

## Specifics of Lithuanian language localization for mobile phones

### *Guidelines for implementors*

#### *Lithuanian Mobile Phone User Group*

## Contents

Scope.....	1
User interface translation.....	1
Keypad layout.....	1
12-key keypads.....	1
Full size keypads.....	3
Language specific special characters.....	3
Ordering of lists.....	3
Predictive text input.....	6
SMS messages and Lithuanian letter repertoire.....	6
Support of legacy character encodings for web browsing.....	6
Support of legacy character encodings for email reading.....	6

## Scope

There are few cases that need to be addressed for a phone to be usable with specific language:

1. User interface (i.e. menu) language translation.
2. Keypad layout allowing to input letters and special characters specific for particular language.
3. Ordering of lists (e.g. phone book) should follow national alphabet.
4. Predictive text input dictionary (e.g. T9) should be available.
5. SMS should be sent using full letter repertoire for a specific language
6. Support of legacy character encodings for web browsing.
7. Support of legacy character encodings for email reading.

## User interface translation

This issue is usually already addressed, i.e. all phones that are selling in Lithuania have menu translated. However it should be seen as an obvious issue that these translations should be tested, as translating of separate words out of context doesn't result in usable translation.

## Keypad layout

### **12-key keypads**

ETSI ES 202 130 document describes keyboard layouts for 12-key keypads for European languages. However the layout described in that document isn't suitable for Lithuanian language because letters with diacritical marks are assigned to the end of key-press sequence. This breaks expected Lithuanian alphabet order for a user and makes it difficult to reach and use these letters. Occurrence of these letters is rather high in Lithuanian language text.

As for the access for visually impaired users, it should be noted that most telephones allow to reach numbers by pressing and holding a key once and using key-press sequence to reach a number

isn't necessary.

Following standard alphabet order for key-press sequences allows for users to find a letter in expected place and provides better user experience. For example, alphabet order for key-press sequences is currently used by Nokia® phones for Lithuanian language input, by Sony-Ericsson® phones for Polish language input and so on.

### Lithuanian 12-key keypad alphabetic assignment

Key tag	Letter	ISO/IEC 10646 identifier	ISO/IEC 10646 name
2aąbcč	a	U+0061	LATIN SMALL LETTER A
	ą	U+0105	LATIN SMALL LETTER A WITH OGONEK
	b	U+0062	LATIN SMALL LETTER B
	c	U+0063	LATIN SMALL LETTER C
	č	U+010D	LATIN SMALL LETTER C WITH CARON
	2	U+0032	DIGIT TWO
3deėėf	d	U+0064	LATIN SMALL LETTER D
	e	U+0065	LATIN SMALL LETTER E
	ė	U+0119	LATIN SMALL LETTER E WITH OGONEK
	ė	U+0117	LATIN SMALL LETTER E WITH DOT ABOVE
	f	U+0066	LATIN SMALL LETTER F
	3	U+0033	DIGIT THREE
4ghiiy	g	U+0067	LATIN SMALL LETTER G
	h	U+0068	LATIN SMALL LETTER H
	i	U+0069	LATIN SMALL LETTER I
	į	U+012F	LATIN SMALL LETTER I WITH OGONEK
	y	U+0079	LATIN SMALL LETTER Y
	4	U+0034	DIGIT FOUR
5jkl	j	U+006A	LATIN SMALL LETTER J
	k	U+006B	LATIN SMALL LETTER K
	l	U+006C	LATIN SMALL LETTER L
	5	U+0035	DIGIT FIVE
6mno	m	U+006D	LATIN SMALL LETTER M
	n	U+006E	LATIN SMALL LETTER N
	o	U+006F	LATIN SMALL LETTER O
	6	U+0036	DIGIT SIX
7pqrsš	p	U+0070	LATIN SMALL LETTER P
	q	U+0071	LATIN SMALL LETTER Q
	r	U+0072	LATIN SMALL LETTER R
	s	U+0073	LATIN SMALL LETTER S
	š	U+0161	LATIN SMALL LETTER S WITH CARON
	7	U+0037	DIGIT SEVEN
8tuųūv	t	U+0074	LATIN SMALL LETTER T
	u	U+0075	LATIN SMALL LETTER U
	ų	U+0173	LATIN SMALL LETTER U WITH OGONEK
	ū	U+016B	LATIN SMALL LETTER U WITH MACRON
	v	U+0076	LATIN SMALL LETTER V
	8	U+0038	DIGIT EIGHT
9wxyzž	w	U+0077	LATIN SMALL LETTER W
	x	U+0078	LATIN SMALL LETTER X
	y	U+0079	LATIN SMALL LETTER Y
	z	U+007A	LATIN SMALL LETTER Z
	ž	U+017E	LATIN SMALL LETTER Z WITH CARON
	9	U+0039	DIGIT NINE

Note: Ideally keys should have tags as noted above, especially in case of soft keys on a visual display. However adopting key tags for hardware keys isn't high priority issue.

There may be additional letters and symbols in the key sequences following numbers. Full German, Polish, Latvian, Scandinavian language letter repertoires are of particular interest to Lithuanian users. This is different from minority languages as outlined in ETSI ES 202 130 as national minority users may choose their own language input method, and letters after numbers isn't adequate input method for all languages/scripts.

### **Full size keypads**

Keyboard layouts used by Windows® operating system or X.Org® software for computers usually assume more keys than typical mobile device keyboard. Keyboards having too few keys may use dead keys. For example, one vendor has chosen to use Chr as dead key and multiple key-presses of non-accented letter key to input accented letters.

Link to standard AŽERTY computer keyboard layout description:

<http://aldona.mii.lt/pms/kalba/klav/lithkeyb4.html>

Link to other popular computer keyboard layouts: <http://webcvs.freedesktop.org/xkeyboard-config/xkeyboard-config/symbols/lt?revision=1.9&view=markup>

More details on Lithuanian layout for full size mobile device keyboards will be provided in separate document or in the next version of this document.

### **Language specific special characters**

ETSI ES 202 130 doesn't cover these characters. U+201E is used as opening quote in Lithuanian language. U+201C is used as closing quote (right, not left). U+0300, U+0301, U+0303 combining accents are rarely used and may not be relevant for telephone users at the time of this writing, though the situation may change in the future with more advanced telephones.

<b>Character</b>	<b>ISO/IEC 10646 identifier</b>	<b>ISO/IEC 10646 name</b>
”	U+201E	DOUBLE LOW-9 QUOTATION MARK
“	U+201C	LEFT DOUBLE QUOTATION MARK
–	U+2013	EN DASH
	U+0300	COMBINING GRAVE ACCENT
	U+0301	COMBINING ACUTE ACCENT
	U+0303	COMBINING TILDE

### **Ordering of lists**

Ordering of lists (e.g. phone book) should follow national alphabetic ordering and not general ordering.

#### **Lithuanian letter repertoire listed in Lithuanian alphabetic order**

<b>Letter</b>	<b>ISO/IEC 10646 identifier</b>	<b>ISO/IEC 10646 name</b>	<b>Usage type</b>	<b>Notes</b>
A	U+0041	LATIN CAPITAL LETTER A	A	*
a	U+0061	LATIN SMALL LETTER A	A	*

Ą	U+0104	LATIN CAPITAL LETTER A WITH OGONEK	A	*
ą	U+0105	LATIN SMALL LETTER A WITH OGONEK	A	*
B	U+0042	LATIN CAPITAL LETTER B	A	
b	U+0062	LATIN SMALL LETTER B	A	
C	U+0043	LATIN CAPITAL LETTER C	A	
c	U+0063	LATIN SMALL LETTER C	A	
Č	U+010C	LATIN CAPITAL LETTER C WITH CARON	A	
č	U+010D	LATIN SMALL LETTER C WITH CARON	A	
D	U+0044	LATIN CAPITAL LETTER D	A	
d	U+0064	LATIN SMALL LETTER D	A	
E	U+0045	LATIN CAPITAL LETTER E	A	*
e	U+0065	LATIN SMALL LETTER E	A	*
Ę	U+0118	LATIN CAPITAL LETTER E WITH OGONEK	A	*
ę	U+0119	LATIN SMALL LETTER E WITH OGONEK	A	*
Ě	U+0116	LATIN CAPITAL LETTER E WITH DOT ABOVE	A	*
ě	U+0117	LATIN SMALL LETTER E WITH DOT ABOVE	A	*
F	U+0046	LATIN CAPITAL LETTER F	A	
f	U+0066	LATIN SMALL LETTER F	A	
G	U+0047	LATIN CAPITAL LETTER G	A	
g	U+0067	LATIN SMALL LETTER G	A	
H	U+0048	LATIN CAPITAL LETTER H	A	
h	U+0068	LATIN SMALL LETTER H	A	
I	U+0049	LATIN CAPITAL LETTER I	A	*
i	U+0069	LATIN SMALL LETTER I	A	*
Ĳ	U+012E	LATIN CAPITAL LETTER I WITH OGONEK	A	*
į	U+012F	LATIN SMALL LETTER I WITH OGONEK	A	*
Y	U+0059	LATIN CAPITAL LETTER Y	A	*
y	U+0079	LATIN SMALL LETTER Y	A	*
J	U+004A	LATIN CAPITAL LETTER J	A	
j	U+006A	LATIN SMALL LETTER J	A	
K	U+004B	LATIN CAPITAL LETTER K	A	
k	U+006B	LATIN SMALL LETTER K	A	
L	U+004C	LATIN CAPITAL LETTER L	A	
l	U+006C	LATIN SMALL LETTER L	A	
M	U+004D	LATIN CAPITAL LETTER M	A	
m	U+006D	LATIN SMALL LETTER M	A	
N	U+004E	LATIN CAPITAL LETTER N	A	

n	U+006E	LATIN SMALL LETTER N	A	
O	U+004F	LATIN CAPITAL LETTER O	A	
o	U+006F	LATIN SMALL LETTER O	A	
P	U+0050	LATIN CAPITAL LETTER P	A	
p	U+0070	LATIN SMALL LETTER P	A	
Q	U+0051	LATIN CAPITAL LETTER Q	B	
q	U+0071	LATIN SMALL LETTER Q	B	
R	U+0052	LATIN CAPITAL LETTER R	A	
r	U+0072	LATIN SMALL LETTER R	A	
S	U+0053	LATIN CAPITAL LETTER S	A	
s	U+0073	LATIN SMALL LETTER S	A	
Š	U+0160	LATIN CAPITAL LETTER S WITH CARON	A	
š	U+0161	LATIN SMALL LETTER S WITH CARON	A	
T	U+0054	LATIN CAPITAL LETTER T	A	
t	U+0074	LATIN SMALL LETTER T	A	
U	U+0055	LATIN CAPITAL LETTER U	A	*
u	U+0075	LATIN SMALL LETTER U	A	*
Ų	U+0172	LATIN CAPITAL LETTER U WITH OGONEK	A	*
ų	U+0173	LATIN SMALL LETTER U WITH OGONEK	A	*
Ū	U+016A	LATIN CAPITAL LETTER U WITH MACRON	A	*
ū	U+016B	LATIN SMALL LETTER U WITH MACRON	A	*
V	U+0056	LATIN CAPITAL LETTER V	A	
v	U+0076	LATIN SMALL LETTER V	A	
W	U+0057	LATIN CAPITAL LETTER W	B	
w	U+0077	LATIN SMALL LETTER W	B	
X	U+0058	LATIN CAPITAL LETTER X	B	
x	U+0078	LATIN SMALL LETTER X	B	
Z	U+005A	LATIN CAPITAL LETTER Z	A	
z	U+007A	LATIN SMALL LETTER Z	A	
Ž	U+017D	LATIN CAPITAL LETTER Z WITH CARON	A	
ž	U+017E	LATIN SMALL LETTER Z WITH CARON	A	

Ordering is defined by Lithuanian standard LST 1285:1993.

Adjacent letters marked with \* are considered equivalent in ordering.

The same letters of different case are equivalent in ordering too.

In case of words differing in case or accents only (marked with \*) order in the table above may be used as secondary weight.

Usage type A means Lithuanian letters. Usage type B means necessary foreign letters frequently used for foreign words and names.

## **Predictive text input**

There is T9 predictive text input dictionary for Lithuanian and it should be installed and available for telephones localized for Lithuanian market.

## **SMS messages and Lithuanian letter repertoire**

Using GSM Phase 2+ with ISO/IEC 10464 encoding is required to provide full Lithuanian letter repertoire for SMS messages. As a result message fragment limit is reduced and message cost for user may increase if operator charges fixed price per message fragment. Some very old telephones may have problems displaying ISO/IEC 10464 SMS. Because of all that some users may wish to send SMS messages without diacritical marks.

Some vendors have chosen to provide a user setting to change accented letters to corresponding ones without diacritical marks automatically when sending out Latin script only SMS messages. It should be noted that default setting should be to send SMS messages as is, or a warning should be given to a user that such text altering is taking place. In any case the text altering should not be performed if a user has no option to prevent it.

This isn't relevant to MMS or e-mail messages as MMS and e-mail have few problems using full character repertoire.

## **Support of legacy character encodings for web browsing**

Beside UTF-8, Windows-1257 character encoding should be supported for phones having TCP/IP and HTML/XHTML browser as this encoding is still widely used on the web at the time of this writing.

## **Support of legacy character encodings for email reading**

Beside UTF-8 these character encodings should be supported for e-mail reading for phones having email client: Windows-1257, ISO 8859-13 and ISO 8859-4. There are chances that user may receive e-mail in one of these character encodings in Lithuanian language.

Nokia® is trademark of Nokia. Sony-Ericsson® is trademark of Sony-Ericsson. Windows® is trademark of Microsoft corporation. X.Org® is trademark of X.Org foundation. All other trademarks mentioned in this document are the property of their respective owners