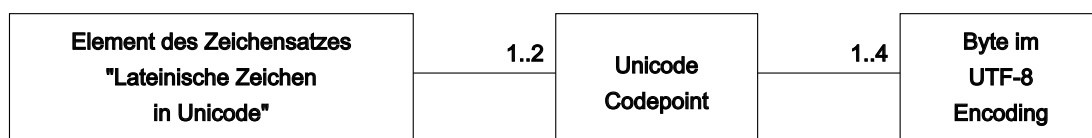
1.2 Encoding

Die Darstellung von Zeichen durch Codepoints ist zu unterscheiden von der Umsetzung der Codepoints in Byte-Folgen für die Speicherung oder Übermittlung von Zeichen. In der maschinenverarbeitbaren Darstellung werden Unicode-Codepoints in üblicherweise einer der drei Darstellungen UTF-8, UTF-16 oder UTF-32 kodiert, wobei UTF-8 das am weitesten verbreitete Encoding ist.

Der UTF-8-Standard verwendet zur Kodierung eines Unicode-Codepoints ein bis vier Bytes. Für die Kodierung des Zeichens 'Ð' (Codepoint-Sequenz 0044+0302) werden beispielsweise *drei* Byte benötigt, da die UTF-8-Darstellung des ersten Codepoints 0044 (des Grundbuchstabens) ein Byte benötigt und die UTF-8-Darstellung des zweiten Codepoints 0302 zwei Bytes.

[Bild 1-1 auf Seite 5](#) stellt den Zusammenhang zwischen den Elementen des hier beschriebenen Zeichensatzes, den Codepoints in Unicodes und der Kodierung in Byte-Folgen in UTF-8 als UML-Klassendiagramm dar.

Bild 1-1 Zusammenhang zwischen Zeichensatz-Element, Codepoint und UTF-8-Encoding



Der Standard Lateinische Zeichen in Unicode legt die Menge der zulässigen Zeichen mit ihren Unicode-Codepoints fest. Er trifft keine über Unicode hinausgehenden Aussagen über die Transformation in Bytefolgen.

1.3 Die Liste der Zeichen

Die nachfolgende Tabelle enthält alle Zeichen des Zeichensatzes Lateinische Zeichen in Unicode. Die Spalte 'Codepoint(s)' gibt die, für das jeweilige Zeichen zu verwendenden, Codepoints an. Die Spalte 'Name' enthält für die Zeichen, die in precomposed Form vorliegen, den Namen des Zeichens und für benannte Sequenzen den Namen der Sequenz, jeweils gemäß Unicode-Standard.

Tabelle 1-1: Liste aller zulässigen Zeichen

Codepoint(s)	Zeichen	Name	Kategorie
0009		CHARACTER TABULATION	OTHER, CONTROL
000A		LINE FEED	OTHER, CONTROL
000D		CARRIAGE RETURN	OTHER, CONTROL
0020		SPACE	SEPARATOR, SPACE
0021	!	EXCLAMATION MARK	PUNCTUATION, OTHER
0022	"	QUOTATION MARK	PUNCTUATION, OTHER
0023	#	NUMBER SIGN	PUNCTUATION, OTHER
0024	\$	DOLLAR SIGN	SYMBOL, CURRENCY
0025	%	PERCENT SIGN	PUNCTUATION, OTHER
0026	&	AMPERSAND	PUNCTUATION, OTHER
0027	'	APOSTROPHE	PUNCTUATION, OTHER
0028	(LEFT PARENTHESIS	PUNCTUATION, OPEN

Codepoint(s)	Zeichen	Name	Kategorie
0029)	RIGHT PARENTHESIS	PUNCTUATION, CLOSE
002A	*	ASTERISK	PUNCTUATION, OTHER
002B	+	PLUS SIGN	SYMBOL, MATH
002C	,	COMMA	PUNCTUATION, OTHER
002D	-	HYPHEN-MINUS	PUNCTUATION, DASH
002E	.	FULL STOP	PUNCTUATION, OTHER
002F	/	SOLIDUS	PUNCTUATION, OTHER
0030	0	DIGIT ZERO	NUMBER, DECIMAL DIGIT
0031	1	DIGIT ONE	NUMBER, DECIMAL DIGIT
0032	2	DIGIT TWO	NUMBER, DECIMAL DIGIT
0033	3	DIGIT THREE	NUMBER, DECIMAL DIGIT
0034	4	DIGIT FOUR	NUMBER, DECIMAL DIGIT
0035	5	DIGIT FIVE	NUMBER, DECIMAL DIGIT
0036	6	DIGIT SIX	NUMBER, DECIMAL DIGIT
0037	7	DIGIT SEVEN	NUMBER, DECIMAL DIGIT
0038	8	DIGIT EIGHT	NUMBER, DECIMAL DIGIT
0039	9	DIGIT NINE	NUMBER, DECIMAL DIGIT
003A	:	COLON	PUNCTUATION, OTHER
003B	;	SEMICOLON	PUNCTUATION, OTHER
003C	<	LESS-THAN SIGN	SYMBOL, MATH
003D	=	EQUALS SIGN	SYMBOL, MATH
003E	>	GREATER-THAN SIGN	SYMBOL, MATH
003F	?	QUESTION MARK	PUNCTUATION, OTHER
0040	@	COMMERCIAL AT	PUNCTUATION, OTHER
0041	A	LATIN CAPITAL LETTER A	LETTER, UPPERCASE
0042	B	LATIN CAPITAL LETTER B	LETTER, UPPERCASE
0043	C	LATIN CAPITAL LETTER C	LETTER, UPPERCASE
0044	D	LATIN CAPITAL LETTER D	LETTER, UPPERCASE
0044+0302	Ď	LATIN CAPITAL LETTER D WITH COMBINING CIRCUMFLEX ACCENT	LETTER,
0045	E	LATIN CAPITAL LETTER E	LETTER, UPPERCASE
0046	F	LATIN CAPITAL LETTER F	LETTER, UPPERCASE
0047	G	LATIN CAPITAL LETTER G	LETTER, UPPERCASE
0048	H	LATIN CAPITAL LETTER H	LETTER, UPPERCASE
0049	I	LATIN CAPITAL LETTER I	LETTER, UPPERCASE

Codepoint(s)	Zeichen	Name	Kategorie
004A	J	LATIN CAPITAL LETTER J	LETTER, UPPERCASE
004A+030C	Ĵ	LATIN CAPITAL LETTER J WITH COMBINING CARON	LETTER,
004B	K	LATIN CAPITAL LETTER K	LETTER, UPPERCASE
004C	L	LATIN CAPITAL LETTER L	LETTER, UPPERCASE
004C+0302	Ľ	LATIN CAPITAL LETTER L WITH COMBINING CIRCUMFLEX ACCENT	LETTER,
004D	M	LATIN CAPITAL LETTER M	LETTER, UPPERCASE
004D+0302	Ĺ	LATIN CAPITAL LETTER M WITH COMBINING CIRCUMFLEX ACCENT	LETTER,
004E	N	LATIN CAPITAL LETTER N	LETTER, UPPERCASE
004E+0302	Ñ	LATIN CAPITAL LETTER N WITH COMBINING CIRCUMFLEX ACCENT	LETTER,
004F	O	LATIN CAPITAL LETTER O	LETTER, UPPERCASE
0050	P	LATIN CAPITAL LETTER P	LETTER, UPPERCASE
0051	Q	LATIN CAPITAL LETTER Q	LETTER, UPPERCASE
0052	R	LATIN CAPITAL LETTER R	LETTER, UPPERCASE
0053	S	LATIN CAPITAL LETTER S	LETTER, UPPERCASE
0054	T	LATIN CAPITAL LETTER T	LETTER, UPPERCASE
0055	U	LATIN CAPITAL LETTER U	LETTER, UPPERCASE
0056	V	LATIN CAPITAL LETTER V	LETTER, UPPERCASE
0057	W	LATIN CAPITAL LETTER W	LETTER, UPPERCASE
0058	X	LATIN CAPITAL LETTER X	LETTER, UPPERCASE
0059	Y	LATIN CAPITAL LETTER Y	LETTER, UPPERCASE
005A	Z	LATIN CAPITAL LETTER Z	LETTER, UPPERCASE
005B	[LEFT SQUARE BRACKET	PUNCTUATION, OPEN
005C	\	REVERSE SOLIDUS	PUNCTUATION, OTHER
005D]	RIGHT SQUARE BRACKET	PUNCTUATION, CLOSE
005E	^	CIRCUMFLEX ACCENT	SYMBOL, MODIFIER
005F	_	LOW LINE	PUNCTUATION, CONNECTOR
0060	`	GRAVE ACCENT	SYMBOL, MODIFIER
0061	a	LATIN SMALL LETTER A	LETTER, LOWERCASE
0062	b	LATIN SMALL LETTER B	LETTER, LOWERCASE
0063	c	LATIN SMALL LETTER C	LETTER, LOWERCASE
0064	d	LATIN SMALL LETTER D	LETTER, LOWERCASE

Codepoint(s)	Zeichen	Name	Kategorie
0064+0302	â	LATIN SMALL LETTER D WITH COMBINING CIRCUMFLEX ACCENT	LETTER,
0065	e	LATIN SMALL LETTER E	LETTER, LOWERCASE
0066	f	LATIN SMALL LETTER F	LETTER, LOWERCASE
0067	g	LATIN SMALL LETTER G	LETTER, LOWERCASE
0068	h	LATIN SMALL LETTER H	LETTER, LOWERCASE
0069	i	LATIN SMALL LETTER I	LETTER, LOWERCASE
006A	j	LATIN SMALL LETTER J	LETTER, LOWERCASE
006B	k	LATIN SMALL LETTER K	LETTER, LOWERCASE
006C	l	LATIN SMALL LETTER L	LETTER, LOWERCASE
006C+0302	î	LATIN SMALL LETTER L WITH COMBINING CIRCUMFLEX ACCENT	LETTER,
006D	m	LATIN SMALL LETTER M	LETTER, LOWERCASE
006D+0302	ġ	LATIN SMALL LETTER M WITH COMBINING CIRCUMFLEX ACCENT	LETTER,
006E	n	LATIN SMALL LETTER N	LETTER, LOWERCASE
006E+0302	ņ	LATIN SMALL LETTER N WITH COMBINING CIRCUMFLEX ACCENT	LETTER,
006F	o	LATIN SMALL LETTER O	LETTER, LOWERCASE
0070	p	LATIN SMALL LETTER P	LETTER, LOWERCASE
0071	q	LATIN SMALL LETTER Q	LETTER, LOWERCASE
0072	r	LATIN SMALL LETTER R	LETTER, LOWERCASE
0073	s	LATIN SMALL LETTER S	LETTER, LOWERCASE
0074	t	LATIN SMALL LETTER T	LETTER, LOWERCASE
0075	u	LATIN SMALL LETTER U	LETTER, LOWERCASE
0076	v	LATIN SMALL LETTER V	LETTER, LOWERCASE
0077	w	LATIN SMALL LETTER W	LETTER, LOWERCASE
0078	x	LATIN SMALL LETTER X	LETTER, LOWERCASE
0079	y	LATIN SMALL LETTER Y	LETTER, LOWERCASE
007A	z	LATIN SMALL LETTER Z	LETTER, LOWERCASE
007B	{	LEFT CURLY BRACKET	PUNCTUATION, OPEN
007C		VERTICAL LINE	SYMBOL, MATH
007D	}	RIGHT CURLY BRACKET	PUNCTUATION, CLOSE
007E	~	TILDE	SYMBOL, MATH
00A1	¡	INVERTED EXCLAMATION MARK	PUNCTUATION, OTHER
00A2	¢	CENT SIGN	SYMBOL, CURRENCY

Codepoint(s)	Zeichen	Name	Kategorie
00A3	£	POUND SIGN	SYMBOL, CURRENCY
00A4	¤	CURRENCY SIGN	SYMBOL, CURRENCY
00A5	¥	YEN SIGN	SYMBOL, CURRENCY
00A6		BROKEN BAR	SYMBOL, OTHER
00A7	§	SECTION SIGN	SYMBOL, OTHER
00A8	¨	DIAERESIS	SYMBOL, MODIFIER
00A9	©	COPYRIGHT SIGN	SYMBOL, OTHER
00AA	ª	FEMININE ORDINAL INDICATOR	LETTER, LOWERCASE
00AB	«	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK	PUNCTUATION, INITIAL QUOTE
00AC	¬	NOT SIGN	SYMBOL, MATH
00AE	®	REGISTERED SIGN	SYMBOL, OTHER
00AF	¯	MACRON	SYMBOL, MODIFIER
00B0	°	DEGREE SIGN	SYMBOL, OTHER
00B1	±	PLUS-MINUS SIGN	SYMBOL, MATH
00B2	²	SUPERSCRIPIT TWO	NUMBER, OTHER
00B3	³	SUPERSCRIPIT THREE	NUMBER, OTHER
00B4	´	ACUTE ACCENT	SYMBOL, MODIFIER
00B5	µ	MICRO SIGN	LETTER, LOWERCASE
00B6	¶	PILCROW SIGN	SYMBOL, OTHER
00B7	·	MIDDLE DOT	PUNCTUATION, OTHER
00B8	¸	CEDILLA	SYMBOL, MODIFIER
00B9	¹	SUPERSCRIPIT ONE	NUMBER, OTHER
00BA	º	MASCULINE ORDINAL INDICATOR	LETTER, LOWERCASE
00BB	»	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK	PUNCTUATION, FINAL QUOTE
00BC	¼	VULGAR FRACTION ONE QUARTER	NUMBER, OTHER
00BD	½	VULGAR FRACTION ONE HALF	NUMBER, OTHER
00BE	¾	VULGAR FRACTION THREE QUARTERS	NUMBER, OTHER
00BF	¿	INVERTED QUESTION MARK	PUNCTUATION, OTHER
00C0	À	LATIN CAPITAL LETTER A WITH GRAVE	LETTER, UPPERCASE
00C1	Á	LATIN CAPITAL LETTER A WITH ACUTE	LETTER, UPPERCASE
00C2	Â	LATIN CAPITAL LETTER A WITH CIRCUMFLEX	LETTER, UPPERCASE
00C3	Ã	LATIN CAPITAL LETTER A WITH TILDE	LETTER, UPPERCASE

Codepoint(s)	Zeichen	Name	Kategorie
00C4	Ä	LATIN CAPITAL LETTER A WITH DIAERESIS	LETTER, UPPERCASE
00C5	Å	LATIN CAPITAL LETTER A WITH RING ABOVE	LETTER, UPPERCASE
00C6	Æ	LATIN CAPITAL LETTER AE	LETTER, UPPERCASE
00C7	Ç	LATIN CAPITAL LETTER C WITH CEDILLA	LETTER, UPPERCASE
00C8	È	LATIN CAPITAL LETTER E WITH GRAVE	LETTER, UPPERCASE
00C9	É	LATIN CAPITAL LETTER E WITH ACUTE	LETTER, UPPERCASE
00CA	Ê	LATIN CAPITAL LETTER E WITH CIRCUMFLEX	LETTER, UPPERCASE
00CB	Ë	LATIN CAPITAL LETTER E WITH DIAERESIS	LETTER, UPPERCASE
00CC	Ì	LATIN CAPITAL LETTER I WITH GRAVE	LETTER, UPPERCASE
00CD	Í	LATIN CAPITAL LETTER I WITH ACUTE	LETTER, UPPERCASE
00CE	Î	LATIN CAPITAL LETTER I WITH CIRCUMFLEX	LETTER, UPPERCASE
00CF	Ï	LATIN CAPITAL LETTER I WITH DIAERESIS	LETTER, UPPERCASE
00D0	Ð	LATIN CAPITAL LETTER ETH	LETTER, UPPERCASE
00D1	Ñ	LATIN CAPITAL LETTER N WITH TILDE	LETTER, UPPERCASE
00D2	Ò	LATIN CAPITAL LETTER O WITH GRAVE	LETTER, UPPERCASE
00D3	Ó	LATIN CAPITAL LETTER O WITH ACUTE	LETTER, UPPERCASE
00D4	Ô	LATIN CAPITAL LETTER O WITH CIRCUMFLEX	LETTER, UPPERCASE
00D5	Õ	LATIN CAPITAL LETTER O WITH TILDE	LETTER, UPPERCASE
00D6	Ö	LATIN CAPITAL LETTER O WITH DIAERESIS	LETTER, UPPERCASE
00D7	×	MULTIPLICATION SIGN	SYMBOL, MATH
00D8	Ø	LATIN CAPITAL LETTER O WITH STROKE	LETTER, UPPERCASE
00D9	Ù	LATIN CAPITAL LETTER U WITH GRAVE	LETTER, UPPERCASE
00DA	Ú	LATIN CAPITAL LETTER U WITH ACUTE	LETTER, UPPERCASE
00DB	Û	LATIN CAPITAL LETTER U WITH CIRCUMFLEX	LETTER, UPPERCASE
00DC	Ü	LATIN CAPITAL LETTER U WITH DIAERESIS	LETTER, UPPERCASE
00DD	Ý	LATIN CAPITAL LETTER Y WITH ACUTE	LETTER, UPPERCASE
00DE	Þ	LATIN CAPITAL LETTER THORN	LETTER, UPPERCASE
00DF	ß	LATIN SMALL LETTER SHARP S	LETTER, LOWERCASE
00E0	à	LATIN SMALL LETTER A WITH GRAVE	LETTER, LOWERCASE

Codepoint(s)	Zeichen	Name	Kategorie
00E1	á	LATIN SMALL LETTER A WITH ACUTE	LETTER, LOWERCASE
00E2	â	LATIN SMALL LETTER A WITH CIRCUMFLEX	LETTER, LOWERCASE
00E3	ã	LATIN SMALL LETTER A WITH TILDE	LETTER, LOWERCASE
00E4	ä	LATIN SMALL LETTER A WITH DIAERESIS	LETTER, LOWERCASE
00E5	å	LATIN SMALL LETTER A WITH RING ABOVE	LETTER, LOWERCASE
00E6	æ	LATIN SMALL LETTER AE	LETTER, LOWERCASE
00E7	ç	LATIN SMALL LETTER C WITH CEDILLA	LETTER, LOWERCASE
00E8	è	LATIN SMALL LETTER E WITH GRAVE	LETTER, LOWERCASE
00E9	é	LATIN SMALL LETTER E WITH ACUTE	LETTER, LOWERCASE
00EA	ê	LATIN SMALL LETTER E WITH CIRCUMFLEX	LETTER, LOWERCASE
00EB	ë	LATIN SMALL LETTER E WITH DIAERESIS	LETTER, LOWERCASE
00EC	ì	LATIN SMALL LETTER I WITH GRAVE	LETTER, LOWERCASE
00ED	í	LATIN SMALL LETTER I WITH ACUTE	LETTER, LOWERCASE
00EE	î	LATIN SMALL LETTER I WITH CIRCUMFLEX	LETTER, LOWERCASE
00EF	ï	LATIN SMALL LETTER I WITH DIAERESIS	LETTER, LOWERCASE
00F0	ð	LATIN SMALL LETTER ETH	LETTER, LOWERCASE
00F1	ñ	LATIN SMALL LETTER N WITH TILDE	LETTER, LOWERCASE
00F2	ò	LATIN SMALL LETTER O WITH GRAVE	LETTER, LOWERCASE
00F3	ó	LATIN SMALL LETTER O WITH ACUTE	LETTER, LOWERCASE
00F4	ô	LATIN SMALL LETTER O WITH CIRCUMFLEX	LETTER, LOWERCASE
00F5	õ	LATIN SMALL LETTER O WITH TILDE	LETTER, LOWERCASE
00F6	ö	LATIN SMALL LETTER O WITH DIAERESIS	LETTER, LOWERCASE
00F7	÷	DIVISION SIGN	SYMBOL, MATH
00F8	ø	LATIN SMALL LETTER O WITH STROKE	LETTER, LOWERCASE
00F9	ù	LATIN SMALL LETTER U WITH GRAVE	LETTER, LOWERCASE
00FA	ú	LATIN SMALL LETTER U WITH ACUTE	LETTER, LOWERCASE
00FB	û	LATIN SMALL LETTER U WITH CIRCUMFLEX	LETTER, LOWERCASE
00FC	ü	LATIN SMALL LETTER U WITH DIAERESIS	LETTER, LOWERCASE
00FD	ý	LATIN SMALL LETTER Y WITH ACUTE	LETTER, LOWERCASE
00FE	þ	LATIN SMALL LETTER THORN	LETTER, LOWERCASE
00FF	ÿ	LATIN SMALL LETTER Y WITH DIAERESIS	LETTER, LOWERCASE

Codepoint(s)	Zeichen	Name	Kategorie
0100	Ā	LATIN CAPITAL LETTER A WITH MACRON	LETTER, UPPERCASE
0101	ā	LATIN SMALL LETTER A WITH MACRON	LETTER, LOWERCASE
0102	Ă	LATIN CAPITAL LETTER A WITH BREVE	LETTER, UPPERCASE
0103	ă	LATIN SMALL LETTER A WITH BREVE	LETTER, LOWERCASE
0104	Ą	LATIN CAPITAL LETTER A WITH OGONEK	LETTER, UPPERCASE
0105	ą	LATIN SMALL LETTER A WITH OGONEK	LETTER, LOWERCASE
0106	Ć	LATIN CAPITAL LETTER C WITH ACUTE	LETTER, UPPERCASE
0107	ć	LATIN SMALL LETTER C WITH ACUTE	LETTER, LOWERCASE
010A	Ĉ	LATIN CAPITAL LETTER C WITH DOT ABOVE	LETTER, UPPERCASE
010B	ĉ	LATIN SMALL LETTER C WITH DOT ABOVE	LETTER, LOWERCASE
010C	Č	LATIN CAPITAL LETTER C WITH CARON	LETTER, UPPERCASE
010D	č	LATIN SMALL LETTER C WITH CARON	LETTER, LOWERCASE
010E	Ď	LATIN CAPITAL LETTER D WITH CARON	LETTER, UPPERCASE
010F	ď	LATIN SMALL LETTER D WITH CARON	LETTER, LOWERCASE
0110	Ð	LATIN CAPITAL LETTER D WITH STROKE	LETTER, UPPERCASE
0111	ð	LATIN SMALL LETTER D WITH STROKE	LETTER, LOWERCASE
0112	Ē	LATIN CAPITAL LETTER E WITH MACRON	LETTER, UPPERCASE
0113	ē	LATIN SMALL LETTER E WITH MACRON	LETTER, LOWERCASE
0114	Ĕ	LATIN CAPITAL LETTER E WITH BREVE	LETTER, UPPERCASE
0115	ĕ	LATIN SMALL LETTER E WITH BREVE	LETTER, LOWERCASE
0116	Ė	LATIN CAPITAL LETTER E WITH DOT ABOVE	LETTER, UPPERCASE
0117	ė	LATIN SMALL LETTER E WITH DOT ABOVE	LETTER, LOWERCASE
0118	Ę	LATIN CAPITAL LETTER E WITH OGONEK	LETTER, UPPERCASE
0119	ę	LATIN SMALL LETTER E WITH OGONEK	LETTER, LOWERCASE
011A	Ě	LATIN CAPITAL LETTER E WITH CARON	LETTER, UPPERCASE
011B	ě	LATIN SMALL LETTER E WITH CARON	LETTER, LOWERCASE
011E	Ĝ	LATIN CAPITAL LETTER G WITH BREVE	LETTER, UPPERCASE
011F	ĝ	LATIN SMALL LETTER G WITH BREVE	LETTER, LOWERCASE
0120	Ğ	LATIN CAPITAL LETTER G WITH DOT ABOVE	LETTER, UPPERCASE
0121	ğ	LATIN SMALL LETTER G WITH DOT ABOVE	LETTER, LOWERCASE
0122	Ġ	LATIN CAPITAL LETTER G WITH CEDILLA	LETTER, UPPERCASE
0123	ġ	LATIN SMALL LETTER G WITH CEDILLA	LETTER, LOWERCASE
0126	Ĥ	LATIN CAPITAL LETTER H WITH STROKE	LETTER, UPPERCASE

Codepoint(s)	Zeichen	Name	Kategorie
0127	ħ	LATIN SMALL LETTER H WITH STROKE	LETTER, LOWERCASE
0128	Ī	LATIN CAPITAL LETTER I WITH TILDE	LETTER, UPPERCASE
0129	ī	LATIN SMALL LETTER I WITH TILDE	LETTER, LOWERCASE
012A	Ī	LATIN CAPITAL LETTER I WITH MACRON	LETTER, UPPERCASE
012B	ī	LATIN SMALL LETTER I WITH MACRON	LETTER, LOWERCASE
012C	İ	LATIN CAPITAL LETTER I WITH BREVE	LETTER, UPPERCASE
012D	ı	LATIN SMALL LETTER I WITH BREVE	LETTER, LOWERCASE
012E	Į	LATIN CAPITAL LETTER I WITH OGONEK	LETTER, UPPERCASE
012F	į	LATIN SMALL LETTER I WITH OGONEK	LETTER, LOWERCASE
0130	İ	LATIN CAPITAL LETTER I WITH DOT ABOVE	LETTER, UPPERCASE
0131	ı	LATIN SMALL LETTER DOTLESS I	LETTER, LOWERCASE
0134	Ĵ	LATIN CAPITAL LETTER J WITH CIRCUMFLEX	LETTER, UPPERCASE
0135	ĵ	LATIN SMALL LETTER J WITH CIRCUMFLEX	LETTER, LOWERCASE
0136	Ķ	LATIN CAPITAL LETTER K WITH CEDILLA	LETTER, UPPERCASE
0137	ķ	LATIN SMALL LETTER K WITH CEDILLA	LETTER, LOWERCASE
0138	κ	LATIN SMALL LETTER KRA	LETTER, LOWERCASE
0139	Ĺ	LATIN CAPITAL LETTER L WITH ACUTE	LETTER, UPPERCASE
013A	ĺ	LATIN SMALL LETTER L WITH ACUTE	LETTER, LOWERCASE
013B	Ľ	LATIN CAPITAL LETTER L WITH CEDILLA	LETTER, UPPERCASE
013C	ļ	LATIN SMALL LETTER L WITH CEDILLA	LETTER, LOWERCASE
013D	Ľ	LATIN CAPITAL LETTER L WITH CARON	LETTER, UPPERCASE
013E	ľ	LATIN SMALL LETTER L WITH CARON	LETTER, LOWERCASE
013F	Ł	LATIN CAPITAL LETTER L WITH MIDDLE DOT	LETTER, UPPERCASE
0140	ł	LATIN SMALL LETTER L WITH MIDDLE DOT	LETTER, LOWERCASE
0141	Ł	LATIN CAPITAL LETTER L WITH STROKE	LETTER, UPPERCASE
0142	ł	LATIN SMALL LETTER L WITH STROKE	LETTER, LOWERCASE
0143	Ń	LATIN CAPITAL LETTER N WITH ACUTE	LETTER, UPPERCASE
0144	ń	LATIN SMALL LETTER N WITH ACUTE	LETTER, LOWERCASE
0145	Ŋ	LATIN CAPITAL LETTER N WITH CEDILLA	LETTER, UPPERCASE
0146	ņ	LATIN SMALL LETTER N WITH CEDILLA	LETTER, LOWERCASE
0147	Ń	LATIN CAPITAL LETTER N WITH CARON	LETTER, UPPERCASE
0148	ň	LATIN SMALL LETTER N WITH CARON	LETTER, LOWERCASE

Codepoint(s)	Zeichen	Name	Kategorie
0149	ñ	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE	LETTER, LOWERCASE
014A	Ɔ	LATIN CAPITAL LETTER ENG	LETTER, UPPERCASE
014B	ɲ	LATIN SMALL LETTER ENG	LETTER, LOWERCASE
014C	Ō	LATIN CAPITAL LETTER O WITH MACRON	LETTER, UPPERCASE
014D	ō	LATIN SMALL LETTER O WITH MACRON	LETTER, LOWERCASE
014E	Ŏ	LATIN CAPITAL LETTER O WITH BREVE	LETTER, UPPERCASE
014F	ŏ	LATIN SMALL LETTER O WITH BREVE	LETTER, LOWERCASE
0150	Ŏ̇	LATIN CAPITAL LETTER O WITH DOUBLE ACUTE	LETTER, UPPERCASE
0151	ȯ	LATIN SMALL LETTER O WITH DOUBLE ACUTE	LETTER, LOWERCASE
0152	Œ	LATIN CAPITAL LIGATURE OE	LETTER, UPPERCASE
0153	œ	LATIN SMALL LIGATURE OE	LETTER, LOWERCASE
0154	Ŕ	LATIN CAPITAL LETTER R WITH ACUTE	LETTER, UPPERCASE
0155	ř	LATIN SMALL LETTER R WITH ACUTE	LETTER, LOWERCASE
0156	Ŗ	LATIN CAPITAL LETTER R WITH CEDILLA	LETTER, UPPERCASE
0157	ŗ	LATIN SMALL LETTER R WITH CEDILLA	LETTER, LOWERCASE
0158	Ř	LATIN CAPITAL LETTER R WITH CARON	LETTER, UPPERCASE
0159	ř	LATIN SMALL LETTER R WITH CARON	LETTER, LOWERCASE
015A	Ś	LATIN CAPITAL LETTER S WITH ACUTE	LETTER, UPPERCASE
015B	ś	LATIN SMALL LETTER S WITH ACUTE	LETTER, LOWERCASE
015E	Ş	LATIN CAPITAL LETTER S WITH CEDILLA	LETTER, UPPERCASE
015F	ş	LATIN SMALL LETTER S WITH CEDILLA	LETTER, LOWERCASE
0160	Š	LATIN CAPITAL LETTER S WITH CARON	LETTER, UPPERCASE
0161	š	LATIN SMALL LETTER S WITH CARON	LETTER, LOWERCASE
0162	Ţ	LATIN CAPITAL LETTER T WITH CEDILLA	LETTER, UPPERCASE
0163	ţ	LATIN SMALL LETTER T WITH CEDILLA	LETTER, LOWERCASE
0164	Ť	LATIN CAPITAL LETTER T WITH CARON	LETTER, UPPERCASE
0165	ť	LATIN SMALL LETTER T WITH CARON	LETTER, LOWERCASE
0166	Ƨ	LATIN CAPITAL LETTER T WITH STROKE	LETTER, UPPERCASE
0167	Ƨ̇	LATIN SMALL LETTER T WITH STROKE	LETTER, LOWERCASE
0168	Ŭ	LATIN CAPITAL LETTER U WITH TILDE	LETTER, UPPERCASE
0169	ů	LATIN SMALL LETTER U WITH TILDE	LETTER, LOWERCASE
016A	Ū	LATIN CAPITAL LETTER U WITH MACRON	LETTER, UPPERCASE
016B	ū	LATIN SMALL LETTER U WITH MACRON	LETTER, LOWERCASE

Codepoint(s)	Zeichen	Name	Kategorie
016E	Ū	LATIN CAPITAL LETTER U WITH RING ABOVE	LETTER, UPPERCASE
016F	ū	LATIN SMALL LETTER U WITH RING ABOVE	LETTER, LOWERCASE
0170	Ů	LATIN CAPITAL LETTER U WITH DOUBLE ACUTE	LETTER, UPPERCASE
0171	ů	LATIN SMALL LETTER U WITH DOUBLE ACUTE	LETTER, LOWERCASE
0172	Ů	LATIN CAPITAL LETTER U WITH OGONEK	LETTER, UPPERCASE
0173	ů	LATIN SMALL LETTER U WITH OGONEK	LETTER, LOWERCASE
0174	Ŵ	LATIN CAPITAL LETTER W WITH CIRCUMFLEX	LETTER, UPPERCASE
0175	ŵ	LATIN SMALL LETTER W WITH CIRCUMFLEX	LETTER, LOWERCASE
0176	Ŷ	LATIN CAPITAL LETTER Y WITH CIRCUMFLEX	LETTER, UPPERCASE
0177	ŷ	LATIN SMALL LETTER Y WITH CIRCUMFLEX	LETTER, LOWERCASE
0178	ÿ	LATIN CAPITAL LETTER Y WITH DIAERESIS	LETTER, UPPERCASE
0179	Ž	LATIN CAPITAL LETTER Z WITH ACUTE	LETTER, UPPERCASE
017A	ž	LATIN SMALL LETTER Z WITH ACUTE	LETTER, LOWERCASE
017B	Ž	LATIN CAPITAL LETTER Z WITH DOT ABOVE	LETTER, UPPERCASE
017C	ž	LATIN SMALL LETTER Z WITH DOT ABOVE	LETTER, LOWERCASE
017D	Ž	LATIN CAPITAL LETTER Z WITH CARON	LETTER, UPPERCASE
017E	ž	LATIN SMALL LETTER Z WITH CARON	LETTER, LOWERCASE
018F	ø	LATIN CAPITAL LETTER SCHWA	LETTER, UPPERCASE
01A0	Ɔ	LATIN CAPITAL LETTER O WITH HORN	LETTER, UPPERCASE
01A1	ɔ	LATIN SMALL LETTER O WITH HORN	LETTER, LOWERCASE
01AF	Ɔ	LATIN CAPITAL LETTER U WITH HORN	LETTER, UPPERCASE
01B0	ɔ	LATIN SMALL LETTER U WITH HORN	LETTER, LOWERCASE
01B7	Ɔ	LATIN CAPITAL LETTER EZH	LETTER, UPPERCASE
01CD	Ǻ	LATIN CAPITAL LETTER A WITH CARON	LETTER, UPPERCASE
01CE	ǻ	LATIN SMALL LETTER A WITH CARON	LETTER, LOWERCASE
01CF	Ǫ	LATIN CAPITAL LETTER I WITH CARON	LETTER, UPPERCASE
01D0	ǫ	LATIN SMALL LETTER I WITH CARON	LETTER, LOWERCASE
01D1	Ǿ	LATIN CAPITAL LETTER O WITH CARON	LETTER, UPPERCASE
01D2	ǿ	LATIN SMALL LETTER O WITH CARON	LETTER, LOWERCASE

Codepoint(s)	Zeichen	Name	Kategorie
01D3	Ů	LATIN CAPITAL LETTER U WITH CARON	LETTER, UPPERCASE
01D4	ů	LATIN SMALL LETTER U WITH CARON	LETTER, LOWERCASE
01DE	Ā	LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON	LETTER, UPPERCASE
01DF	ā	LATIN SMALL LETTER A WITH DIAERESIS AND MACRON	LETTER, LOWERCASE
01E4	Ɔ	LATIN CAPITAL LETTER G WITH STROKE	LETTER, UPPERCASE
01E5	Ɠ	LATIN SMALL LETTER G WITH STROKE	LETTER, LOWERCASE
01E6	Ǧ	LATIN CAPITAL LETTER G WITH CARON	LETTER, UPPERCASE
01E7	ǧ	LATIN SMALL LETTER G WITH CARON	LETTER, LOWERCASE
01E8	Ǧ	LATIN CAPITAL LETTER K WITH CARON	LETTER, UPPERCASE
01E9	ǧ	LATIN SMALL LETTER K WITH CARON	LETTER, LOWERCASE
01EA	Ɔ	LATIN CAPITAL LETTER O WITH OGONEK	LETTER, UPPERCASE
01EB	ɔ	LATIN SMALL LETTER O WITH OGONEK	LETTER, LOWERCASE
01EC	Ō	LATIN CAPITAL LETTER O WITH OGONEK AND MACRON	LETTER, UPPERCASE
01ED	ō	LATIN SMALL LETTER O WITH OGONEK AND MACRON	LETTER, LOWERCASE
01EE	ǰ	LATIN CAPITAL LETTER EZH WITH CARON	LETTER, UPPERCASE
01EF	Ǳ	LATIN SMALL LETTER EZH WITH CARON	LETTER, LOWERCASE
01F0	ǰ	LATIN SMALL LETTER J WITH CARON	LETTER, LOWERCASE
01F4	Ĝ	LATIN CAPITAL LETTER G WITH ACUTE	LETTER, UPPERCASE
01F5	ĝ	LATIN SMALL LETTER G WITH ACUTE	LETTER, LOWERCASE
01FA	Ā	LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE	LETTER, UPPERCASE
01FB	á	LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE	LETTER, LOWERCASE
01FC	Æ	LATIN CAPITAL LETTER AE WITH ACUTE	LETTER, UPPERCASE
01FD	æ	LATIN SMALL LETTER AE WITH ACUTE	LETTER, LOWERCASE
01FE	Œ	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE	LETTER, UPPERCASE
01FF	œ	LATIN SMALL LETTER O WITH STROKE AND ACUTE	LETTER, LOWERCASE
0218	Ş	LATIN CAPITAL LETTER S WITH COMMA BELOW	LETTER, UPPERCASE
0219	ş	LATIN SMALL LETTER S WITH COMMA BELOW	LETTER, LOWERCASE
021A	Ț	LATIN CAPITAL LETTER T WITH COMMA BELOW	LETTER, UPPERCASE

Codepoint(s)	Zeichen	Name	Kategorie
021B	‡	LATIN SMALL LETTER T WITH COMMA BELOW	LETTER, LOWERCASE
021E	Ĥ	LATIN CAPITAL LETTER H WITH CARON	LETTER, UPPERCASE
021F	ĥ	LATIN SMALL LETTER H WITH CARON	LETTER, LOWERCASE
022A	Ŏ	LATIN CAPITAL LETTER O WITH DIAERESIS AND MACRON	LETTER, UPPERCASE
022B	õ	LATIN SMALL LETTER O WITH DIAERESIS AND MACRON	LETTER, LOWERCASE
022E	Ô	LATIN CAPITAL LETTER O WITH DOT ABOVE	LETTER, UPPERCASE
022F	ó	LATIN SMALL LETTER O WITH DOT ABOVE	LETTER, LOWERCASE
0230	Ŏ	LATIN CAPITAL LETTER O WITH DOT ABOVE AND MACRON	LETTER, UPPERCASE
0231	õ	LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON	LETTER, LOWERCASE
0232	ÿ	LATIN CAPITAL LETTER Y WITH MACRON	LETTER, UPPERCASE
0233	ÿ	LATIN SMALL LETTER Y WITH MACRON	LETTER, LOWERCASE
0259	ə	LATIN SMALL LETTER SCHWA	LETTER, LOWERCASE
0292	Ʒ	LATIN SMALL LETTER EZH	LETTER, LOWERCASE
1E02	Ā	LATIN CAPITAL LETTER A WITH DOT ABOVE	LETTER, UPPERCASE
1E03	ā	LATIN SMALL LETTER A WITH DOT ABOVE	LETTER, LOWERCASE
1E0A	Ď	LATIN CAPITAL LETTER D WITH DOT ABOVE	LETTER, UPPERCASE
1E0B	ď	LATIN SMALL LETTER D WITH DOT ABOVE	LETTER, LOWERCASE
1E10	Ď	LATIN CAPITAL LETTER D WITH CEDILLA	LETTER, UPPERCASE
1E11	ď	LATIN SMALL LETTER D WITH CEDILLA	LETTER, LOWERCASE
1E1E	Ħ	LATIN CAPITAL LETTER F WITH DOT ABOVE	LETTER, UPPERCASE
1E1F	ƒ	LATIN SMALL LETTER F WITH DOT ABOVE	LETTER, LOWERCASE
1E20	Ĝ	LATIN CAPITAL LETTER G WITH MACRON	LETTER, UPPERCASE
1E21	ĝ	LATIN SMALL LETTER G WITH MACRON	LETTER, LOWERCASE
1E24	Ħ	LATIN CAPITAL LETTER H WITH DOT BELOW	LETTER, UPPERCASE
1E25	ħ	LATIN SMALL LETTER H WITH DOT BELOW	LETTER, LOWERCASE
1E26	Ĥ	LATIN CAPITAL LETTER H WITH DIAERESIS	LETTER, UPPERCASE
1E27	ĥ	LATIN SMALL LETTER H WITH DIAERESIS	LETTER, LOWERCASE
1E30	Ķ	LATIN CAPITAL LETTER K WITH ACUTE	LETTER, UPPERCASE

Codepoint(s)	Zeichen	Name	Kategorie
1E31	ĳ	LATIN SMALL LETTER K WITH ACUTE	LETTER, LOWERCASE
1E40	Ĭ	LATIN CAPITAL LETTER M WITH DOT ABOVE	LETTER, UPPERCASE
1E41	ĩ	LATIN SMALL LETTER M WITH DOT ABOVE	LETTER, LOWERCASE
1E44	Ñ	LATIN CAPITAL LETTER N WITH DOT ABOVE	LETTER, UPPERCASE
1E45	ñ	LATIN SMALL LETTER N WITH DOT ABOVE	LETTER, LOWERCASE
1E56	Ĵ	LATIN CAPITAL LETTER P WITH DOT ABOVE	LETTER, UPPERCASE
1E57	ĵ	LATIN SMALL LETTER P WITH DOT ABOVE	LETTER, LOWERCASE
1E60	Ŝ	LATIN CAPITAL LETTER S WITH DOT ABOVE	LETTER, UPPERCASE
1E61	ŝ	LATIN SMALL LETTER S WITH DOT ABOVE	LETTER, LOWERCASE
1E62	Ș	LATIN CAPITAL LETTER S WITH DOT BELOW	LETTER, UPPERCASE
1E63	ș	LATIN SMALL LETTER S WITH DOT BELOW	LETTER, LOWERCASE
1E6A	Ţ	LATIN CAPITAL LETTER T WITH DOT ABOVE	LETTER, UPPERCASE
1E6B	ţ	LATIN SMALL LETTER T WITH DOT ABOVE	LETTER, LOWERCASE
1E80	Ŵ	LATIN CAPITAL LETTER W WITH GRAVE	LETTER, UPPERCASE
1E81	ŵ	LATIN SMALL LETTER W WITH GRAVE	LETTER, LOWERCASE
1E82	Ŷ	LATIN CAPITAL LETTER W WITH ACUTE	LETTER, UPPERCASE
1E83	ŷ	LATIN SMALL LETTER W WITH ACUTE	LETTER, LOWERCASE
1E84	Ÿ	LATIN CAPITAL LETTER W WITH DIAERESIS	LETTER, UPPERCASE
1E85	ÿ	LATIN SMALL LETTER W WITH DIAERESIS	LETTER, LOWERCASE
1E8C	Ẋ	LATIN CAPITAL LETTER X WITH DIAERESIS	LETTER, UPPERCASE
1E8D	ẋ	LATIN SMALL LETTER X WITH DIAERESIS	LETTER, LOWERCASE
1E8E	Ỳ	LATIN CAPITAL LETTER Y WITH DOT ABOVE	LETTER, UPPERCASE
1E8F	ỳ	LATIN SMALL LETTER Y WITH DOT ABOVE	LETTER, LOWERCASE
1E90	Ẑ	LATIN CAPITAL LETTER Z WITH CIRCUMFLEX	LETTER, UPPERCASE
1E91	ẑ	LATIN SMALL LETTER Z WITH CIRCUMFLEX	LETTER, LOWERCASE
1E92	Ẓ	LATIN CAPITAL LETTER Z WITH DOT BELOW	LETTER, UPPERCASE

Codepoint(s)	Zeichen	Name	Kategorie
1E93	Ƶ	LATIN SMALL LETTER Z WITH DOT BELOW	LETTER, LOWERCASE
1E9E	ß	LATIN CAPITAL LETTER SHARP S	LETTER, UPPERCASE
1EA0	Ȧ	LATIN CAPITAL LETTER A WITH DOT BELOW	LETTER, UPPERCASE
1EA1	ȧ	LATIN SMALL LETTER A WITH DOT BELOW	LETTER, LOWERCASE
1EA2	Ȧ̃	LATIN CAPITAL LETTER A WITH HOOK ABOVE	LETTER, UPPERCASE
1EA3	ȧ̃	LATIN SMALL LETTER A WITH HOOK ABOVE	LETTER, LOWERCASE
1EA4	Ȧ̂	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE	LETTER, UPPERCASE
1EA5	ȧ̂	LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE	LETTER, LOWERCASE
1EA6	Ȧ̂̄	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE	LETTER, UPPERCASE
1EA7	ȧ̂̄	LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE	LETTER, LOWERCASE
1EAA	Ȧ̂̃	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE	LETTER, UPPERCASE
1EAB	ȧ̂̃	LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE	LETTER, LOWERCASE
1EAC	Ȧ̂̇	LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW	LETTER, UPPERCASE
1EAE	Ȧ̂̆	LATIN CAPITAL LETTER A WITH BREVE AND ACUTE	LETTER, UPPERCASE
1EAF	ȧ̂̆	LATIN SMALL LETTER A WITH BREVE AND ACUTE	LETTER, LOWERCASE
1EB0	Ȧ̂̄̆	LATIN CAPITAL LETTER A WITH BREVE AND GRAVE	LETTER, UPPERCASE
1EB1	ȧ̂̄̆	LATIN SMALL LETTER A WITH BREVE AND GRAVE	LETTER, LOWERCASE
1EB2	Ȧ̂̃̆	LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE	LETTER, UPPERCASE
1EB3	ȧ̂̃̆	LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE	LETTER, LOWERCASE
1EB4	Ȧ̂̃̇	LATIN CAPITAL LETTER A WITH BREVE AND TILDE	LETTER, UPPERCASE
1EB5	ȧ̂̃̇	LATIN SMALL LETTER A WITH BREVE AND TILDE	LETTER, LOWERCASE
1EB6	Ȧ̂̇̆	LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW	LETTER, UPPERCASE

Codepoint(s)	Zeichen	Name	Kategorie
1EB7	ă	LATIN SMALL LETTER A WITH BREVE AND DOT BELOW	LETTER, LOWERCASE
1EB8	Ě	LATIN CAPITAL LETTER E WITH DOT BELOW	LETTER, UPPERCASE
1EB9	ẹ	LATIN SMALL LETTER E WITH DOT BELOW	LETTER, LOWERCASE
1EBA	Ě̂	LATIN CAPITAL LETTER E WITH HOOK ABOVE	LETTER, UPPERCASE
1EBB	ě̂	LATIN SMALL LETTER E WITH HOOK ABOVE	LETTER, LOWERCASE
1EBC	Ě̃	LATIN CAPITAL LETTER E WITH TILDE	LETTER, UPPERCASE
1EBD	ě̃	LATIN SMALL LETTER E WITH TILDE	LETTER, LOWERCASE
1EBE	Ě́	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND ACUTE	LETTER, UPPERCASE
1EBF	ě́	LATIN SMALL LETTER E WITH CIRCUMFLEX AND ACUTE	LETTER, LOWERCASE
1EC0	Ě̀	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND GRAVE	LETTER, UPPERCASE
1EC1	ě̀	LATIN SMALL LETTER E WITH CIRCUMFLEX AND GRAVE	LETTER, LOWERCASE
1EC4	Ě̄	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND TILDE	LETTER, UPPERCASE
1EC5	ě̄	LATIN SMALL LETTER E WITH CIRCUMFLEX AND TILDE	LETTER, LOWERCASE
1EC6	Ě̇	LATIN CAPITAL LETTER E WITH CIRCUMFLEX AND DOT BELOW	LETTER, UPPERCASE
1EC7	ẹ̇	LATIN SMALL LETTER E WITH CIRCUMFLEX AND DOT BELOW	LETTER, LOWERCASE
1EC8	Ĭ	LATIN CAPITAL LETTER I WITH HOOK ABOVE	LETTER, UPPERCASE
1EC9	ĩ	LATIN SMALL LETTER I WITH HOOK ABOVE	LETTER, LOWERCASE
1ECA	ı̇	LATIN CAPITAL LETTER I WITH DOT BELOW	LETTER, UPPERCASE
1ECB	ı̇	LATIN SMALL LETTER I WITH DOT BELOW	LETTER, LOWERCASE
1ECC	Ȫ	LATIN CAPITAL LETTER O WITH DOT BELOW	LETTER, UPPERCASE
1ECD	Ƚ	LATIN SMALL LETTER O WITH DOT BELOW	LETTER, LOWERCASE
1ECE	Ȫ̂	LATIN CAPITAL LETTER O WITH HOOK ABOVE	LETTER, UPPERCASE
1ECF	Ƚ̂	LATIN SMALL LETTER O WITH HOOK ABOVE	LETTER, LOWERCASE

Codepoint(s)	Zeichen	Name	Kategorie
1ED0	Ó	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND ACUTE	LETTER, UPPERCASE
1ED1	ó	LATIN SMALL LETTER O WITH CIRCUMFLEX AND ACUTE	LETTER, LOWERCASE
1ED2	Ô	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND GRAVE	LETTER, UPPERCASE
1ED3	ô	LATIN SMALL LETTER O WITH CIRCUMFLEX AND GRAVE	LETTER, LOWERCASE
1ED6	Õ	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND TILDE	LETTER, UPPERCASE
1ED7	õ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND TILDE	LETTER, LOWERCASE
1ED8	Ȯ	LATIN CAPITAL LETTER O WITH CIRCUMFLEX AND DOT BELOW	LETTER, UPPERCASE
1ED9	ȯ	LATIN SMALL LETTER O WITH CIRCUMFLEX AND DOT BELOW	LETTER, LOWERCASE
1EDA	Ó	LATIN CAPITAL LETTER O WITH HORN AND ACUTE	LETTER, UPPERCASE
1EDB	ó	LATIN SMALL LETTER O WITH HORN AND ACUTE	LETTER, LOWERCASE
1EDC	Ò	LATIN CAPITAL LETTER O WITH HORN AND GRAVE	LETTER, UPPERCASE
1EDD	ò	LATIN SMALL LETTER O WITH HORN AND GRAVE	LETTER, LOWERCASE
1EE4	Ȯ	LATIN CAPITAL LETTER U WITH DOT BELOW	LETTER, UPPERCASE
1EE5	ȯ	LATIN SMALL LETTER U WITH DOT BELOW	LETTER, LOWERCASE
1EE6	Û	LATIN CAPITAL LETTER U WITH HOOK ABOVE	LETTER, UPPERCASE
1EE7	û	LATIN SMALL LETTER U WITH HOOK ABOVE	LETTER, LOWERCASE
1EE8	Ú	LATIN CAPITAL LETTER U WITH HORN AND ACUTE	LETTER, UPPERCASE
1EE9	ú	LATIN SMALL LETTER U WITH HORN AND ACUTE	LETTER, LOWERCASE
1EEA	Û	LATIN CAPITAL LETTER U WITH HORN AND GRAVE	LETTER, UPPERCASE
1EEB	û	LATIN SMALL LETTER U WITH HORN AND GRAVE	LETTER, LOWERCASE
1EEC	Û	LATIN CAPITAL LETTER U WITH HORN AND HOOK ABOVE	LETTER, UPPERCASE
1EED	û	LATIN SMALL LETTER U WITH HORN AND HOOK ABOVE	LETTER, LOWERCASE

Codepoint(s)	Zeichen	Name	Kategorie
1EEE	Ū	LATIN CAPITAL LETTER U WITH HORN AND TILDE	LETTER, UPPERCASE
1EEF	ū	LATIN SMALL LETTER U WITH HORN AND TILDE	LETTER, LOWERCASE
1EF0	Ụ̄	LATIN CAPITAL LETTER U WITH HORN AND DOT BELOW	LETTER, UPPERCASE
1EF1	ụ̄	LATIN SMALL LETTER U WITH HORN AND DOT BELOW	LETTER, LOWERCASE
1EF2	Ỳ	LATIN CAPITAL LETTER Y WITH GRAVE	LETTER, UPPERCASE
1EF3	ỳ	LATIN SMALL LETTER Y WITH GRAVE	LETTER, LOWERCASE
1EF4	Ỵ̀	LATIN CAPITAL LETTER Y WITH DOT BELOW	LETTER, UPPERCASE
1EF5	ỵ̀	LATIN SMALL LETTER Y WITH DOT BELOW	LETTER, LOWERCASE
1EF6	Ỳ̂	LATIN CAPITAL LETTER Y WITH HOOK ABOVE	LETTER, UPPERCASE
1EF7	ỳ̂	LATIN SMALL LETTER Y WITH HOOK ABOVE	LETTER, LOWERCASE
1EF8	Ỳ̃	LATIN CAPITAL LETTER Y WITH TILDE	LETTER, UPPERCASE
1EF9	ỳ̃	LATIN SMALL LETTER Y WITH TILDE	LETTER, LOWERCASE
20AC	€	EURO SIGN	SYMBOL, CURRENCY

2. DER DATENTYP STRING.LATIN

Der in diesem Standard beschriebene Zeichensatz kann eins-zu-eins in einem Datentyp in XML Schema dargestellt werden. Der nachfolgend beschriebene Datentyp `string.Latin` schränkt den allgemeinen Schema-Datentyp `xs:string` auf die in [Abschnitt 1 auf Seite 4](#) beschriebenen Codepoints bzw. Codepoint-Sequenzen ein. Entsprechend der Erläuterungen des vorhergehenden Kapitels erzwingt der Datentyp damit die Verwendung der Unicode-Normalform NFC.

Der Datentyp wird in Form einer XML Schema-Datei `latinchars.xsd` sowie als UML-Modell zur Verwendung in den Modellen von XÖV-Standards bereitgestellt.

2.1 Datentyp in XML Schema

Der Datentyp `String.Latin` im Namespace `http://xoev.de/latinchars/1_1/datatypes` wird durch die KoSIT unter `http://xoev.de/latinchars/1_1/datatypes/latinchars.xsd` bereitgestellt. Die Einbindung in eigene XML Schema-Dateien muss über `xs:import` erfolgen.

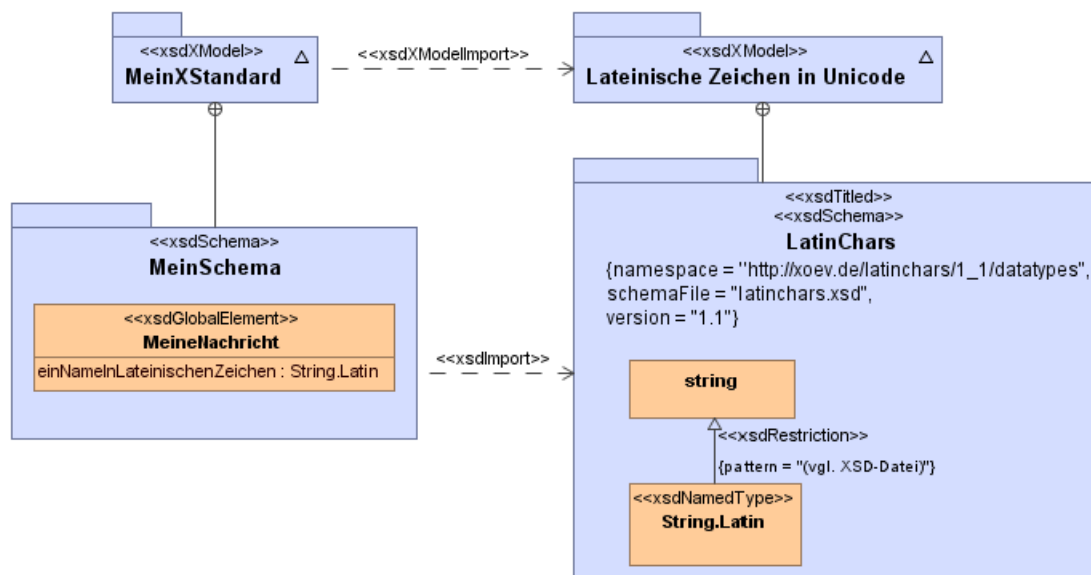
Der Datentyp ist als *simpleType* und *Restriction* von `xs:string` realisiert. Die Einschränkung auf die zulässigen Codepoints bzw. Codepoint-Sequenzen ist über ein entsprechendes *pattern* hinterlegt.

2.2 Datentyp als XÖV-UML-Modell

Zur Verwendung des Datentypen `String.Latin` in XÖV-UML-Modellen stellt die KoSIT ein entsprechendes UML-Modell bereit.

Die Einbindung des UML-Modells erfolgt nach den im [XÖV-Handbuch¹](#) beschriebenen Regeln. [Bild 2-1 auf Seite 23](#) stellt die Einbindung mit den notwendigen Stereotypen dar.

Bild 2-1 XÖV-UML-Pakete



1. www.xoev.de/sixcms/media.php/13/2010-03-02-Handbuch-final.pdf

A Die maschinenlesbare Zeichensatzdatei

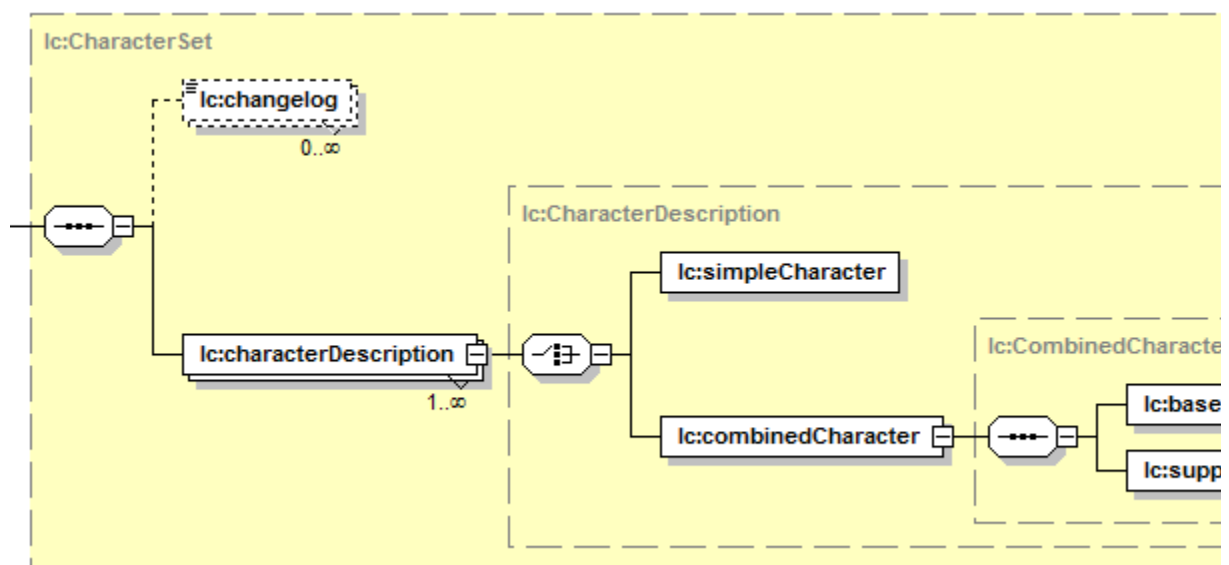
Die Datei `latinchars.xml` enthält den Zeichensatz der lateinischen Zeichen in maschinenlesbarer Form als XML-Datei. Im normativen Teil beinhaltet sie den Inhalt von [Tabelle 1-1 auf Seite 5](#). Darüber hinaus enthält sie weitere, nicht-normative Inhalte. Gegenwärtig beinhaltet der nicht-normative Teil die Abbildungsregeln von den lateinischen Zeichen in Unicode auf andere Zeichensätze gemäß der Empfehlungen der Arbeitsgruppe XInnere zu Identifikationsverfahren.

Die folgenden Abschnitte beschreiben die Struktur der Datei. Für die vollständige technische Beschreibung wird auf die Schema-Datei `charset.xsd` verwiesen.

A.1 Normativer Teil

[Bild A-1 auf Seite 24](#) zeigt den normativen Teil der Elementstruktur der Datei.

Bild A-1 Dokumentstruktur (normativer Teil)



A.1.1 Element `characterSet`

Das Wurzelement `characterSet` beschreibt alle Zeichen bzw. benannte Sequenzen des Zeichensatzes.

Eigenschaft	Art	Multiplizität
<code>changelog</code>	Element	0..*
<i>Änderungen dieser Datei gegenüber früheren Versionen sind durch <code>changeLog</code>-Einträge gekennzeichnet. Jeder <code>changeLog</code>-Eintrag enthält die Attribute <code>date</code> und <code>version</code>, sowie die Beschreibung der Änderung als XML-Content.</i>		
<code>characterDescription</code>	Element	1..*
<i>Für jedes Zeichen / jede benannte Sequenz des Zeichensatzes enthält das <code>CharacterSet</code> ein eigenes Element <code>characterDescription</code>.</i>		
<code>created</code>	Attribut	1

Eigenschaft	Art	Multiplizität
<i>Das Datum, an diese Datei erzeugt wurde.</i>		
name	Attribut	1
<i>Der Name des Standards "Lateinische Zeichen in Unicode"</i>		
version	Attribut	1
<i>Version des Standards "Lateinische Zeichen in Unicode", zu dem diese Datei gehört.</i>		

A.1.2 Element **characterDescription**

Die Beschreibung eines Zeichens / einer benannten Sequenz des Zeichensatzes.

Eigenschaft	Art	Multiplizität
simpleCharacter	Element	0..1
<i>Beschreibung des Zeichens, sofern es als eigenes Zeichen mit eigenem Codepoint in Unicode enthalten ist. Im Attribut simpleCharacter/codepoint ist der Codepoint des Zeichens alphanumerisch angegeben.</i>		
combinedCharacter	Element	0..1
<i>Beschreibung des Zeichens, falls es nicht mit eigenem Codepoint in Unicode enthalten ist, sondern als Kombination einfacher Zeichen dargestellt werden muss. In den Attributen combinedCharacter/base/codepoint und combinedCharacter/supplement/codepoint sind die Codepoint des Basiszeichens und der diakritischen Zeichens jeweils alphanumerisch angegeben.</i>		
<i>Mögliche Angaben über die Abbildung dieses Zeichens in einen anderen Zeichensatz</i>		
name	Attribut	0..1
<i>Der Name dieses Zeichens bzw. der Name dieser Named Sequence</i>		
remark	Attribut	0..1
<i>Bemerkungen zu diesem Eintrag</i>		
codepoints	Attribut	0..1
<i>Die Sequenz der alphanumerischen Codepoints zu diesem Zeichen, durch '+' getrennt. Für einen simpleCharacter entspricht dieses Attribut dem Wert in simpleCharacter/codepoint, für einen combinedCharacter der Verkettung von combinedCharacter/base/codepoint, '+' und combinedCharacter/supplement/codepoint.</i>		
blockName	Attribut	0..1
<i>Bezeichnung des Blocks (entsprechend des Unicode-Standards) zu dem dieses Zeichen gehört.</i>		
baseName	Attribut	1
<i>Bezeichnung des Basiszeichens für Buchstaben mit diakritischen Zeichen.</i>		
supplementName	Attribut	0..1
<i>Bezeichnung des Supplements (entsprechend des Unicode-Standards) zu dem dieses Zeichen gehört.</i>		
category	Attribut	0..1
<i>Bezeichnung der Kategorie (entsprechend des Unicode-Standards) zu dem dieses Zeichen gehört (z. B. LETTER).</i>		
subcategory	Attribut	0..1

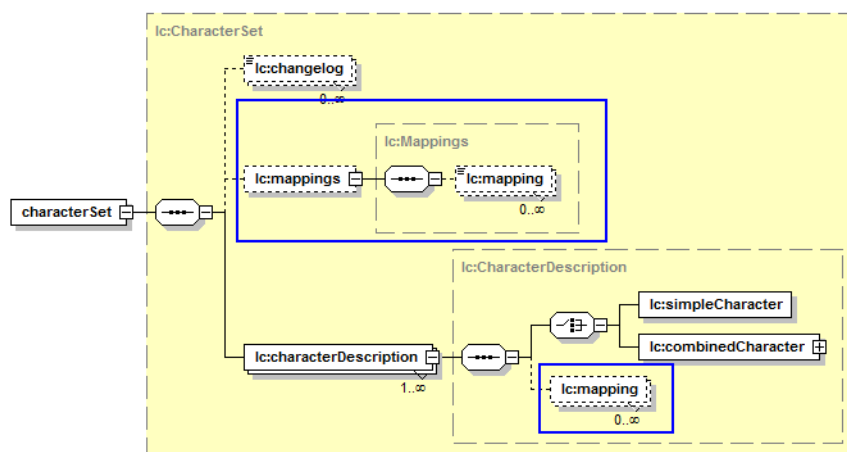
Eigenschaft	Art	Multiplizität
<i>Bezeichnung der Unterkategorie (entsprechend des Unicode-Standards) zu dem dieses Zeichen gehört (z. B. UPPERCASE)</i>		
theChars	Attribut	1
<i>Die UTF-8-Kodierung dieses Zeichens.</i>		
insert	Attribut	0..1
<i>Version des Standards, ab der dieses Zeichen aufgenommen wurde.</i>		

A.2 Nicht-normativer Teil

Bild A-2 auf Seite 26 zeigt (blau hervorgehoben) den nicht-normativen Teil der Elementstruktur der Datei.

Der nicht-normative Teil kodiert die Abbildungsregeln von den lateinischen Zeichen in Unicode auf andere Zeichensätze gemäß der Empfehlungen der Arbeitsgruppe XInneres zu Identifikationsverfahren.

Bild A-2 Dokumentstruktur (nichtnormativer Teil)



A.2.1 Element mappings (in characterSet)

In diese Element wird deklariert, für welche anderen Zeichensätze Abbildungen für die jeweiligen Zeichen dieses Zeichensatzes definiert werden.

Eigenschaft	Art	Multiplizität
codelist	Element	0..*
<i>Bezeichnung des Zeichensatzes, für den nachfolgend eine Abbildung definiert wird.</i>		

A.2.2 Element mapping (in characterDescription)

Abbildungsregel eines Zeichen / eine benannte Sequenz dieses Zeichensatzes auf ein Zeichen / eine Sequenz von Zeichen in einem anderen Zeichensatz.

Eigenschaft	Art	Multiplizität
codelist	Attribut	1
<i>Bezeichnung des Zeichensatzes, für den diese Abbildung definiert wird.</i>		
codepoints	Attribut	0..1
<i>Die Sequenz der alphanumerischen Codepoints im durch codelist bezeichneten Zeichensatz, die das aktuelle Zeichen in diesem kodieren.</i>		
transliterated	Attribut	0..1
<i>Falls dieses Element mit true angegeben ist, so gibt es keine direkte Entsprechung des Zeichens im durch codelist bezeichneten Zeichensatz, der Wert des Elements codepoints ist eine Transliteration.</i>		
theChars	Attribut	0..1
<i>Die Abbildung des Zeichens (dargestellt in UTF-8).</i>		