

ST.26 - ANNEX IV

CHARACTER SUBSET FROM THE UNICODE BASIC LATIN CODE TABLE

Final Draft

Proposal presented by the SEQL Task Force for consideration and adoption at the CWS/4

The ampersand character (0026) is only permitted as part of a predefined entity or as part of a numeric character reference (&#nnnn;). The quotation mark (0022), the apostrophe (0027), the less-than sign (003C), and the greater-than sign (003E) are not permitted and must be represented by their predefined entities.

Unicode code point	Character	Name
0020		SPACE
0021	!	EXCLAMATION MARK
0023	#	NUMBER SIGN
0024	\$	DOLLAR SIGN
0025	%	PERCENT SIGN
0026	&	AMPERSAND
0028	(LEFT PARENTHESIS
0029)	RIGHT PARENTHESIS
002A	*	ASTERISK
002B	+	PLUS SIGN
002C	,	COMMA
002D	-	HYPHEN-MINUS
002E	.	FULL STOP
002F	/	SOLIDUS
0030	0	DIGIT ZERO
0031	1	DIGIT ONE
0032	2	DIGIT TWO
0033	3	DIGIT THREE
0034	4	DIGIT FOUR
0035	5	DIGIT FIVE
0036	6	DIGIT SIX
0037	7	DIGIT SEVEN
0038	8	DIGIT EIGHT
0039	9	DIGIT NINE
003A	:	COLON
003B	;	SEMICOLON
003D	=	EQUALS SIGN
003F	?	QUESTION MARK
0040	@	COMMERCIAL AT
0041	A	LATIN CAPITAL LETTER A
0042	B	LATIN CAPITAL LETTER B
0043	C	LATIN CAPITAL LETTER C
0044	D	LATIN CAPITAL LETTER D
0045	E	LATIN CAPITAL LETTER E
0046	F	LATIN CAPITAL LETTER F
0047	G	LATIN CAPITAL LETTER G
0048	H	LATIN CAPITAL LETTER H
0049	I	LATIN CAPITAL LETTER I
004A	J	LATIN CAPITAL LETTER J
004B	K	LATIN CAPITAL LETTER K
004C	L	LATIN CAPITAL LETTER L
004D	M	LATIN CAPITAL LETTER M

Unicode code point	Character	Name
004E	N	LATIN CAPITAL LETTER N
004F	O	LATIN CAPITAL LETTER O
0050	P	LATIN CAPITAL LETTER P
0051	Q	LATIN CAPITAL LETTER Q
0052	R	LATIN CAPITAL LETTER R
0053	S	LATIN CAPITAL LETTER S
0054	T	LATIN CAPITAL LETTER T
0055	U	LATIN CAPITAL LETTER U
0056	V	LATIN CAPITAL LETTER V
0057	W	LATIN CAPITAL LETTER W
0058	X	LATIN CAPITAL LETTER X
0059	Y	LATIN CAPITAL LETTER Y
005A	Z	LATIN CAPITAL LETTER Z
005B	[LEFT SQUARE BRACKET
005C	\	REVERSE SOLIDUS
005D]	RIGHT SQUARE BRACKET
005E	^	CIRCUMFLEX ACCENT
005F	_	LOW LINE
0060	`	GRAVE ACCENT
0061	a	LATIN SMALL LETTER A
0062	b	LATIN SMALL LETTER B
0063	c	LATIN SMALL LETTER C
0064	d	LATIN SMALL LETTER D
0065	e	LATIN SMALL LETTER E
0066	f	LATIN SMALL LETTER F
0067	g	LATIN SMALL LETTER G
0068	h	LATIN SMALL LETTER H
0069	i	LATIN SMALL LETTER I
006A	j	LATIN SMALL LETTER J
006B	k	LATIN SMALL LETTER K
006C	l	LATIN SMALL LETTER L
006D	m	LATIN SMALL LETTER M
006E	n	LATIN SMALL LETTER N
006F	o	LATIN SMALL LETTER O
0070	p	LATIN SMALL LETTER P
0071	q	LATIN SMALL LETTER Q
0072	r	LATIN SMALL LETTER R
0073	s	LATIN SMALL LETTER S
0074	t	LATIN SMALL LETTER T
0075	u	LATIN SMALL LETTER U
0076	v	LATIN SMALL LETTER V
0077	w	LATIN SMALL LETTER W
0078	x	LATIN SMALL LETTER X
0079	y	LATIN SMALL LETTER Y
007A	z	LATIN SMALL LETTER Z
007B	{	LEFT CURLY BRACKET
007C		VERTICAL LINE
007D	}	RIGHT CURLY BRACKET
007E	~	TILDE

[Annex V to ST.26 follows]

ST.26 - ANNEX V

ADDITIONAL DATA EXCHANGE REQUIREMENTS (FOR PATENT OFFICES ONLY)

Final Draft

Proposal presented by the SEQL Task Force for consideration and adoption at the CWS/4

In the context of data exchange with database providers (INSID members), the Patent Offices should populate for each sequence the element `INSDSeq_other-seqids` with one `INSDSeqid` containing a reference to the corresponding published patent and the sequence identification number in the following format:

`pat|{office code}|{publication number}|{document kind code}|{sequence identification number}`

where office code is the code of the IP office publishing the patent document as set forth in ST.3; document kind code is the code for the identification of different kinds of patent documents as set forth in ST.16; publication number is the publication number of the application or patent; and Sequence identification number is the number of the sequence in that application or patent.

Example:

`pat|WO|2013999999|A1|123456`

Which would be translated into a valid XML instance as:

```
<INSDSeq_other-seqids>  
  < INSDSeqid>pat |WO| 2013999999 | A1 | 123456</INSDSeqid>  
</INSDSeq_other-seqids>
```

Where "123456" is the 123456th sequence from the WO publication no. 2013999999 (A1).

[End of Annex II and of document]