



Яко апостолом единоправнии и словенских стран учителя,
Кирилле и Мефодие богомудрии, Владыку всех молите,
вся языки словенския утвердiti в Православии и единомыслии,
умирити мир и спасти души наша.
Тропарь равноапостольных Мефодия и Кирилла

ЦЕРКОВНОСЛАВЯНСКАЯ АЗБУКА

КОДИРОВАНИЕ ДЛЯ ИСПОЛЬЗОВАНИЯ В ЭЛЕКТРОННЫХ ДОКУМЕНТАХ

Версия CU v1.1

Составление таблицы кодирования азбуки церковнославянского языка для использования в текстах электронных документов и на этой основе создание электронных программ шрифтов и раскладки клавиатуры разработаны в исполнении повеления Христа Спасителя: «Идите и научите все народы...», в продолжении труда святых равноапостольных Кирилла и Мефодия просветителей словенских, в процессе развития и сохранения культурного наследия Святой Православной Руси, в полученном от прародителей и родителей почтении и любви ко Христу и Богу, Церкви и Отечеству.

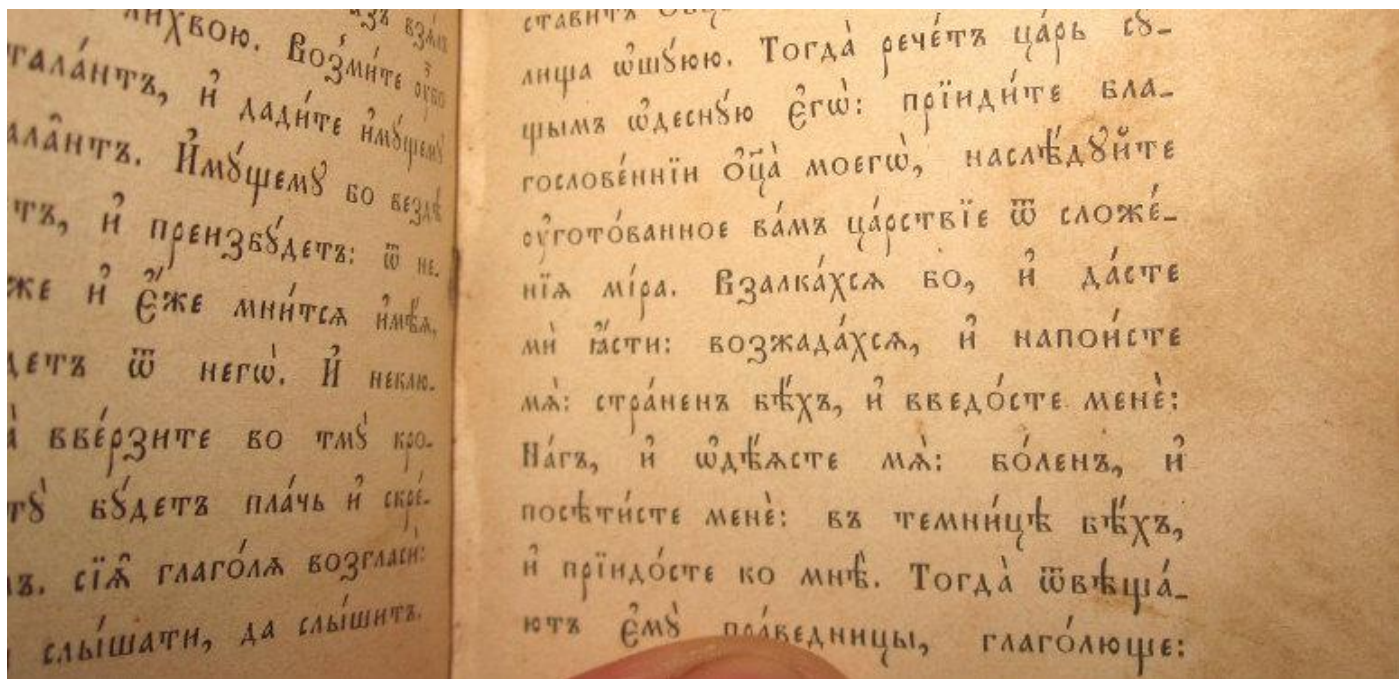


Фото 1. Фрагмент Евангелия. Москва, Синодальная типография, XIX в., мини формат.

Описание

Кодирование азбуки ЦС языка, а также соответствующего разработанного автором данной статьи шрифта и раскладки клавиатуры для набора текста соответствует The Unicode Standard, Version 9.0.
<http://www.unicode.org>

Юникод (Unicode) – один из ряда стандартов кодирования символов, где каждому символу присваивается индивидуальный код, одинаково распознаваемый различными вычислительными системами, использующими этот стандарт.

Азбука, а также соответствующий шрифт и раскладка клавиатуры располагаются в следующих областях

символов юникод стандарта, табл. 1.

Basic Latin. Latin-1 Supplement. Latin Extended-A. Spacing Modifier Letters. Combining Diacritical Marks. Cyrillic. Cyrillic Extended-A. Cyrillic Extended-B. Supplemental Punctuation. General Punctuation. Miscellaneous Symbols. Miscellaneous Symbols and Pictographs.

Информацию о шрифте ЦС языка «Панти» (Panteley v.1.5); о раскладке клавиатуры ЦС языка «Церковнославянский» kbl-cu.v1.51; по фонетической азбуке ЦС языка посмотреть на странице разработки: <http://vcsl.dn.ua/patr/soft/font/slavonic.htm>.

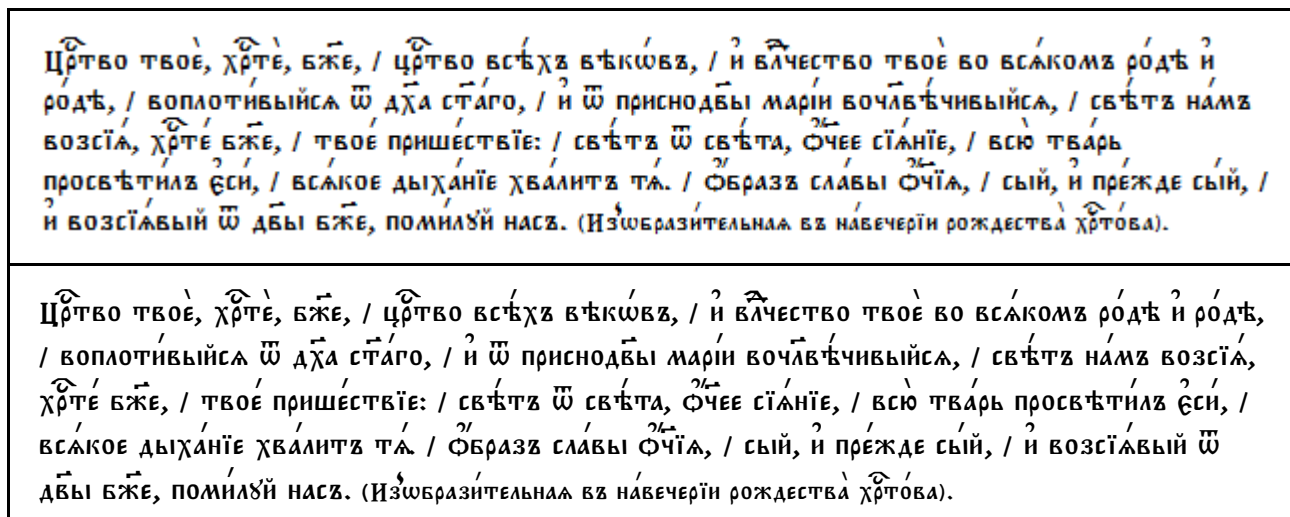


Таблица 1. Фото (вверху) и текст (внизу). Образец текста с заданным шрифтом **Panty v.1.4**

Если образец шрифта снизу отличается от его фото сверху, то для дальнейшего чтения страницы с полным отображением информации необходимо:

или обновить на компьютере браузер,

или установить на компьютер шрифт «Panty», загрузив со страницы разработки:

<http://vcsl.dn.ua/patr/soft/font/slavonic.htm>.

Каждая буква алфавита (табл. 2), знак пунктуации (табл. 3), надстрочный знак (табл. 4) и буква под титлом (табл. 5) церковнославянского языка имеют свой юникод (набор цифр и букв), в соответствии с которым кодируется шрифт и компьютерная раскладка, предназначенная для набора текста на ЦС языке. В другие системы кодирования юникод, в том числе, и ЦС языка, автоматически конвертируется специальными входящими в шрифты или программы таблицами. В ниже представленных таблицах показано соотношение символов церковнославянского языка, их кода, звучания, названия, букв современного алфавита русского языка и наименования соответствующих клавиш раскладки русского языка «Русская» или их сочетаний (схема 1). Следует отметить, что представленные ниже таблицы включают не все знаки, используемые в ЦС языке. Часть из отсутствующих в данном обзоре знаков можно просмотреть через Таблицу символов операционной системы компьютера или текстового редактора в шрифте Panty, отсутствующие там иные необходимо программировать в шрифт из таблиц Юникода. Отсутствующие в Юникоде знаки можно добавлять, подавая заявку в консорциум Юникода или программировать в системе фонетической азбуки*.

Для понимания структуры таблиц рассмотрим, например, ЦС букву «а» (табл. 2), которая имеет шифр юникода – U+0430, называется «аз», соответствует букве русского языка «а», под титлом «ā» обозначает число 1, набирается путем нажатия на клавишу «а», соответствующую раскладке русского языка «Русская» (схема 1). Для набора заглавной «А» с шифром юникода – U+0410 используется комбинация клавиш Shift + «А». В раскладке в качестве дополнительной используются также правая клавиша AltGr или её аналог справа Alt + Ctrl, в зависимости от программного обеспечения компьютера. А так же используется сочетание клавиш AltGr и Shift.

Символ	Уникод	Число ЦС	Чтение	Название буквы ЦС	Соответств. буква рус. языка	Клавиши раскладки рус. языка «Русская»
А а	0410 0430	1	[а]	аз	А а	А а
Б б	0411 0431		[б]	буки	Б б	Б б
В в	0412 0432	2	[в]	вѣди	В в	В в
Г г	0413 0433	3	[г]	глаголь	Г г	Г г
Д д	0414 0434	4	[д]	добрѡ	Д д	Д д
Е е	0415 0435	5	[е]	есть	Е е	Е е
Ё ё	0404 0454		[е]	есть широкое (якорное)	Не исп.	Э э
Ж ж	0416 0436		[ж']	живѣте	Ж ж	Ж ж
З з	0405 0455	6	[дз']	зелѡ	Не исп.	Ё ё
З з	0417 0437	7	[з]	земля	З з	З з
И и	0418 0438	8	[и]	иже (8-ричное)		И и
Й й	0419 0439		[й]	иже краткое		Й й
Ї ї	04c0 0308 04cf 0308	10	[и]	иже (10-ричное) комбиниров.	Не исп.	Shift+1 1 (комбин. двух знаков)
К к	041A 043A	20	[к]	како	К к	К к
Ѧ ѧ	040B 045B			гервь (для транслит. глаголицы)	Не исп.	Shift AltGr+в а
Л л	041B 043B	30	[л]	люди	Л л	Л л
М м	041C 043C	40	[м]	мыслѣте	М м	М м
Н н	041D 043D	50	[н]	наш	Н н	Н н
О о	041E 043E		[о]	он	О о	О о
Ѡ ѡ	047A 047B	70	[о]	он широкое	Не исп.	Shift+0 0

Ω ω	0460 0461		[o]	oméга	Не исп.	Shift+6 6
Ϟ ϟ	A64C A64D		[o]	ОМéга дополняется в раскладке на одной клавише знаками 0486 0487	Не исп.	AltGr+c AltGr+m
Ϡ ϡ	047C 047D		[o]	oméга (торж.)	Не исп.	Shift+«-» «-»
Ϣ ϣ	047E 047F	800	[от]	от	Не исп.	Shift+7 7
Π π	041F 043F	80	[п]	поко́й	П п	П п
Ρ ρ	0420 0440	100	[р]	рцы	Р р	Р р
Ϛ ϛ	0421 0441	200	[с]	сло́во	С с	С с
Τ τ	0422 0442	300	[т]	твёрдо	Т т	Т т
Υ υ	0423 0443	400	[у]	ук	У у	AltGr+2 AltGr+3
Ου ου	0478 0479		[у]	оник	Не исп.	Shift+3 3
Ϛ ϛ	A64A A64B		[у]	ук	Не исп.	У у
Φ φ	0424 0444	500	[ф]	ферт	Ф ф	Ф ф
Χ χ	0425 0445	600	[х]	хер	Х х	Х х
Ц ц	0426 0446	900	[ц']	цы	Ц ц	Ц ц
Ϙ ϙ	0427 0447	90	[ч']	червь	Ч ч	Ч ч
Ш ш	0428 0448		[ш']	ша	Ш ш	Ш ш
Щ щ	0429 0449		[ш'т']	ща	Щ щ	Щ щ
Ъ ъ	042A 044A		[ъ]	ер	Ъ ъ	Ъ ъ
Ы ы	042B 044B		[ы]	еры́	Ы ы	Ы ы
Ь ь	042C 044C		[ь]	ерь	Ь ь	Shift+2 2
Ѧ ѧ	0462 0463		[æ], [ие]	ять		Shift+5 5
Э э	042D 044D			? есть мал.	Э э	AltGrShft+э AltGrShft+«/»

Ю ю	042E 044E		[йу]	ю	Ю ю	Ю ю
Я я	042F 044F			Не исп.	Я я	
Ў ѳ	0464 0465		[йэ]	Е йотиров. Редко исп.	Не исп.	AltGrShft+ь AltGrShft+б
Ѧ ѧ	A656 A657		[йа]	А йотированное	Не исп.	Shift+4 4
Ѩ ѩ	0466 0467		[эн]	Малый юс	Не исп.	Я я
Ѫ ѫ	046A 046B		[он]	Большой юс	Не исп.	AltGrShft+ф AltGrShft+ы
Ѭ ѭ	0468 0469		[йэн]	юс малый йотированный	Не исп.	AltGrShft+я AltGrShft+ч
Ѯ ѯ	046C 046D		[йон]	юс большой йотированный	Не исп.	AltGrShft+ю AltGrShft+«.»
Ѳ ѳ	046E 046F	60	[кс]	кси	Не исп.	Shift+8 8
Ѵ ѵ	0470 0471	700	[пс]	пси	Не исп.	Shift+9 9
Ѷ ѷ	0472 0473	9	[θ], [ф]	фита́	Не исп.	Shift+«/» «/»
Ѹ ѹ	0474 0475	400	[и], [в]	йжица	Не исп.	Shift+«=>» «=>»
Ѻ ѻ	0480 0481	(90) арх.		коппа	Не исп.	AltGrShft+c AltGrShft+m

Табл. 2. Кодирование букв церковнославянской азбуки и раскладки.

В ЦС языке существует ряд цифр, представленных отдельными знаками, в таблице представлена цифра, имеющая значение «одна тысяча» (табл. 3).

Символ	Уникод	Число	Клавиши раскладки рус. языка
Ѱ	0482	Одна тысяча (1 000)	AltGr+0

Табл. 3. Кодирование цифр церковнославянской азбуки и раскладки.

Знак	Юникод	Название	Соответствие	Клавиши раскладки рус. языка
,	002C	Запятая	,	Alt+«.,»

.	002E	Простая точка	.	«.»
к.	002E	Малая точка	;	«.»
к.	00B7	Средняя точка	.	AltGr+ë
:	003A	Двоеточие	(:), (;), (...)	AltGr+б
;	037E	Вопросительная	«;»	AltGr+7
!	0021	Удивительная	!	AltGr+1
*	A673	Славянская звездочка	*	AltGr+8
-	002D	Дефис В ЦС на базовой линии	-	AltGr+«-»
	00AD	Мягкий перенос. В ЦС на базовой линии		AltGr+«ш»
/	002F	Косая черта	/	AltGr+«\»
()	0028 0029	Круглые скобки	()	AltGr+я AltGr+ч
[]	005B 005D	Квадратные скобки	[]	AltGr+э AltGr+ж

Табл. 4. Кодирование знаков пунктуации церковнославянской азбуки и раскладки.

Особенностью ЦС языка является частое применение надстрочных знаков. Все употребляемые в современном ЦС языке надстрочные знаки представлены в таблице и закодированы в раскладке (табл. 5).

Знак	Юникод	Название	Клавиши раскладки рус. языка
/	0301	Оксия (острое ударение)	ь
\	0300	Вария (тяжелое ударение)	Ь
˘	0311	Камора или переспомен (облегченное ударение)	AltGr+5
ˇ	0486	Звательцо или псили (простое придыхание)	AltGr+п
̣	0485	Дасия (густое придыхание)	AltGr+ы

Ѡ	0486 + 0301	ИСО (придыхание + оксия) комбинация двух знаков, печатающихся в тексте нажатием одной клавиши	AltGr+a
ѡ	0486 + 0300	Апостроф (придыхание + вария) комбинация двух знаков, печатающихся в тексте нажатием одной клавиши	AltGr+v
Ѣ	030F	Смычек или кендема исп. для буквы ѡ.	AltGr+2
Ѥ	A66F	Взмет	AltGr+«г»
Ѧ	A67E	Кавыка для маркировки слова в его начале или конце (знак с местом в строке)	AltGr+«д»
ѧ	A67C	Кавыка для маркировка слова в его середине (диакритический знак над буквой)	AltGr+«щ»
Ѩ	0306	Краткая	AltGr+«.»
Ѭ	0484	Палатилизация (смягчение)	AltGr+«ь»
Ѯ	A67D	Паерок (замещение ѣ, как маркировка над иной буквой)	AltGr+«л»
ѯ	A67F	Паерок (замещение ѣ, как маркировка вместо буквы ѣ)	В табл.
Ѱ	0483	Простое титло	AltGr+«ф»
Ѳ	0487	Покрытие для буквотитлов	AltGr+«н»

Табл. 5. Кодирование надстрочных знаков церковнославянской азбуки и раскладки.

Со времен, Остромирова Евангелия, встречаются буквы титла, вынесенные над строкой буквы или их сочетания, покрываемые или не покрываемые особым знаком «покрытие». В раскладке представлены пять ныне часто используемых букв титлов. Одной клавишей раскладки они печатаются в тексте с дополнительно запрограммированным знаком «покрытие» U+0487. Клавишей «Bksp» покрытие, при использовании шрифта «Panty», может удаляться, далее так же и сама буква. Остальные буквотитлы находятся по таблице шрифта и печатаются с дополнительным покрытием или в разновидности исторической формы – без покрытия.

Знак	Уникод	Название	Комб. знак	Уникод	Раскладка рус. языка
Ѱ	2DE2	г	Ѱ	U+2DE2 U+0487	AltGr+«у»
Ѳ	2DE3	д	Ѳ	U+2DE3 (совмещен в одном знаке)	AltGr+«к»

о	2DEA	о	ⲟ	U+2DEA U+0487	AltGr+«й»
р	2DEC	р	Ⲡ	U+2DEC U+0487	AltGr+«ц»
с	2DED	с	ⲡ	U+2DED U+0487	AltGr+«е»
х	2DEF	х	Ⲣ	U+2DEF (совмещен в одном знаке)	В таблице
з	2DE5	з	ⲣ	U+2DE5 (совмещен в одном знаке)	AltGr+«з»

Табл. 6. Кодирование букв титлов современной церковнославянской азбуки и раскладки.

знак	юникод	название	
Ⲟ	2626	Православный крест	AltGr+«р»

Табл. 7. Кодирование знаков церковнославянской раскладки.

Особенности кодирования буквы «и» десятичной

Буква «и» десятичная в ЦС текстах со времени возникновения ЦС азбуки встречается в трёх основных формах: без точки «и», с точкой «і» и с двумя точками сверху «ї». При этом в каждом пятом случае точка или две точки заменяются надстрочными знаками. Во избежание путаницы в кодировании данной буквы и в дальнейшей электронной обработке всего объема текстов ЦС языка целесообразным является использование одного кода для данной буквы во всех случаях общего стандарта ЦС языка всей истории его развития «і» 04CF. С последующим добавлением в тексте необходимого надстрочного знака. В том числе, особого компонента буквы, точки или двух точек, выполняющих в данном случае функцию надстрочного знака, предназначенного для лучшего опознания (читаемости) «тощей» буквы-палочки, приданием ей точкой или двумя точками дополнительной высоты.

Вопросы дальнейшего обсуждения

Наряду с тем, что проведенные предварительные испытания позволяют сделать вывод о соответствии в основном требованиям и к кодированию ЦС азбуки, и к созданному на её основе шрифту, необходимо отметить, что при этом существует ряд проблем, требующих решения. Эти проблемы подразделяются на две связанные группы, области филологии и области программирования азбуки, относящиеся к стандартизации (приведению к определённой норме) ЦС языка и к программному кодированию букв и знаков.

К основным проблемам программирования относится необходимость определения и стандартизации кодирования букв: десятичной «и», прописной якорной «Е», а также знаков пунктуации: малой точки, удивительной и иных дополнительных.

1) Буква «и» десятичная – «ї» в ЦС текстах со времени возникновения ЦС азбуки встречается в трёх формах: без точки, с одной точкой и с двумя точками сверху. При этом в каждом пятом случае точка или две точки заменяются надстрочными знаками. Если запрограммировать ї 0457 с двумя точками в одном буквенном знаке, тогда этими точками в сочетании с надстрочными знаками ударениями и титлами будет искажаться должный вид текста. В ряде программ имеется функция OpenType, автоматически программно заменяющая «і» с двумя точками на необходимую палочку при сочетании с надстрочными знаками, но OpenType применяется не во всех программах.

Эта проблема решается использованием для кодирования «ї» двух знаков: палочки «і» 04CF и верхнего

диакритического знака, двух точек « ѣ » 0308, не имеющего своего места в ряду, но располагающегося над предыдущей буквой. В случае сочетания «ї» с иными надстрочными знаками «две точки» просто не печатаются.

Необходима экспертиза и стандартизация.

2) Якорное «є» 0454 употребляется в словах ЦС языка, заменяя обычную «е» 0415, при этом несет на себе информационную нагрузку в области числа и склонения слова. Логически отсюда вытекает, что при наборе прописными буквами должно применять её пару «Є» 0404. Но такой особой буквы нет в ЦС языке. А в текстах применяется заглавная пара к «е» 0454 – «Е» 0415. Если при этом использовать в ЦС тексте из заглавных букв букву «Е» 0415, то в определённых случаях неизбежно будет искажен смысл текста.

Решение проблемы видится в маркировке дополнительным знаком графемы заглавной буквы «Є» 0404, для её отличия от заглавной «Е» 0415.

Необходима экспертиза и стандартизация.

3) Малая точка отличается от обычной тем, что следующее предложение (всегда в середине абзаца) пишется со строчной, а не с заглавной буквы. Анализ текстов показывает, что применение строчной буквы вместо заглавной определяется не впереди стоящей точкой, а особенностями построения текста ЦС языка. Точка всегда одна и та же.

Решение проблемы не в кодировании особой точки, которой таковая не является, а в понимании особенностей построения и обозначения начала предложений в середине абзаца ЦС языка. А также в определении нормы современного ЦС языка, потому что в части допущенных ИС РПЦ к публикации текстов эта строчная заменяется прописной, в соответствии с современной пунктуацией.

Необходима экспертиза и стандартизация.

4) Вопросительный знак «?» русского языка в ЦС языке «удивительная» обозначается вертикально расположенными точкой и запятой «;». Проблема возникает при автоматизации программными средствами перевода ЦС текста на русский язык и при конструировании системы программирования шрифта, где этот знак также имеется, но с другим значением.

Решение проблемы заключается в использовании в качестве «удивительной» знака не русского и латинского «;» 003В, а такого же вида, но из другой области и с иным кодом «;» 037Е. Что необходимо учитывать при программировании шрифтов и раскладок, а также процессе создания иных программ.

Необходима экспертиза и стандартизация.

Решение указанных проблем позволит сделать еще один шаг к увеличению автоматизации компьютерной обработки ЦС текстов, таких например, как проверка правописания, поиск, использование словарей и иного. Что необходимо в области образования и науки, а также востребовано и в иных сферах, связанных с текстовой информацией на ЦС языке.

Е-почта для замечаний и предложений edit@vcsl.dn.ua, csl@ukrpost.ua

Калашников Ю.В. Церковнославянская азбука. Кодирование CU. Краткое описание. [Электронный ресурс] // Электронные материалы журнала «Праведность и мир». ISSN 2078-8932. URL: <http://vcsl.dn.ua/patr/soft/font/abc-cu.pdf> (дата обновления: 2016.11.18).