

**Proposal to identify the Lithuanian Alphabet
as a Collection in the ISO/IEC 10646,
including the named sequences for the accented letters
that have no pre-composed form of encoding (also in TUS)**

Expert contribution by: Erkki I. Kolehmainen (member of CEN/ISSS CDFG),
based on input from the Lithuanian Language Institute
Affiliation: RILF, representing SFS; also a Liaison to Unicode
Date: 12 April 2006

1 Lithuanian Lettering

Lithuanian is a language of the Indo-European language group. Lithuanian is also one of the oldest living languages. Presently it is used by approximately 4 million people (about 3.5 million in Lithuania and about 0.5 million elsewhere, mostly in the USA). Lithuanian is the only language that has an official language status in the Republic of Lithuania, a Member of the European Union.

In 1547 the first Lithuanian book was published in Königsberg. Lithuanian language orthographic norms were finally worked out in the beginning of the XX century together with the restoration of Lithuanian statehood in 1918. The contemporary Lithuanian main alphabet is based on Latin (excluding Q, W and X) along with extra 18 letters with diacritics: A, Č, E, È, Į, Š, Ū, Ū, Ž (9 capital and 9 small). Due to the complicated phonetics of the Lithuanian language, the position of the stress (accent) in a word is indicated with the diacritics: grave, tilde and acute. Therefore, in addition to the 64 letters (32 capital and 32 small), 68 additional accented letters (34 capital and 34 small) are used in contemporary Lithuanian (see ADDENDUM I). The word is written with additional accented letters in order to: a) indicate the proper pronunciation of the word; b) distinguish the ambiguous words, which have the same basic spelling, but different pronunciation and meaning. Accented letters are essential for publishing Lithuanian schoolbooks, educational textbooks, dictionaries, linguistic publications, official and legal documents as well as for terminological data bases and web. Accented letters are used in media to avoid ambiguity in expressions; together with non-accented letters they comprise the integral part of the writing system of contemporary Lithuanian.

2 Lithuanian Letters and the ISO/IEC 10646 (and TUS)

During World War II Lithuania was annexed by the Soviet Union up until 1990. As a consequence, Lithuania's own national interests and cultural needs were not properly represented in international organisations and in the evolving world of information technology (IT).

The Lithuanian National Body has earlier submitted a request to encode all of the Lithuanian accented letters as pre-composed in the ISO/IEC 10646 and the Unicode Standard. This request, however, was rejected as conflicting with the then established normalization scheme, and since all the characters can be encoded as decomposed.

The issue was discussed at the last CEN/ISSS CDFG (Cultural Diversity Focus Group) meeting in Sophia Antipolis on 16-17 March 2006, where an action item was recorded for an expedient registration of the Lithuanian characters to facilitate a formal reference for support requirements, be it for input, processing, or rendering. **Thus, the proposal is to include such a registration already in the forthcoming Amendment 3.**

Note. There are three national 8-bit single-byte code tables in Lithuania for encoding of accented letters adopted as national standards: 1) the basic code table; 2) code table for Windows with some extra symbols in 8 and 9 columns; 3) code table for DOS with some extra drawing symbols. The basic code table is shown in Addendum III.

ADDENDUM I

The Lithuanian letter repertoire. The accented letters have a colored background. The letters identified by named sequences are numbered (corresponding to Addendum II).

A 0041	Ą 0104	À 00C0	Á 00C1	Ã 00C3	#1	#3			
a 0061	ą 0105	à 00E0	á 00E1	ã 00E3	#2	#4			
B 0042									
b 0062									
C 0043	Č 010C								
c 0063	č 010D								
D 0044									
d 0064									
E 0045	Ę 0118	È 0116	È 00C8	É 00C9	Ë 1EBC	#5	#7	#9	#11
e 0065	ę 0119	è 0117	è 00E8	é 00E9	ë 1EBD	#6	#8	#10	#12
F 0046									
f 0066									
G 0047									
g 0067									
H 0048									
h 0068									
I 0049	Į 012E	Ì 00CC	Í 00CD	Ï 0128	#16	#18			
I 0069	į 012F	#13	#14	#15	#17	#19			
Y 0059	Ý 00DD	ÿ 1EF8							
y 0079	ý 00FD	ÿ 1EF9							
J 004A	#20								
j 006A	#21								
K 004B									
k 006B									
L 004C	#22								
l 006C	#23								
M 004D	#24								
m 006D	#25								

N 004E	Ñ 00D1								
n 006E	ñ 00F1								
O 004F	Ò 00D2	Ó 00D3	Õ 00D5						
o 006F	ò 00F2	ó 00F3	õ 00F5						
P 0050									
p 0070									
R 0052	#26								
r 0072	#27								
S 0053	Š 0160								
s 0073	š 0161								
T 0054									
t 0074									
U 0055	Ů 0172	Ū 016A	Ù 00D9	Ú 00DA	Û 0168	#28	#30	#32	#34
u 0075	ů 0173	ū 016B	ù 00F9	ú 00FA	û 0169	#29	#31	#33	#35
V 0056									
v 0076									
Z 005A	Ž 017D								
z 007A	ž 017E								

ADDENDUM II

The Lithuanian letters identified by named sequences

#	Code Points	Name (Named character sequence)
1	0104 0301	LATIN CAPITAL LETTER A WITH OGONEK AND ACUTE
2	0105 0301	LATIN SMALL LETTER A WITH OGONEK AND ACUTE
3	0104 0303	LATIN CAPITAL LETTER A WITH OGONEK AND TILDE
4	0105 0303	LATIN SMALL LETTER A WITH OGONEK AND TILDE
5	0118 0301	LATIN CAPITAL LETTER E WITH OGONEK AND ACUTE
6	0119 0301	LATIN SMALL LETTER E WITH OGONEK AND ACUTE
7	0118 0303	LATIN CAPITAL LETTER E WITH OGONEK AND TILDE
8	0119 0303	LATIN SMALL LETTER E WITH OGONEK AND TILDE
9	0116 0301	LATIN CAPITAL LETTER E WITH DOT ABOVE AND ACUTE
10	0117 0301	LATIN SMALL LETTER E WITH DOT ABOVE AND ACUTE
11	0116 0303	LATIN CAPITAL LETTER E WITH DOT ABOVE AND TILDE
12	0117 0303	LATIN SMALL LETTER E WITH DOT ABOVE AND TILDE
13	0069 0307 0300	LATIN SMALL LETTER I WITH DOT ABOVE AND GRAVE
14	0069 0307 0301	LATIN SMALL LETTER I WITH DOT ABOVE AND ACUTE
15	0069 0307 0303	LATIN SMALL LETTER I WITH DOT ABOVE AND TILDE
16	012E 0301	LATIN CAPITAL LETTER I WITH OGONEK AND ACUTE
17	012F 0307 0301	LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND ACUTE
18	012E 0303	LATIN CAPITAL LETTER I WITH OGONEK AND TILDE
19	012F 0307 0303	LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND TILDE
20	004A 0303	LATIN CAPITAL LETTER J WITH TILDE
21	006A 0307 0303	LATIN SMALL LETTER J WITH DOT ABOVE AND TILDE
22	004C 0303	LATIN CAPITAL LETTER L WITH TILDE
23	006C 0303	LATIN SMALL LETTER L WITH TILDE
24	004D 0303	LATIN CAPITAL LETTER M WITH TILDE
25	006D 0303	LATIN SMALL LETTER M WITH TILDE
26	0052 0303	LATIN CAPITAL LETTER R WITH TILDE
27	0072 0303	LATIN SMALL LETTER R WITH TILDE
28	0172 0301	LATIN CAPITAL LETTER U WITH OGONEK AND ACUTE
29	0173 0301	LATIN SMALL LETTER U WITH OGONEK AND ACUTE
30	0172 0303	LATIN CAPITAL LETTER U WITH OGONEK AND TILDE
31	0173 0303	LATIN SMALL LETTER U WITH OGONEK AND TILDE
32	016A 0301	LATIN CAPITAL LETTER U WITH MACRON AND ACUTE
33	016B 0301	LATIN SMALL LETTER U WITH MACRON AND ACUTE
34	016A 0303	LATIN CAPITAL LETTER U WITH MACRON AND TILDE
35	016B 0303	LATIN SMALL LETTER U WITH MACRON AND TILDE

ADDENDUM III

Code table from Lithuanian Standard LST 1564:2000 *Information technology – 8-bit single-byte character coding – Lithuanian accented letters.*

X \ Y	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0	@	P	`	p			NBSP	ĩ	Ą	Š	ą	š
1			!	1	A	Q	a	q			Ą	ą	Į	į	ı	ĩ
2			"	2	B	R	b	r			É	é	À	Ò	à	ò
3			#	3	C	S	c	s			Ë	ë	Á	Ó	á	ó
4			\$	4	D	T	d	t			ĩ	'	Ä	Ý	ä	ý
5			%	5	E	U	e	u			Ĺ	ĺ	Ã	Õ	ã	õ
6			&	6	F	V	f	v			Ń	ņ	Ę	Ö	ę	ö
7			'	7	G	W	g	w			ĩ	ĵ	Ą	Ū	ą	ũ
8			(8	H	X	h	x			Ě	é	Č	U	č	u
9)	9	I	Y	i	y			Ñ	ñ	É	Ù	é	ù
A			*	:	J	Z	j	z			Ě	ě	È	Ú	è	ú
B			+	;	K	[k	{			Ŕ	ř	È	Ū	è	ū
C			,	<	L	\	l				Ú	ú	Ě	Ü	ě	ü
D			-	=	M]	m	}			SHY	Ū	Ì	Ÿ	ì	ÿ
E			.	>	N	^	n	~			Ŭ	ũ	Í	Ž	í	ž
F			/	?	O	_	o				Ú	ú	Į	β	į	ũ