

Стандардизација
старословенског
ћириличног
писма и његова
регистрација у
Уникоду

УВОД

Конзорцијум Уникод је непрофитна организација, која је основана у циљу развоја, ширења, промовисања и коришћења УНИКОД-СТАНДАРДА који одређује приказ текста у савременим компјутерским програмима. Уникод-стандард је прихваћен од водећих светских информатичких компанија као што су: Apple, HP, IBM, Just System, Microsoft, Oracle, SAP, Sun, Sybase, Unisys и многих других. Сваки језик, односно његово писмо, има своју кодну страну (кодни систем), на којој се налазе сва слова, цифре и интерпункцијски знаци, карактеристични за то писмо. До појаве Уникода кодни системи су били у колизији: различити системи су користили или исто место (код) за различите знаке, или различито место (код) за исте знаке. Није била могућа конверзија међу различитим системима. Уникод-стандард користи јединствени код за сваки карактер (слово, цифру, знак интерпункције и неки други симбол) без обзира на платформу (оперативни систем рачунара), компјутерски програм и језик. Уникод-стандард је постао општеприхваћен и у њему су данас регистровани скоро сви језици света, односно њихова писма.

Када је у питању ћирилица, Уникод је регистровао комплетну савремену ћирилицу (покрива сва писма свих народа који је користе) и само она старословенска слова која се не налазе у савременој ћирилици [17 курентних (Ѡ, Ћ, Ћ, Љ, Љ, Њ, Џ, Џ, Џ, Ѡ, Ѡ, Ѡ, Ѡ), 17 верзалних (Ѡ, Ѓ, Ѓ, Љ, Љ, Њ, Џ, Џ, Ѡ, Ѡ, Ѡ, Ѡ), 4 дијакритичка знака, 3 симбола за цифре)– види на www.unicode.org – U0400 Cyrillic] Она, у ствари, припадају грађанском (нецрквеном) писму (такозвана „гражданка“), које је модификовано у 18. веку (Уникод-кодови од 0460 до 0481). Зато слово „а“ из савремене ћирилице и слово аз „а“ из старословенског имају исти Уникод код (0430). То је случај и са осталим словима која се поклапају. Последица овога јесте та да се у истом фонту не могу истовремено имати савремена и стара ћирилица. **СТАРΟΣЛОВЕНСКО ПИСМО КОЈЕ БИ ДОДАТНО ПОКРИВАЛО И НАЦИОНАЛНЕ РЕДАКЦИЈЕ ЗАПРАВО УОПШТЕ НИЈЕ РЕГИСТРОВАНО.** Коришћење садашњег решења за примену на старословенско писмо, чак и за најпростију употребу, није могуће због разних проблема:

- Од стандардних слова недостају: Ѓ, 8, Ц, Ѓ, Ѓ, а такође и ређе коришћена: α, д, ε, ι, λ, μ, η, τ, ѱ и недефинисана ѡ, ѣ, ѣ, ѣ
- Имена (у савременој и старој ћирилици) и облици слова су различити.
- Нека имају исти облик, старо ћерв „ћ“ и савремено „ћ“ али различит изговор.
- Нека имају различит облик, старо „ј“ и савремено „ј“ али исти изговор. Додатна конфузија је што старо „ј“ није настало од „ј“ већ од малог јуса „ј“, па му је место у старој и новој азбуци различито.
- Нека од њих су узбучена другачије у савременој и старој ћирилици па је сортирање погрешно.

За потпуну и удобну примену старословенског писма потребно је регистровати многобројна слова, лигатуре, надметнута слова са титлом и без ње, велики број интерпункцијских и дијакритичких знакова и све старословенске цифре. Пошто то није регистровано, односно пошто не постоји стандард, сви постојећи фонтови су рађени у „сопственом“ кодном систему, који је различит од других. Зато је основни проблем некомпатибилност у употреби старословенских фонтова у различитим срединама, што спречава читање и коришћење старих текстова у електронској форми ван средине у којој су настали. Интернет, односно међународна комуникација готово је онемогућена.

ПРИНЦИПИ

1. Писмо неког језика је потпуно регистровано у Уникоду ако, и само ако су сви атрибути тог писма регистровани. У случају старословенског писма, то значи:

а) Сва мала и велика слова (не варијанте писања истог слова, односно глифови)

Иако Уникод не региструје глифове већ слова, другачији је случај са старословенским. Писање глифова, чак и у истом реду, је био стандардни начин писања (на пример: „о“, „о“, „Ѡ“, или „а“ и „Ѡ“, или „т“ и „т“, или „оѣ“, „8“ и „8“ итд.). То није био изузетак већ правило писања (види пример писања). Због тога предлагемо проширену азбуку у нашем предлогу. Остатак слова треба сматрати глифовима који припадају делу Уникода за приватну употребу. На тај начин

би и најсложенији старословенски текст био Уникод–текст, потпуно компатибилан са свим фонтовима који поштују стандард (Уникод).

- б) **Сва надметнута слова са титлом и без титле** (писана изнад слова или између слова) за мала и велика слова у складу са решењем No2 (види Композитна слова).
- ц) **Сви надредни знаци** за мала и велика слова и знаци интерпункције у складу са решењем No2 (види Композитна слова)
- ц) **Све лигатуре** за мала и велика слова (види на пример www.unicode.org – UFB00 Ligatures for Latin, Armenian and Hebrew Ligatures)
- д) **Све цифре** (види на пример www.unicode.org – U10140-Ancient Greek Numbers)

2. Редослед Уникод–кодова треба да буде у складу са азбуком због сортирања.

Имајући у виду принципе, није могуће извршити додавање на оно што је до сада регистровано. Потребно је извршити ново регистровање из два разлога. Први разлог је тај што у садашњој регистрацији нема места за убацивање великог броја нових кодова. Други разлог је тај што, чак и да има места, постојећи ћирилични кодови би морали потпуно да се преуреде како би се омогућило сортирање. Чињеница је да се сада не може мењати редослед јер је ћирилична кодна страна већ у широкој употреби.

Из свега наведеног следи да је потребно:

1. **Утврдити стандард за старословенско писмо (тачно дефинисати сва слова, надметнута слова, надредне знаке, лигатуре и цифре које репрезентују старословенски језик)**
2. **Утврдити азбуку (ако постоје различита мишљења о томе)**
3. **Такав заједнички предлог, у име свих институција чији представници буду учествовали на конференцији, упутити Уникоду са захтевом да старословенско писмо треба регистровати одвојено и независно од садашње ситуације како је регистрована ћирилица.**

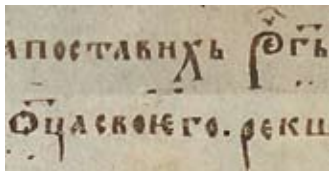
У том случају можемо регистровати старословенско писмо у потпуности, у складу са оба принципа, и можемо имати савремену и старословенску ћирилицу у истом фонту.

Према нашем предлогу, имајући у виду принципе за регистровање писма, примере писања и принципе израде композитних слова, потребно је регистровати 1.087 карактера (види прилог). Глифови (у жутим пољима) и симболи треба да се кодирају у делу Уникода који је за приватну употребу (Private Use Area), о чему треба такође направити интерни стандард како би старословенски фонтови били компатибилни.

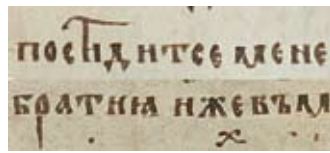
Преглед карактера за регистровање

Курентна слова	слова	глифови
Слова	75	234
Надметнута слова без титле (писана изнад слова)	65	34
Надметнута слова без титле (писана између два слова)	65	34
Надметнута слова са титлом (писана изнад слова)	65	34
Надметнута слова са титлом (писана између два слова)	65	34
Дијакритички знаци и титле (писана изнад слова)	68	
Титле (писане између два слова)	13	
Лигатуре	85	7
Верзална слова		
Слова	71	121
Надметнута слова без титле (писана изнад слова)	65	34
Надметнута слова без титле (писана између два слова)	65	34
Надметнута слова са титлом (писана изнад слова)	65	34
Надметнута слова са титлом (писана између два слова)	65	34
Дијакритички знаци и титле (писана изнад слова)	68	
Титле (писане између два слова)	13	
Лигатуре	111	26
Цифре	63	30
Тотални број јединствених карактера	1.087	690

Пример писања:



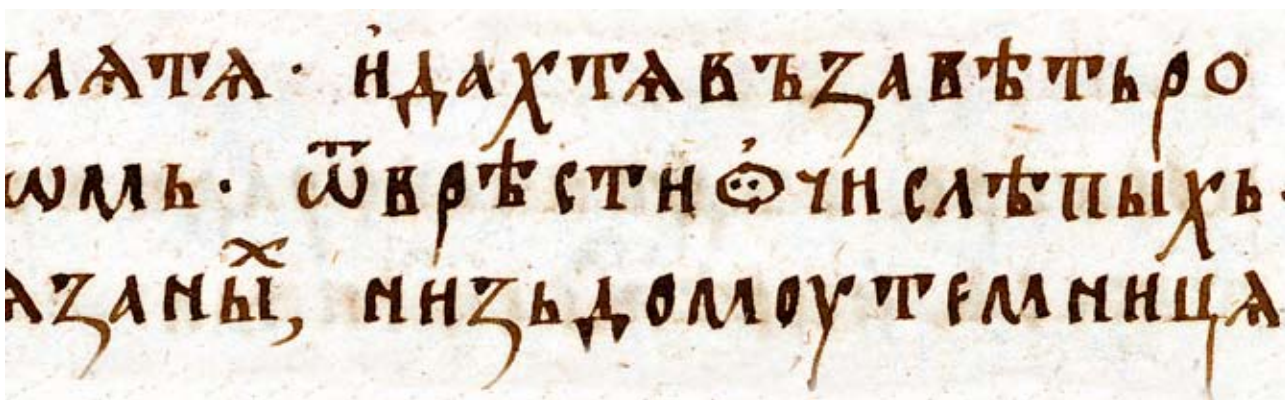
Лигатура „ре“ **ⱪ**, слова „ре“ **ре**.



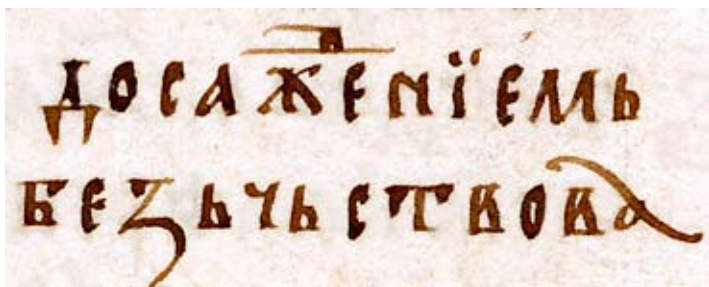
Лигатура „ти“ **Ѡ**, слова „ти“ **ТИ**.



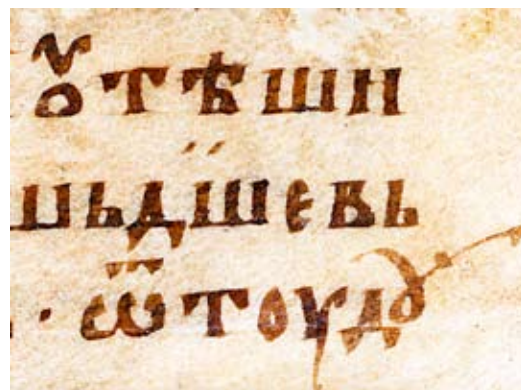
Слово „он широко 2“ **Ɱ**, „он широко“ **О**, „он“ **Ɱ**.



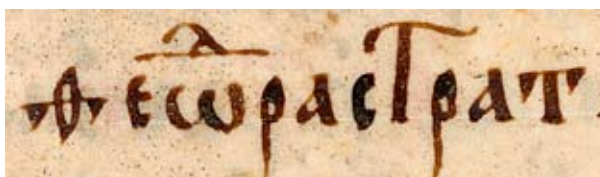
Слово „он широко“ **О**, „он двооко“ **Ɱ**, „он“ **Ɱ**.



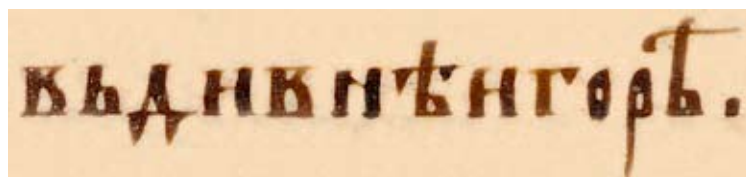
Слово „аз“ **А**, „алфа 5“ **λ**.



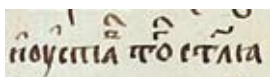
Слово „ук“ **Ѡ**, „оник“ **Ɱ**, „ук два 2“ **Ѡ**.



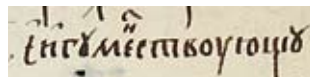
Слово „тврдо високо“ **Ѡ**, „тврдо“ **Т**.



Слово „јат“ **Ѣ**, „јат високи“ **Ѣ**.



Слово „тврдо 2“ **Ш**, „тврдо 3“ **П**.



Слово „јота“ **Ѡ**, „иже“ **Н**.

Композитна слова

Да би се направило слово са дијакритичким знаком изнад њега (композитни карактер) постоје две могућности. Показаћемо то на примеру конструкције курентног слова *udieresis* “ü” и верзалног *Udieresis* “Û”.

1. У Уникоду имамо дефинисано слово “u” са кодом (0075), дијакритички знак дијерезис “” са кодом (00A8) и композитно слово удијерезис “ü” са кодом (00FC). Дијакритички знак нема неки одређен положај, већ га типограф у композитном слову тачно подешава у односу на ширину и висину слова.

За верзално композитно слово Удијерезис “Û” са кодом (00DC), користимо слово “U” са кодом (0055) и исти дијакритички знак дијерезис “” са кодом (00A8) као и у случају курентног слова. Типограф га поново тачно подешава за нову ширину и висину.

Видимо да у оба случаја није битан положај дијакритичког знака у свом пољу (00A8) већ да тај положај одређујемо у композитном слову.

Сва остала слова са дијакритичким знацима се конструишу на исти начин. Када куцате, куцате директно “ü” у једном потезу. Последица овог приступа је да за сва слова са дијерезисом имате два кода, односно два места у фонту (слово и композитно слово) и само један код (место) за сам дијерезис. То значи да вам за 12 слова (ä, ë, ï, ö, ü, ÿ, Ä, Ë, Ì, Ö, Û, ÿ) треба 25 Уникод кодова, односно места у фонту.

u	”	ü	U	”	Û
0075	00A8	00FC	0055	00A8	00DC

2. Други приступ је упрошћен и са типографског становишта лошији. Као и у горњем случају имамо слово “u” са кодом (0075) и дијакритички знак дијерезис “” са кодом (00A8), али немамо композитно слово. Уместо тога када правимо дијерезис позиционирамо га тачно за висину курентног слова, стављамо да му је ширина нула и померамо га улево како би дошао изнад слова. Када куцамо, куцамо прво “u” па онда дијерезис које ће се појавити тачно изнад “u” и добијемо “ü”.

За верзал имамо слово “U” са кодом (0055) али сада морамо узети други дијакритички знак Дијерезис “” са кодом (F6CB). Дијерезис сада подешавамо за висину верзалног слова, стављамо ширину нула и померамо га улево како би дошао изнад “U”. Када куцамо, куцамо прво “U” па онда дијерезис које ће се појавити тачно изнад “U” и добијемо “Û”.

У оба случаја, ако смо добро подесили положаје дијакритичких знакова, добићемо добар резултат. Тако за истих 12 слова (ä, ë, ï, ö, ü, ÿ, Ä, Ë, Ì, Ö, Û, ÿ) треба 14 кодова, односно места у фонту, 12 за слова и 2 за дијакритичке знакове, што је значајно мање него у првом случају (25).

У чему је проблем оваквог приступа. Како смо позиционирали положај дијерезиса за ширину слова “u” а постоји само један дијерезис, то ће положај дијерезиса за слово “i” које је уже бити погрешан.

u	+	”		=	ü		U	+	”		=	Û	but for idieresis	l	+	”		=	ï	
0075		00A8					0055		F6CB						0131		00A8			

Да закључимо:

- Решење No1 је типографски оправдано и сви романски језици као и савремена ћирилица су урађени на тај начин. У случају старословенског писма то би, због многобројних надметнутих слова и надредних знакова, захтевало огroman број кодних места (више десетина хиљада).
- Решење No2 је типографски мање оправдано али захтева знатно мање кодних места.
- Решење No3 не постоји.