Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document DRAFT 3 – 2011-12-05
Title: Proposal to add Lithuanian accented letters to the UCS
Source: Vilnius University, the State Commission of the Lithuanian language, Lithuanian Standards Board
Status: Lithuanian Standards Board Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2011-XX-XX

1. Introduction

The alphabet of standard modern Lithuanian contains only nine letters with diacritical marks: Ąą, Čč, Ęę, Ėė, Įį, Šš, Ųų, Ūū, Žž. In addition to these, there are optional **accent marks**, which signify word stress together with its syllabic intonation (in a manner similar to that of the accent signs in ancient Greek). Although most of ordinary Lithuanian writing is done without using stress marks – in fact with the nine letters with phonetic diacritics shown above – the stress signs are used in dictionaries, grammars, manuals for language study, and other such materials. Even in the ordinary texts optional accent marks can be, and are, occasionally used to prevent ambiguity.

In standard modern Lithuanian there are three accent (stress) signs, which look like those in ancient (polytonic) Greek: **acute** (´), **grave** (`), and **circumflex** (˜). Like in polytonic Greek, the circumflex has a wavy shape, so it looks identical to Latin sign called tilde. A short stressed syllable is marked with the grave sign; long stressed syllables can have either acute or circumflex tone.

When stress marks fall on ordinary Latin letters, this is not much of a problem, as most of these composite characters are provided by the standard Unicode sets: á, à, ã, é, è, ẽ, etc. However, when the stress marks fall on letters already having a phonetic or etymological diacritic of their own (ą̃, ū́, ė́, etc.), this becomes a problem, as such composite characters are not provided by Unicode. There are certain orthographic rules about the use of the accent signs: when a diphthong is stressed, the acute is usually written on the first letter of the diphthong, and the circumflex on the second. The second component of a diphthong is not necessarily a vowel, but may also be a resonant consonant (l, m, n, and r). This rule gives us some more letter-accent combinations not provided by Unicode: l̃, m̃, r̃ (the letter ñ, luckily, already exists). There are some rare cases when the orthographically written letter j functions as the second component of a circumflexed diphthong, so some scholars have proposed also the sign j̃. In addition, a few scholars have expressed the request that the accent sign on the letter i should not cancel out the letter's dot, so that the ordinary characters ì, í, ĩ have been deemed insufficient, and special variants of accented i's with both the dot and the accent have to be introduced.

The problem of already accented letters receiving stress marks is comparable to that in German, which uses only Ää, Öö, Üü, and ß; but in some special texts, such as dictionaries, there is the need to distinguish, say, the words 'übersétzen' (to translate) and 'ǘbersetzen' (to transfer), by adding the acute-shaped stress mark. This effectively brings into existence a wholly new character, ǘ, which is part of the Unicode standard. The principle is similar in Lithuanian, but here we have more vowels with diacritic marks than in German, and more stress marks, and perhaps the need to use stress marks is more frequent. As another analogy, although there are only 24 letters in the Greek alphabet, the number of letter-accent combinations in the Unicode Greek Extended range, for polytonic Greek, is well over two hundred.

All in all, the Lithuanian stress marks give us 68 letters (including lower and upper case), 35 of them are not included in the Unicode. So, there is an urgent need to include them in the UCS.

2. Proposed characters

Assumed code point	Symbol	Name
U+HH00	Ą́	LATIN CAPITAL LETTER A WITH OGONEK AND ACUTE
U+HH01	ą́	LATIN SMALL LETTER A WITH OGONEK AND ACUTE
U+HH02	Ą̃	LATIN CAPITAL LETTER A WITH OGONEK AND TILDE
U+HH03	ą̃	LATIN SMALL LETTER A WITH OGONEK AND TILDE
U+HH04	Ę́	LATIN CAPITAL LETTER E WITH OGONEK AND ACUTE
U+HH05	ę́	LATIN SMALL LETTER E WITH OGONEK AND ACUTE
U+HH06	Ę̃	LATIN CAPITAL LETTER E WITH OGONEK AND TILDE
U+HH07	ę̃	LATIN SMALL LETTER E WITH OGONEK AND TILDE
U+HH08	Ė́	LATIN CAPITAL LETTER E WITH DOT ABOVE AND ACUTE
U+HH09	ė́	LATIN SMALL LETTER E WITH DOT ABOVE AND ACUTE
U+HH0A	Ė̃	LATIN CAPITAL LETTER E WITH DOT ABOVE AND TILDE
U+HH0B	ė̃	LATIN SMALL LETTER E WITH DOT ABOVE AND TILDE
U+HH0C	i̇̀	LATIN SMALL LETTER I WITH DOT ABOVE AND GRAVE
U+HH0D	i̇́	LATIN SMALL LETTER I WITH DOT ABOVE AND ACUTE
U+HH0E	i̇̃	LATIN SMALL LETTER I WITH DOT ABOVE AND TILDE
U+HH0F	Į́	LATIN CAPITAL LETTER I WITH OGONEK AND ACUTE
U+HH10	į̇́	LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND ACUTE
U+HH11	Į̃	LATIN CAPITAL LETTER I WITH OGONEK AND TILDE
U+HH12	į̇̃	LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND TILDE
U+HH13	J̃	LATIN CAPITAL LETTER J WITH TILDE
U+HH14	j̇̃	LATIN SMALL LETTER J WITH DOT ABOVE AND TILDE
U+HH15	L̃	LATIN CAPITAL LETTER L WITH TILDE
U+HH16	l̃	LATIN SMALL LETTER L WITH TILDE
U+HH17	M̃	LATIN CAPITAL LETTER M WITH TILDE
U+HH18	m̃	LATIN SMALL LETTER M WITH TILDE
U+HH19	R̃	LATIN CAPITAL LETTER R WITH TILDE
U+HH1A	r̃	LATIN SMALL LETTER R WITH TILDE
U+HH1B	Ų́	LATIN CAPITAL LETTER U WITH OGONEK AND ACUTE
U+HH1C	ų́	LATIN SMALL LETTER U WITH OGONEK AND ACUTE
U+HH1D	Ų̃	LATIN CAPITAL LETTER U WITH OGONEK AND TILDE
U+HH1E	ų̃	LATIN SMALL LETTER U WITH OGONEK AND TILDE
U+HH1F	Ū́	LATIN CAPITAL LETTER U WITH MACRON AND ACUTE
U+HH20	ū́	LATIN SMALL LETTER U WITH MACRON AND ACUTE
U+HH21	Ū̃	LATIN CAPITAL LETTER U WITH MACRON AND TILDE
U+HH22	ū̃	LATIN SMALL LETTER U WITH MACRON AND TILDE

Properties:

```
U+HH00;LATIN CAPITAL LETTER A WITH OGONEK AND ACUTE;Lu;0;L;0104
     0301;;;;N;;;;U+HH01;
U+HH01;LATIN SMALL LETTER A WITH OGONEK AND ACUTE;Ll;0;L;0105
     0301;;;;N;;;U+HH00;;U+HH00
U+HH02;LATIN CAPITAL LETTER A WITH OGONEK AND TILDE;Lu;0;L;0104
     0303;;;;N;;;;U+HH03;
U+HH03;LATIN SMALL LETTER A WITH OGONEK AND TILDE;Ll;0;L;0105
     0303;;;;N;;;U+HH02;;U+HH02
U+HH04;LATIN CAPITAL LETTER E WITH OGONEK AND ACUTE;Lu;0;L;0118
     0301;;;;N;;;;U+HH05;
U+HH05;LATIN SMALL LETTER E WITH OGONEK AND ACUTE;Ll;0;L;0119
     0301;;;;N;;;U+HH04;;U+HH04
U+HH06;LATIN CAPITAL LETTER E WITH OGONEK AND TILDE;Lu;0;L;0118
     0303;;;;N;;;;U+HH07;
U+HH07;LATIN SMALL LETTER E WITH OGONEK AND TILDE;Ll;0;L;0119
     0303;;;;N;;;U+HH06;;U+HH06
U+HH08;LATIN CAPITAL LETTER E WITH DOT ABOVE AND ACUTE;Lu;0;L;0116
     0301;;;;N;;;;U+HH09;
U+HH09;LATIN SMALL LETTER E WITH DOT ABOVE AND ACUTE;Ll;0;L;0117
     0301;;;;N;;;U+HH08;;U+HH08
U+HH0A;LATIN CAPITAL LETTER E WITH DOT ABOVE AND TILDE;Lu;0;L;0116
     0303;;;;N;;;;U+HH0B;
U+HH0B;LATIN SMALL LETTER E WITH DOT ABOVE AND TILDE;Ll;0;L;0117
     0303;;;;N;;;U+HH0A;;U+HH0A
U+HH0C;LATIN SMALL LETTER I WITH DOT ABOVE AND GRAVE;Ll;0;L;0069 0307
     0300;;;;N;;;00CC;; 00CC
U+HH0D;LATIN SMALL LETTER I WITH DOT ABOVE AND ACUTE;Ll;0;L;0069 0307
     0301;;;;N;;;00CD;;00CD
U+HH0E;LATIN SMALL LETTER I WITH DOT ABOVE AND TILDE;Ll;0;L;0069 0307
     0303;;;;N;;;0128;;0128
U+HH0F;LATIN CAPITAL LETTER I WITH OGONEK AND ACUTE;Lu;0;L;012E
     0301;;;;N;;;;U+HH10;
U+HH10;LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND ACUTE;Ll;0;L;012F
     0307 0301;;;;N;;;U+HH0F;;U+HH0F
U+HH11;LATIN CAPITAL LETTER I WITH OGONEK AND TILDE;Lu;0;L;012E
     0303;;;;N;;;;U+HH12;
U+HH12;LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND TILDE;Ll;0;L;012F
     0307 0303;;;;N;;;U+HH11;;U+HH11
U+HH13;LATIN CAPITAL LETTER J WITH TILDE;Lu;0;L;004A 0303;;;;N;;;;U+HH14;
U+HH14;LATIN SMALL LETTER J WITH DOT ABOVE AND TILDE;Ll;0;L;006A 0307
     0303;;;;N;;;U+HH13;;U+HH13
U+HH15;LATIN CAPITAL LETTER L WITH TILDE;Lu;0;L;004C 0303;;;;N;;;;U+HH16;
U+HH16;LATIN SMALL LETTER L WITH TILDE;Ll;0;L;006C
     0303;;;;N;;;U+HH15;;U+HH15
U+HH17;LATIN CAPITAL LETTER M WITH TILDE;Lu;0;L;004D 0303;;;;N;;;;U+HH18;
U+HH18;LATIN SMALL LETTER M WITH TILDE;Ll;0;L;006D
     0303;;;;N;;;U+HH17;;U+HH17
U+HH19;LATIN CAPITAL LETTER R WITH TILDE;Lu;0;L;0052 0303;;;;N;;;;U+HH1A;
```

```
U+HH1A;LATIN SMALL LETTER R WITH TILDE;Ll;0;L;0072
     0303;;;;N;;;U+HH19;;U+HH19
U+HH1B;LATIN CAPITAL LETTER U WITH OGONEK AND ACUTE;Lu;0;L;0172
     0301;;;;N;;;;U+HH1C;
U+HH1C;LATIN SMALL LETTER U WITH OGONEK AND ACUTE;Ll;0;L;0173
     0301;;;;N;;;U+HH1B;;U+HH1B
U+HH1D;LATIN CAPITAL LETTER U WITH OGONEK AND TILDE;Lu;0;L;0172
     0303;;;;N;;;;U+HH1E;
U+HH1E;LATIN SMALL LETTER U WITH OGONEK AND TILDE;Ll;0;L;0173
     0303;;;;N;;;U+HH1D;;U+HH1D
U+HH1F;LATIN CAPITAL LETTER U WITH MACRON AND ACUTE;Lu;0;L;016A
     0301;;;;N;;;;U+HH20;
U+HH20;LATIN SMALL LETTER U WITH MACRON AND ACUTE;Ll;0;L;016B
     0301;;;;N;;;U+HH1F;;U+HH1F
U+HH21;LATIN CAPITAL LETTER U WITH MACRON AND TILDE;Lu;0;L;016A
     0303;;;;N;;;;U+HH22;
U+HH22;LATIN SMALL LETTER U WITH MACRON AND TILDE;Ll;0;L;016B
     0303;;;;N;;;U+HH21;;U+HH21
```

Note. The proposed characters hypothetically are allocated in the range U+HH00..U+HH22.

3. Rationale

3.1. Lithuanian alphabet

Any Lithuanian text and, indeed, any Lithuanian word can be written in two ways: the standard way and the accented way. Over 99% of all Lithuanian writing is done in the standard orthography; the accented orthography is used in some special cases.

The *standard* Lithuanian alphabet includes only those nine (or eighteen, if capital and small forms taken into account separately) letters: Ąą Čč Ęę Ėė Įį Šš Ųų Ūū Žž.

The diacritic marks on these nine letters mean either varieties of pronunciation (such as Lithuanian s = s, but Lithuanian š = English -sh-, German -sch-, etc.), or their historical provenance, that is, pronunciation shades in earlier forms of Lithuanian, nowadays retained only orthographically (the vowels ąęįų were nasal vowels in the 17th century pronunciation; nowadays they are ordinary long vowels, but the retention of the orthographic mark of nasality is traditional and has certain grammatical benefits, among them that of disambiguating homographs).

The diacritical marks on these nine letters of the standard alphabet (the dot on ė, the macron on ū, etc.) are *not* referred to as 'accents' by Lithuanian linguists. In the view of a Lithuanian linguist, or a school pupil, the letters ė, ū, etc., are *not* accented; they are plain letters of the standard alphabet.

These nine letters of the standard alphabet have been included in 8-bit single-byte coded character sets (ISO/IEC 8859-13, MS CP 1257, IBM CP 775, etc.), as well as the Unicode. Thus, the *standard* orthography of Lithuanian has already been taken good care of, and presents no technical problems in computer systems.

With the advent of the OpenType technology and combining accents, every letter of the standard alphabet of Lithuanian can be encoded in two equivalent ways: either as solid (precomposed) characters or as code sequences.

As to the *accented* Lithuanian orthography, the situation is different, and this proposal has been submitted to address exactly that aspect.

3.2. Lithuanian accented letters

Although most of Lithuanian writing is done in the standard alphabet, there are cases when there is a necessity to use additional marks, acute, grave, and (tilde-shaped) circumflex. These are the *accent marks* proper, as understood in Lithuanian schools and linguistics.

Every word in Lithuanian, like in other languages, has a stressed (accented) syllable, and every accented syllable can be either short or long; every long accented syllable can have either acute or circumflex tone (syllable intonation). In this way, three accent marks are used, grave, for short syllables, and acute or (tilde-shaped) circumflex for long syllables with the corresponding tone.

The **phonetic** syllable tones in Lithuanian have been inherited from the (hypothetically reconstructed) Common Proto-Baltic language, which in turn had its syllable tones derived from certain phonetic or prosodic phenomena in Proto-Indo-European. Despite the antiquity of Lithuanian syllable tones, they are still audible in modern Lithuanian.

The use of the accent marks in Lithuanian **writing** came into existence, at first in a somewhat different form than today, in the second half of the 16th century, originally derived from the practices employed in ancient Greek (polytonic) script. By the end of the 19th century, the Lithuanian system of written accents has acquired its present shape and has become fully standardized. Its use is uniform; it is not the same as, say, systems of phonetic transcription used in English dictionaries, where the system employed might vary from dictionary to dictionary, and has to be explained in a foreword. The Lithuanian system of written accents is always uniform; it is taught in schools and is unambiguously understood by every literate Lithuanian. Also, although by its nature it is meant to signify phonetic phenomena, the Lithuanian system of marking accents has its own orthographic rules and conventions, governing the shape and placement of accent marks; so in fact it is an established (albeit optional and of restricted use) *orthographic system*. As such, it is comparable, *mutatis mutandis,* to the use of orthographic systems in the Modern Greek, which can be written in either official *monotonic* or optional *polytonic* orthography.

A similar alternative (linguistic) orthographic system exists in Latvian. For example, the name of the city of Riga is written *Rīga* in the *standard* orthography of Latvian, and *Rĩga* in *linguistic* orthography. In this example the tilde means the Latvian syllable intonation, in a manner similar to Lithuanian.

A similar system exists in Croatian, which language also has syllable intonations, like ancient Greek and the Baltic languages. The Unicode characters from 0200 to 0217 are Croatian analogues of Lithuanian accented letters.

In modern Lithuanian publishing practice, both traditional and electronic, the accented orthography is used in dictionaries, encyclopaedias, grammars, and other editions of this kind, among them the Lithuanian State Terminology Database. The accented orthography is also used in ordinary texts, to disambiguate homographs, to indicate the pronunciation of rare place names, and similar.

Now, what situation do we have with the computer representation of the accented orthographies of these languages?

Both Greek monotonic and polytonic orthographies can be expressed in two equivalent ways, as precomposed characters (Greek Extended range) and as code sequences.

Both Croatian standard and accented orthographies can be expressed in two equivalent ways, as precomposed characters (see the range 0200–0217) and as code sequences.

In the case of Lithuanian, *only the standard orthography* can be fully expressed in both ways, as precomposed characters and as code sequences. *Part* of the letters of the accented orthography can be expressed in both ways; *another part* of the letters of Lithuanian accented orthography can be expressed *only as code sequences*. This creates an *awkward asymmetrical situation*, which entails various technical and logical problems.

The body of scholars who prepared the present proposal is fully aware of the ban on accepting new accented (decomposable) letters into Unicode, effective since 1999. However, we appeal on the authorities of the Unicode to reconsider this individual case, based on the following circumstances:

The situation in Lithuanian orthography is comparable to that of Greek and Croatian; the *precomposed* characters of the accentual orthographies of these languages have been included into Unicode, because they were submitted in time before the deadline in 1999. It is true that the Lithuanian side was, unfortunately, too late to submit the Lithuanian accented characters. But this unlucky oversight puts the Lithuanian language into a disadvantage position, compared to other languages with similar orthographic systems. Is it fair that the privilege of having solid accented letter codes in the Unicode, which was once deservedly granted to Greek and Croatian, now be denied to Lithuanian, on the grounds that the Lithuanian side was (very regrettably) not quick enough to react to the developments of the Unicode policies?

3.3. 8-bit single-byte encoding (National standard code tables)

Lithuanian accented letters were used even in pre-Unicode era. There are three national 8-bit code tables in Lithuania for encoding accented letters. The basic Lithuanian code table defines the basic character repertoire including accented letters. This code table is conformant with ISO/IEC 8859-13, i. e. the codes of all Lithuanian alphabet letters in both tables are the same. Common use and very important graphic characters are retained. The repertoire of this table is optimal for linguistic text processing.

Code table for Windows OS contains the basic repertoire and extra phonetic symbols in 8 and 9 columns. This code table is conformant with Windows-1257 code table.

It is interesting to note that accented letters were used yet in DOS environments. Constructed code table for DOS contains basic repertoire and box drawing symbols and is conformant with IBM CP 775 for Baltic States. All this proves the importance of accented letters for Lithuanian linguistic needs.

Basic code table is shown in Addendum I.

3.4. Multiple-octet encoding in the ISO/IEC 10646

All letters of Lithuanian alphabet are already encoded in the ISO/IEC 10646 (have the UCS code points). The situation with Lithuanian accented letters is more complicated. As it was mentioned, Lithuanian accented letters are Latin script letters with grave accent, acute accent or tilde. So some Lithuanian accented letters are also the letters in other languages. For example, LATIN LETTER A WITH ACUTE is also in Irish, Icelandic, Portuguese, Slovak etc. languages, LATIN LETTER N WITH TILDE is also in Basque, Breton and Spanish languages. Thus they have separate the UCS code points.

All together there are 33 Lithuanian accented letters that have the UCS code points and 35 accented letters have not separate ones. Letters missing UCS code points are shadowed (see below).

À Á Ã Ą́ Ą̃ È É Ẽ Ę́ Ę̃ Ė́ Ė̃ Ì Í Ĩ Į́ Į̃ Ý Ỹ

à á ã ą́ ą̃ è é ẽ ę́ ę̃ ė́ ė̃ i̇̀ i̇́ i̇̃ į̇́ į̇̃ ý ỹ

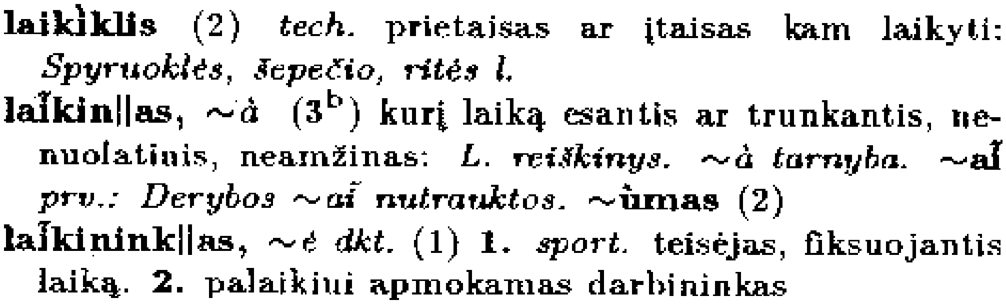
J̃ L̃ M̃ Ñ Ò Ó Õ R̃ Ù Ú Ũ Ų́ Ų̃ Ū́ Ū̃

j̃ l̃ m̃ ñ ò ó õ r̃ ù ú ũ ų́ ų̃ ū́ ū̃

There is another problem with small letter "i" (and "i with ogonek" and "j"). Lithuanian letter "i" is with a dot above. All accented forms of "i" should be also with a dot (see samples in 2.4). In ISO/IEC 10646 all such forms are dotless. For example, LATIN SMALL LETTER I WITH ACUTE. We ought to retain a dot above, in that case, so we should define and name this character with explicit name of dot and diacritic as LATIN SMALL LETTER I WITH DOT ABOVE AND ACUTE.

4. Samples

In [3, p.350]:

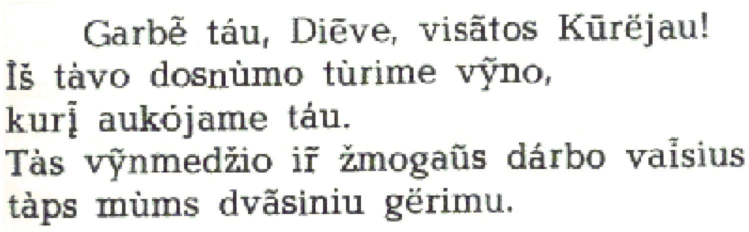


laiki̇̀klis (2) *tech.* prietaisas ar įtaisas kam laikyti: *Spyruoklės, šepečio, ritės l.*

lai̇̃kin||as, *~à* (3^b) kurį laiką esantis ar trunkantis, nenuolatinis, neamžinas: *L. reiškinys. ~à tarnyba.* **~ai̇̃** *prv.: Derybos ~ai̇̃ nutrauktos.* **~ùmas** (2)

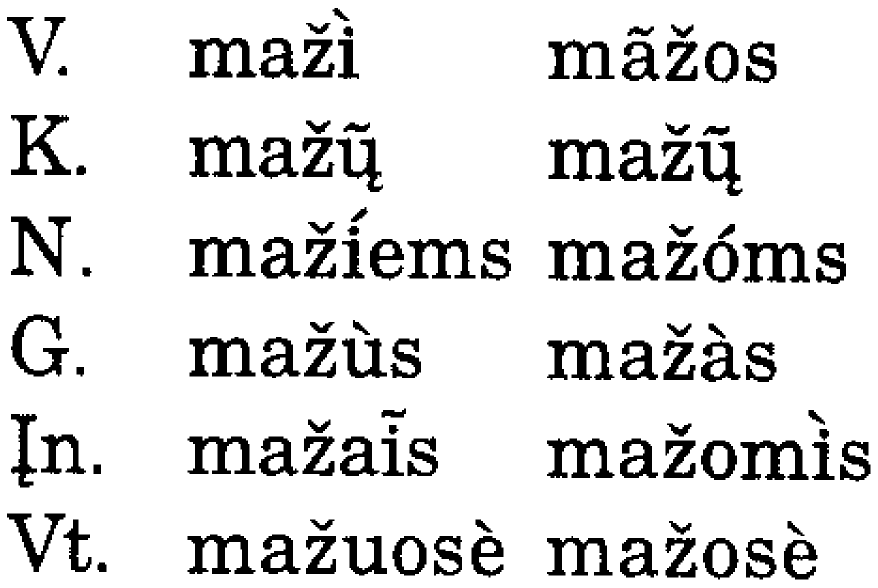
lai̇̃kiṇink||as, *~ė dkt.* (1) **1.** *sport.* teisėjas, fiksuojantis laiką. **2.** palaikiui apmokamas darbininkas

In [13, p.75]:



Garbė̃ táu, Diẽve, visãtos Kūrė̃jau!
Ĩš tàvo dosnùmo tùrime vỹno,
kurį̃ aukójame táu.
Tàs vỹnmedžio ir̃ žmogaũs dárbo vai̇̃sius
tàps mùms dvãsiniu gė̃rimu.

In [4, p.38]. Note the accented "i":



V.	maži̇̀	mãžos
K.	mažų̃	mažų̃
N.	maži̇́ems	mažóms
G.	mažùs	mažàs
Įn.	mažai̇̃s	mažomi̇̀s
Vt.	mažuosè	mažosè

5. Rendering of the sequences

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

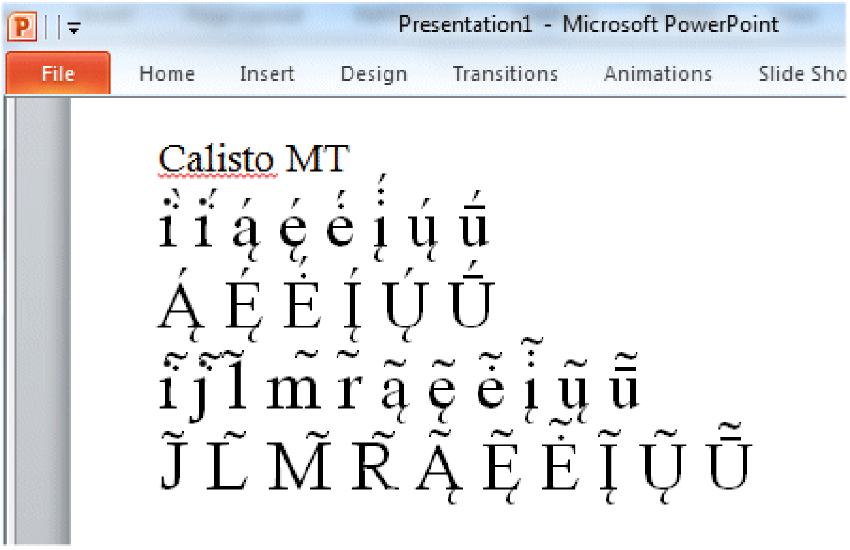
In 2006 Lithuanian accented letters as composite ones were identified by the named sequences. See Unicode Character Database file "NamedSequences.txt"
http://www.unicode.org/Public/UNIDATA/NamedSequences.txt
The extract from this file is shown in Addendum II of this proposal.

WG2 also accepted the inclusion in the standard of an extended collection named 'Contemporary Lithuanian Letters', an extended collection of standalone characters and named sequences, corresponding to the repertoire shown in Addendum I in document N3090:
http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3090.doc

All this means that Lithuanian accented letters should be expressed by sequences. But it is not an option. Text entering, editing and processing is more complicated. Quality typesetting is almost impossible because the rendering of these sequences depends on fonts and applications and is not always suitable. The common user has the impression that the Lithuanian accented letters are in fact missing.

The rendering of accented letters expressed by the sequences in various applications is illustrated below. Some (sometimes almost all) positions of diacritics are incorrect. None of analyzed environments and fonts displays all sequences in suitable way.

MS PowerPoint 2010



Calisto MT

i̇́ i̇̀ ą́ ę́ ė́ į̇́ ų́ ū́

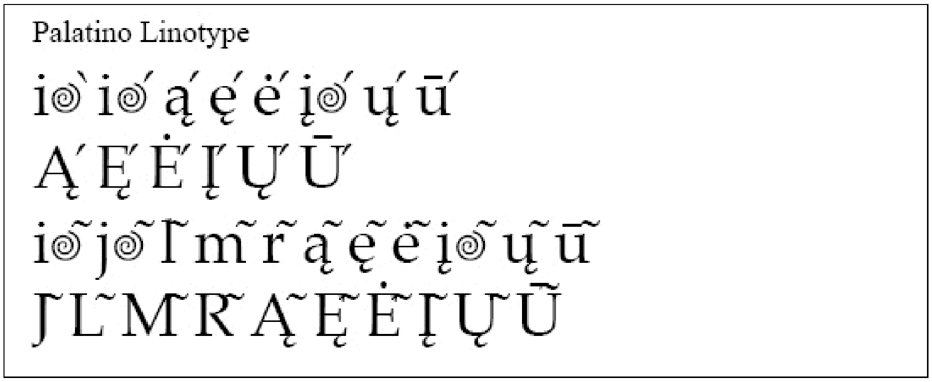
Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃

J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Only 22-24 letters look satisfactory, others are incorrect..

Adobe InDesign CS2



Palatino Linotype

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́

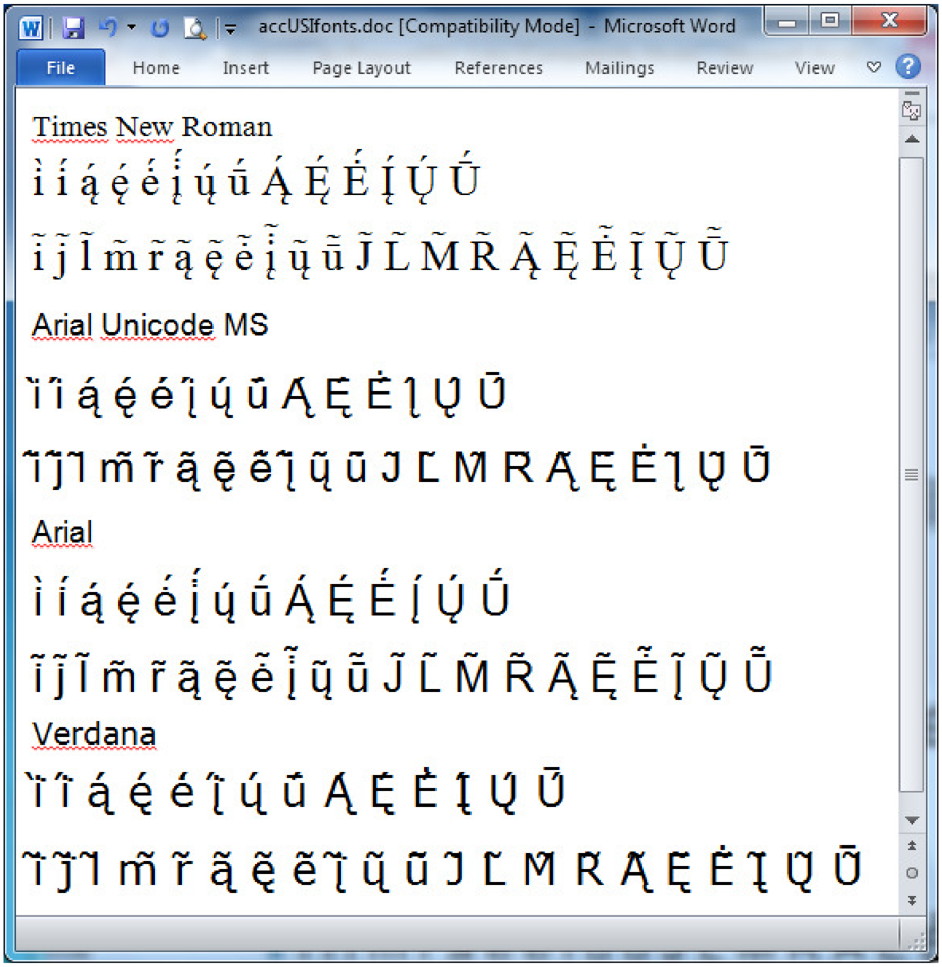
Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃

J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Palatino Linotype has not U+0307 COMBINING DOT ABOVE. Non-existing characters are visualized by strange looking symbol-snail (instead of usual rectangle). All letters are displayed incorrectly.

Microsoft Word 2010



Times New Roman

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Arial Unicode MS

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Arial

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

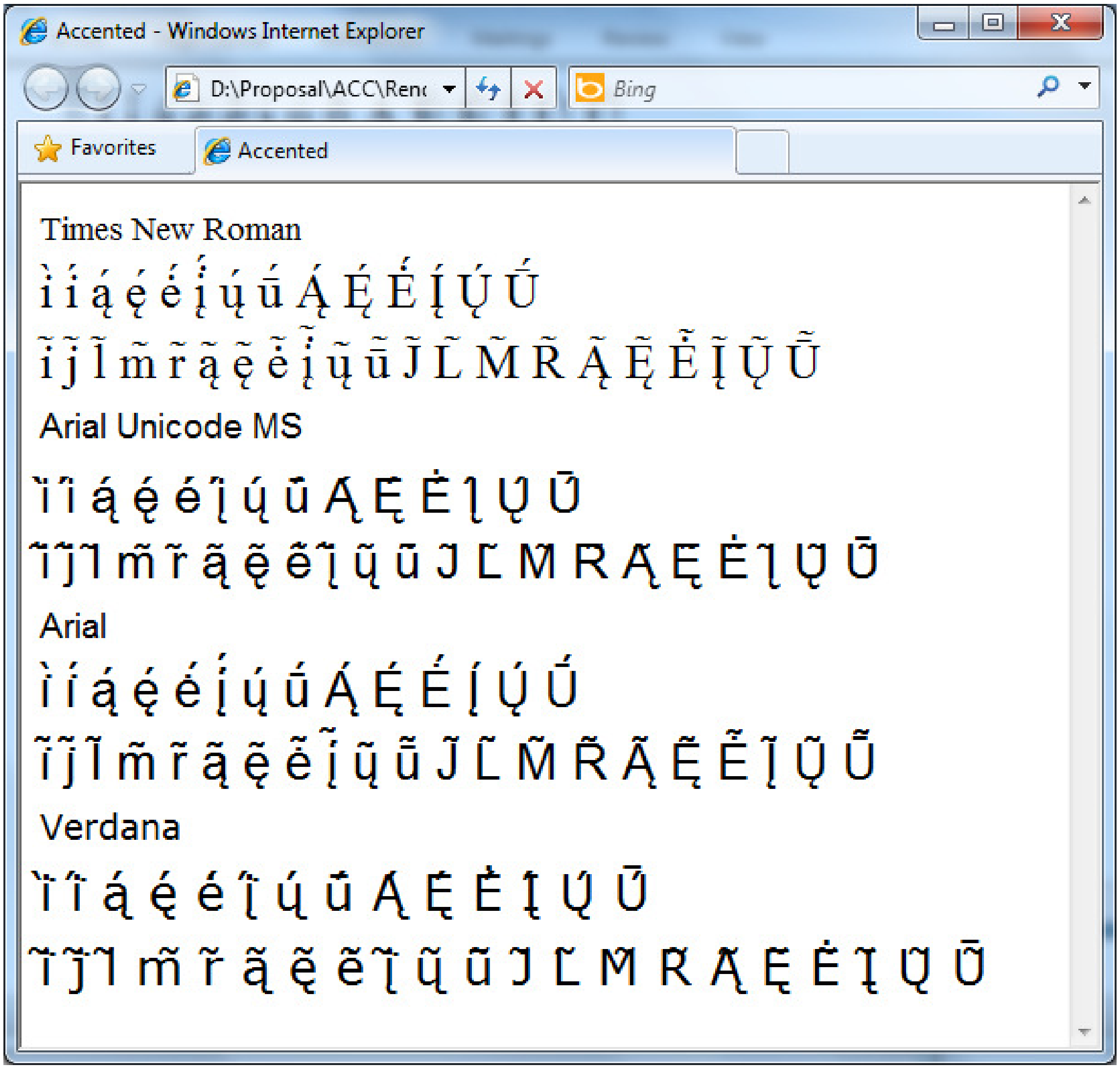
Verdana

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

The positions of diacritics in *Times New Roman* and *Arial* are correct except two letters: small-i-ogonek-and-acute and small-i-ogonek-and-tilde. Almost all letters of *Arial Unicode MS* and *Verdana* are incorrect.

Internet Explorer 8.0 (Windows 7 Professional)



Times New Roman

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Arial Unicode MS

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Arial

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

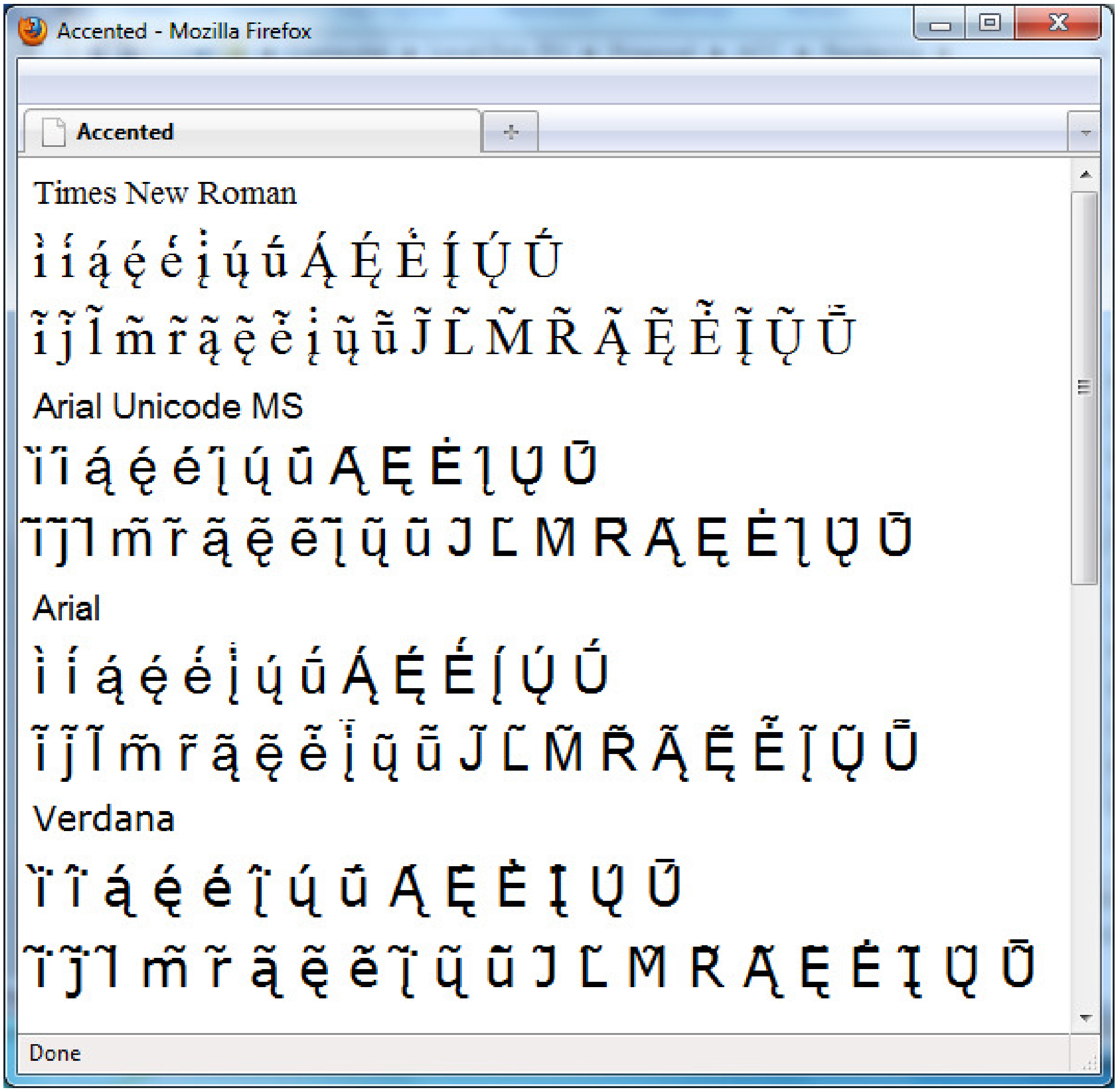
Verdana

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Rendering results of all web browsers are more or less similar to *Microsoft Word* 2010.

Mozilla Firefox 3.6.3 (Windows 7 Professional)



Times New Roman

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Arial Unicode MS

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Arial

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

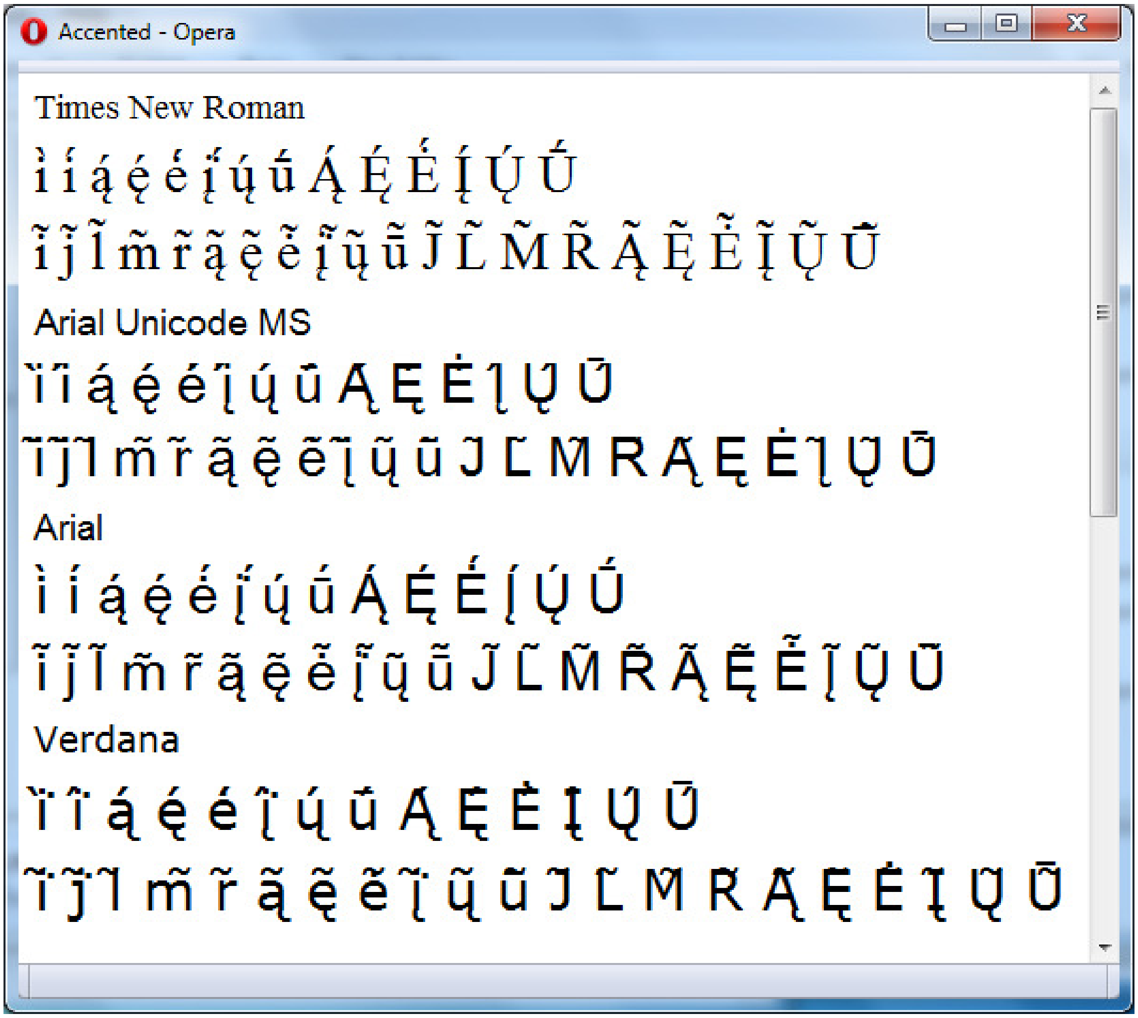
i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Verdana

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Opera 11.50 (Windows 7 Professional)



Times New Roman

i̇̀ i̇́ ą́ ę́ ė́ į́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Arial Unicode MS

i̇̀ i̇́ ą́ ę́ ė́ į́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Arial

i̇̀ i̇́ ą́ ę́ ė́ į́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

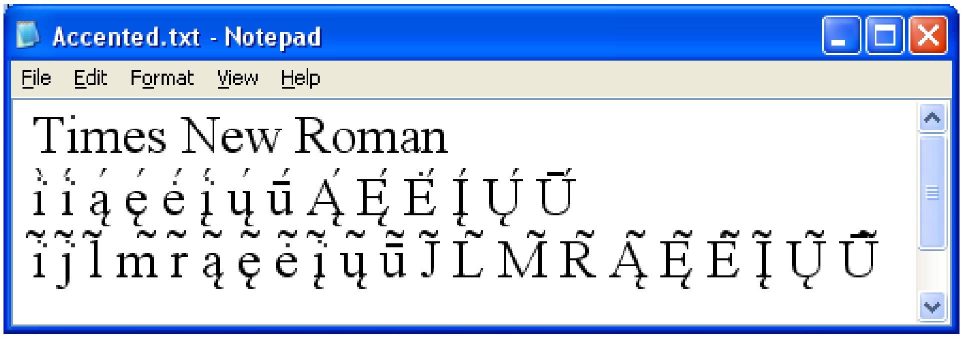
i̇̃ j̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Verdana

i̇̀ i̇́ ą́ ę́ ė́ į́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

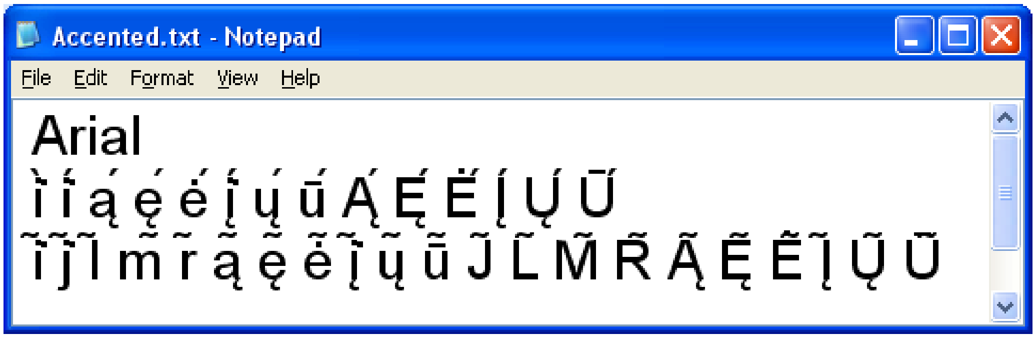
Notepad (Microsoft Windows XP Professional, SP3)



Times New Roman

i̇̀ i̇́ ą́ ę́ ė́ į́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃



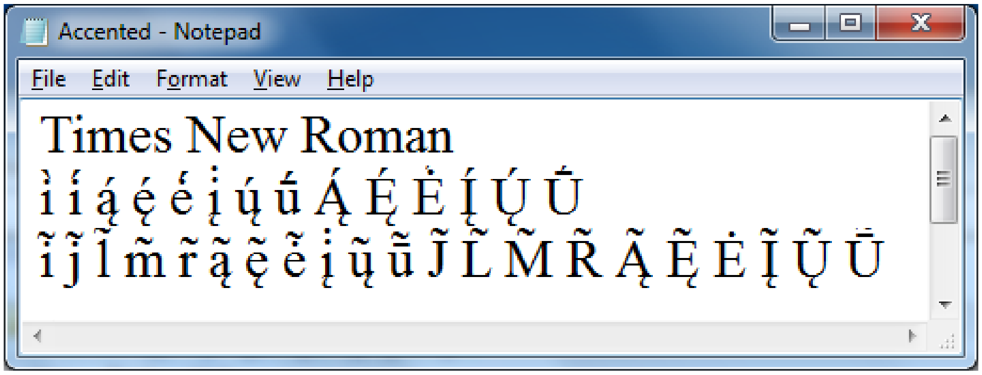
Arial

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

We see the results of *Notepad* depend on operating system. Almost all positions of diacritics in Microsoft Windows XP Professional are incorrect.

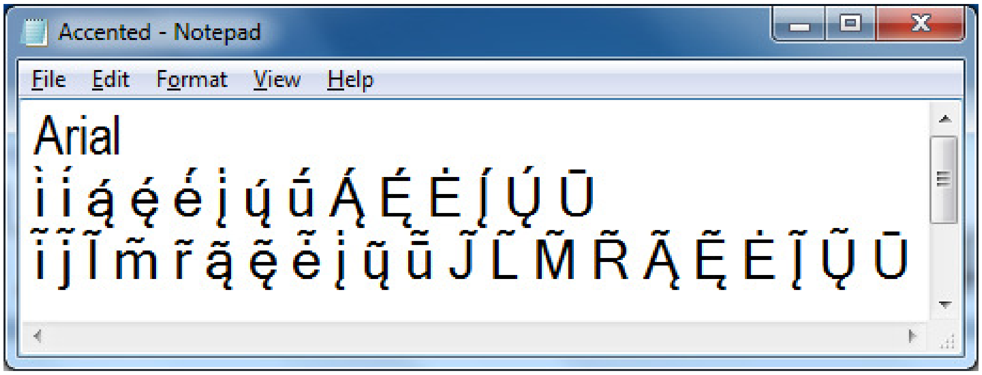
Notepad (Windows 7 Professional)



Times New Roman

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃



Arial

i̇̀ i̇́ ą́ ę́ ė́ į̇́ ų́ ū́ Ą́ Ę́ Ė́ Į́ Ų́ Ū́

i̇̃ j̇̃ l̃ m̃ r̃ ą̃ ę̃ ė̃ į̇̃ ų̃ ū̃ J̃ L̃ M̃ R̃ Ą̃ Ę̃ Ė̃ Į̃ Ų̃ Ū̃

Meanwhile the results of *Notepad* in Windows 7 Professional are correct (except 5 letters for *Times New Roman* and 6 letters for *Arial*).

6. References

1. M. Daukša, *Kathechismas (1595)* and *Postilla catholicka (1599).*

2. *Lietuvių kalbos žodynas*, I–XVIII t. [*Dictionary of Lithuanian Language,* I–XVIII volumes], Vilnius, 1956–1997.

3. *Dabartinės lietuvių kalbos žodynas*, vyr. red. St. Keinys [*Dictionary of Modern Lithuanian Language,* ed. by St. Keinys], Vilnius, Mokslo ir enciklopedijų leidykla, 1993.

4. Adelė Laigonaitė, Zigmas Zinkevičius, *Lietuvių kalba. Mokomoji knyga X klasei* [*Lithuanian Language. Textbook for X form*], Kaunas, Sviesa, 1997.

5. S. Matulaitienė, *Skaitiniai. Vadovėlis VI klasei* [*Lithuanian Texts. Textbook for VI form*], Kaunas, Šviesa, 1990.

6. *Lithuanian Grammar*, ed. by V. Ambrazas, Vilnius, Baltos lankos, 1997.

7. T. Mathiassen, A *Short Grammar of Lithuanian*, Slavica Publishers, Columbus, Ohio, 1996.

8. M. Ramonienė, I. Press, *Colloquial Lithuanian*, London and New York, Routledge, 1996.

9. V. Tumasonis. *Encoding of Lithuanian Accented Letters*. Proceedings of GLDV'99. Multilingual Corpora: Encoding, Structuring, Analysis. Frankfurt/Main, 1999.

10. B. Piesarskas, *Dvitomis anglų-lietuvių kalbų žodynas* [*English-Lithuanian Dictionary, 2* volumes], Vilnius, Alma littera, 2004.

11. *Vokiečių-lietuvių kalbų žodynas* [*German-Lithuanian Dictionary*], Vilnius, Mokslas, 1989.

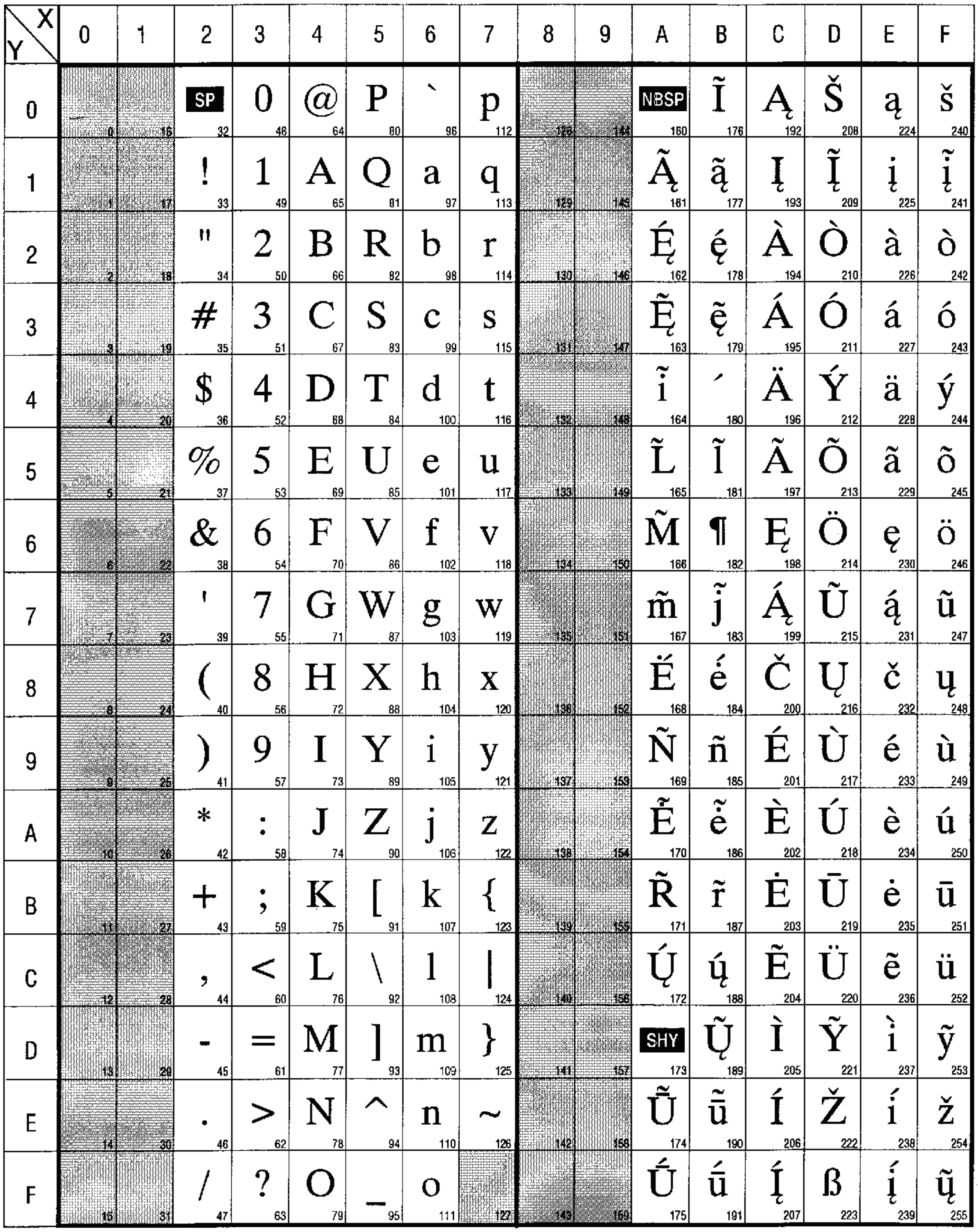
12. A. Parenti, *Italiano-Lituano, Lituano-Italiano*, Garzanti Editore, 1994.

13. Romos Mišiolas. *Gedulinis Mišiolas* [*Missalis Romani. Missale Parvum*], Kaunas - Vilnius, 1982.

14. A. Bendorienė, V. Bogušienė, *Tarptautinių žodžių žodynas* [*Dictionary of International words*], Vilnius, Alma littera, 2008.

Addendum I

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



Y \ X	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	0	16	SP 32	0 48	@ 64	P 80	` 96	p 112	128	144	NBSP 160	Ĩ 176	Ą 192	Š 208	ą 224	š 240
1	1	17	! 33	1 49	A 65	Q 81	a 97	q 113	129	145	Ą̃ 161	ą̃ 177	Į 193	Į̃ 209	į 225	į̃ 241
2	2	18	" 34	2 50	B 66	R 82	b 98	r 114	130	146	Ę́ 162	ę́ 178	À 194	Ò 210	à 226	ò 242
3	3	19	# 35	3 51	C 67	S 83	c 99	s 115	131	147	Ę̃ 163	ę̃ 179	Á 195	Ó 211	á 227	ó 243
4	4	20	$ 36	4 52	D 68	T 84	d 100	t 116	132	148	i̇̃ 164	´ 180	Ä 196	Ý 212	ä 228	ý 244
5	5	21	% 37	5 53	E 69	U 85	e 101	u 117	133	149	L̃ 165	l̃ 181	Ã 197	Õ 213	ã 229	õ 245
6	6	22	& 38	6 54	F 70	V 86	f 102	v 118	134	150	M̃ 166	¶ 182	Ę 198	Ö 214	ę 230	ö 246
7	7	23	' 39	7 55	G 71	W 87	g 103	w 119	135	151	m̃ 167	j̃ 183	Ą́ 199	Ũ 215	ą́ 231	ũ 247
8	8	24	(40	8 56	H 72	X 88	h 104	x 120	136	152	Ė́ 168	ė́ 184	Č 200	Ų 216	č 232	ų 248
9	9	25	) 41	9 57	I 73	Y 89	i 105	y 121	137	153	Ñ 169	ñ 185	É 201	Ù 217	é 233	ù 249
A	10	26	* 42	: 58	J 74	Z 90	j 106	z 122	138	154	Ė̃ 170	ė̃ 186	È 202	Ú 218	è 234	ú 250
B	11	27	+ 43	; 59	K 75	[91	k 107	{ 123	139	155	R̃ 171	r̃ 187	Ė 203	Ū 219	ė 235	ū 251
C	12	28	, 44	< 60	L 76	\ 92	l 108	| 124	140	156	Ų́ 172	ų́ 188	Ẽ 204	Ü 220	ẽ 236	ü 252
D	13	29	- 45	= 61	M 77	] 93	m 109	} 125	141	157	SHY 173	Ų̃ 189	Ì 205	Ỹ 221	ì 237	ỹ 253
E	14	30	. 46	> 62	N 78	^ 94	n 110	~ 126	142	158	Ū̃ 174	ū̃ 190	Í 206	Ž 222	í 238	ž 254
F	15	31	/ 47	? 63	O 79	_ 95	o 111	127	143	159	Ū́ 175	ū́ 191	Į́ 207	ß 223	į́ 239	ų̃ 255

Addendum II

The extract from Unicode Character Database file "NamedSequences.txt"

```
# NamedSequences-6.0.0.txt
# Date: 2010-05-18, 10:48:00 PDT [KW]
#
# Unicode Character Database
# Copyright (c) 1991-2010 Unicode, Inc.
#
# ================================================

# Additions for Lithuanian. Provisional 2006-05-18, Approved 2007-10-19

LATIN CAPITAL LETTER A WITH OGONEK AND ACUTE;0104 0301
LATIN SMALL LETTER A WITH OGONEK AND ACUTE;0105 0301
LATIN CAPITAL LETTER A WITH OGONEK AND TILDE;0104 0303
LATIN SMALL LETTER A WITH OGONEK AND TILDE;0105 0303
LATIN CAPITAL LETTER E WITH OGONEK AND ACUTE;0118 0301
LATIN SMALL LETTER E WITH OGONEK AND ACUTE;0119 0301
LATIN CAPITAL LETTER E WITH OGONEK AND TILDE;0118 0303
LATIN SMALL LETTER E WITH OGONEK AND TILDE;0119 0303
LATIN CAPITAL LETTER E WITH DOT ABOVE AND ACUTE;0116 0301
LATIN SMALL LETTER E WITH DOT ABOVE AND ACUTE;0117 0301
LATIN CAPITAL LETTER E WITH DOT ABOVE AND TILDE;0116 0303
LATIN SMALL LETTER E WITH DOT ABOVE AND TILDE;0117 0303
LATIN SMALL LETTER I WITH DOT ABOVE AND GRAVE;0069 0307 0300
LATIN SMALL LETTER I WITH DOT ABOVE AND ACUTE;0069 0307 0301
LATIN SMALL LETTER I WITH DOT ABOVE AND TILDE;0069 0307 0303
LATIN CAPITAL LETTER I WITH OGONEK AND ACUTE;012E 0301
LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND ACUTE;012F 0307 0301
LATIN CAPITAL LETTER I WITH OGONEK AND TILDE;012E 0303
LATIN SMALL LETTER I WITH OGONEK AND DOT ABOVE AND TILDE;012F 0307 0303
LATIN CAPITAL LETTER J WITH TILDE;004A 0303
LATIN SMALL LETTER J WITH DOT ABOVE AND TILDE;006A 0307 0303
LATIN CAPITAL LETTER L WITH TILDE;004C 0303
LATIN SMALL LETTER L WITH TILDE;006C 0303
LATIN CAPITAL LETTER M WITH TILDE;004D 0303
LATIN SMALL LETTER M WITH TILDE;006D 0303
LATIN CAPITAL LETTER R WITH TILDE;0052 0303
LATIN SMALL LETTER R WITH TILDE;0072 0303
LATIN CAPITAL LETTER U WITH OGONEK AND ACUTE;0172 0301
LATIN SMALL LETTER U WITH OGONEK AND ACUTE;0173 0301
LATIN CAPITAL LETTER U WITH OGONEK AND TILDE;0172 0303
LATIN SMALL LETTER U WITH OGONEK AND TILDE;0173 0303
LATIN CAPITAL LETTER U WITH MACRON AND ACUTE;016A 0301
LATIN SMALL LETTER U WITH MACRON AND ACUTE;016B 0301
LATIN CAPITAL LETTER U WITH MACRON AND TILDE;016A 0303
LATIN SMALL LETTER U WITH MACRON AND TILDE;016B 0303
```

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html for guidelines and details before filling this form.
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest *Roadmaps*.

A. Administrative

1. **Title:** ***Proposal to add Lithuanian accented letters to the UCS***
2. Requester's name: *Lithuanian Standards Board (LST)*
3. Requester type (Member body/Liaison/Individual contribution): *ISO Member*
4. Submission date: *2011-XX-XX*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *Yes*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *No*
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *Yes*
      Name of the existing block: *Latin Extended-D*
2. Number of characters in proposal: *35*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary *X* B.1-Specialized (small collection) B.2-Specialized (large collection)
   C-Major extinct D-Attested extinct E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable usage symbols
4. Is a repertoire including character names provided? *Yes*
   a. If YES, are the names in accordance with the "character naming guidelines"
      in Annex L of P&P document? *Yes*
   b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
      Vladas Tumasonis
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
   of proposed characters attached? *Yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input,
   presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *No*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

[1] Form number: N3902-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? *Yes*
   If YES explain *In 1999 the proposal has not been accepted*
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
   If YES, with whom?
   If YES, available relevant documents:
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *Yes*
   Reference: *See text*
4. The context of use for the proposed characters (type of use; common or rare) *Common*
   Reference: *See text*
5. Are the proposed characters in current use by the user community? Yes
   If YES, where? Reference: *In Lithuania*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *Yes*
   If YES, is a rationale provided?
   If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *No*
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? *No*
   If YES, is a rationale for its inclusion provided?
   If YES, reference:
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? *Yes*
   If YES, is a rationale for its inclusion provided? *Yes*
   If YES, reference: *Is enclosed*
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? *No*
    If YES, is a rationale for its inclusion provided?
    If YES, reference:
11. Does the proposal include use of combining characters and/or use of composite sequences? *No*
    If YES, is a rationale for such use provided?
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    If YES, reference:
12. Does the proposal contain characters with any special properties such as control function or similar semantics? *No*
    If YES, describe in detail (include attachment if necessary)
13. Does the proposal contain any Ideographic compatibility characters? *No*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference:

ISO/IEC JTC1/SC2/WG2 N4191-A

Date: 2010-02-10

ADDENDUM TO THE DOCUMENT *ISO/IEC JTC1/SC2/WG2 N4191*

This Addendum should be considered as addendum to chapter 5 *Rendering of sequences* of the document ISO/IEC JTC1/SC2/WG2 N4191.

Apart from rendering there are other problems when using Combining Character Sequences mechanism. Lithuania request to have all Lithuanian letters as encoded including 35 accented letters identified by the named sequences is fully legitimate and should be supported by all means. Lithuania request is supported by the stakeholders from several parties in Lithuania and abroad (ISO/IEC JTC1/SC2/WG2 N4187, 4188, 4189, 4192). The decision from 1996 that encoding of precomposed Latin letters is no longer necessary severely impacts usage of contemporary Lithuanian language in IT applications.

Currently Lithuania is only European country which is forced to use precomposed sequences for it official language. European Union tries to preserve its linguistic diversity promoting all but especially lesser used languages that might be in danger. Lithuanian belongs to the category due to the highest rate of emigration among EU members. Actually it is a Diaspora nation with more than half a million of Lithuanians living abroad. Schoolchildren in the emigrant's communities are provided with a distant computer aided learning possibilities in order to support and develop their native tongue to such an extent that they were able to study at Lithuanian universities. Accentuation in general and accented texts in particular is crucial for their language skills.

Accented text become of paramount importance also for the native speakers living in the country because of the language change. The stress in Lithuanian is not fixed changing its position depending on the accentuation paradigm. Sometime initial, sometimes final parts of the word are stressed. The latter case is in danger as many native speakers tend to stress only initial positions (as it is easier to pronounce a word in this way). If the endings permanently loose their stress, they will be shortened or disappear altogether as it has already happened for another Baltic language, i.e. Latvian. In order to preserve the stress in the final positions more texts have to be accentuated. Attempts are made to do it automatically using special software tools as "Kirciuokle" (http://donelaitis.vdu.lt). However, such tools can be supported only by the national fonts which are not compatible with popular learning environments, e.g. MOODLE.

Recent developments in the field of digitalization of the cultural heritage and other fields raised new challenges coming and important problems for Lithuanian language not having all letters in Unicode as well:

Common research infrastructures based on linguistic resources. Since the appearance of ESFRI, (the European Strategy Forum on Research Infrastructures that is used as a strategic instrument to develop the scientific integration of Europe and to strengthen its international outreach) a number of projects, initiatives and consortia appeared. Some of them, like CLARIN (Common Language Resources and Technologies Infrastructure) are committed to establish an integrated and interoperable research infrastructure of language resources and its technology enabling e-Humanities. Lithuania is a member of the consortium providing its linguistic resources (text and speech) as well as software tools for natural language processing for an international research community. Since Lithuanian is the oldest living Indo-European language it attracts researchers' attention from all over the world. They use corpora, digital dictionaries and search tools for their research. For their needs accented letters are necessary. Moreover, other languages included in CLARIN infrastructure are fully supported by the ISO/IEC 10646 and the UNICODE standards, leaving Lithuanian language in a worse position.

Search in large text corpora including accented text. As languages tend to shrink with respect to the number of words that are being used in everyday spoken language it is important to compile large text corpora, storing language information as part of national heritage. That is of paramount importance for the lesser used languages, Lithuanian language being one of them. Large corpora serve multiple purposes; therefore they include a great variety of texts, accented texts being part of them. Since Unicode is currently used as the main corpora encoding standard, it is important to have uniform representation for all the letters used (both accented and regular) in the corpus, in order to ensure correct search options. Search of words and their environment (so called concordancing) is one of the basic corpora-related services, implemented either as an on-line service, or as a network service, that can be integrated as a building block in other complex services. Accented letters are important for disambiguation when defining search patterns, as well as in these cases, when corpora are used in machine learning for algorithm training purposes. Moreover, extended search options that include accented letters are vital for research in the field of computational linguistics. For all the above mentioned reasons, search of accented Lithuanian letters should be supported correctly in different operational and design environments, as well as by different application systems.

Search in speech corpora. Taking into account the increasing need of speech analysis and synthesis in different applications (e.g. virtual assistant applications, automatic speech-text recording of medical records, media information, etc.), correct search options should be ensured for speech corpora, including both spoken language and its transcription, the latter including also accented letters. Accents are important for disambiguation, while using speech corpora for algorithm training purposes. Lithuanian speech corpora and speech recognition/synthesis are in a fast development phase right now, and their designers are pointing to the accented letter management problem as one of the most important problems in their design field.

Search in electronic dictionaries. Electronic dictionaries/thesaurus is an important part of the programs for the preservation of the national heritage. Dictionaries explicitly include accented words, and regular dictionary search must support also search patterns with accented letters.

Conventional multi-level search patterns. Accented letters are needed while executing multi-level search, i.e. taking the results of the first search level for narrowing search at the next level. This is normally done by applying the „copy-paste" procedure, resulting in a failed search for numerous standard applications.

Modern learning systems based on machine learning. Modern learning systems are increasingly using different machine learning (artificial intelligence) approaches. In this case, algorithm training is the main component, requiring large corpora, speech corpora, thesaurus, dictionaries and other linguistic resources with correct search procedures implemented. Here, accented letters are especially important in language learning systems.

Accented Lithuanian letters are very widely used. Their usage spans various media and means of information technology. Presently, the Combining Character Sequences mechanism in most IT products is being supported only episodically. Lithuanian accented letters are mostly used in Lithuania, so it would seem that it would be possible to code the missing letters using the PUA defined by the UNICODE and adopting a local Lithuanian standard for that. This is only temporary and palliative means which does not take into account proliferation of IT usage. Without proper inclusion of all

Lithuanian letters into the UNICODE, the emergence of international software that is non-discriminative (i.e. with Lithuanian full sorting order, search engines, etc.) to Lithuanian seems not likely as well, which raises an issue of the compliance with the license agreements and ROI of the same product in different countries.

The optimal way of solving the problem (as was done with ancient Greek) would be **the appointment of UNICODE positions for the missing Lithuanian letters in the new version of the UNICODE and ISO/IEC 10646 standard**. The accented letters of other languages have their own UNICODE coding for a long time already and neither composition sequences nor PUA are being used for their information processing. The same must be done for the rest 35 Lithuanian accented letters.