DATE: 7 June 2013
DOC TYPE: Working Group Document
TITLE: Lithuanian Accents - Follow-Up
SOURCE: Deborah Anderson, Script Encoding Initiative, UC Berkeley
STATUS: Liaison contribution
ACTION: For consideration by WG2
DISTRIBUTION: ISO/IEC JTC1/SC2/WG2

In the "Lithuanian Ad-hoc Report" from Feb. 2012 (http://std.dkuug.dk/JTC1/SC2/WG2/docs/n4242.pdf), one suggestion on how to proceed was:

> Consider the potential role that that the Unicode CLDR project (http://cldr.unicode.org/) might have in capturing requirements for Lithuanian text processing, in driving awareness of those requirements, and in improving implementations in products that make use of CLDR data. Note: In relation to this, Lithuanian experts are encouraged to actively participate in the Unicode CLDR project.

Peter Edberg of Apple Inc. submitted the Lithuanian accented letter combinations to CLDR on 3 May 2012 and they were contained as part of the auxiliary exemplar set for Lithuanian in CLDR release 22 (released September 2012 – see screenshots on page 2).

The UNICODE LOCALE DATA MARKUP LANGUAGE (LDML) spec, which describes the format used in CLDR, characterizes the main and auxiliary exemplar sets as follows:

> The *main* [exemplar] set should contain the minimal set required for users of the language [which usually matches the basic alphabet taught in school], while the *auxiliary* exemplar set is designed to encompass additional characters: those non-native or historical characters that would customarily occur in common publications, dictionaries, and so on.
> (Source: http://www.unicode.org/reports/tr35/tr35-general.html#Character_Elements)

Two typical usages for the main and auxiliary exemplars are:
- determining the extent to which a font can support a given language (ranking the support level): Does it have glyphs & mapping info for the main exemplars, punctuation exemplars and number systems used by the locale? Can it also support the auxiliary exemplars?
- determining the extent to which a keyboard layout can support input of a given language or determining the set of languages that can be typed using a given key layout.
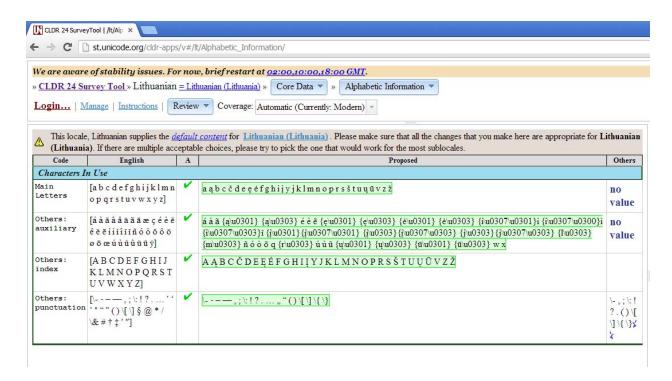
Some implementations may use them for:
- a very basic language filter
- determining which characters to allow without flagging (e.g., in identifiers)
- determining which non-Unicode code pages support a language.

The information contained in CLDR is used widely by various companies, including Apple, Google, IBM, etc. (See further: http://cldr.unicode.org/#TOC-Who-uses-CLDR-).

Below is a screenshot of the Lithuanian accented letter combinations in the CLDR Survey Tool (http://st.unicode.org/cldr-apps/v#/lt/Alphabetic_Information/).

The green entries in the "Proposed" column are the currently approved Lithuanian values.

Close-up of the auxiliary set:

To file additions or corrections, contact CLDR via: http://unicode.org/cldr/trac/newticket.