

# Vnašalni sistem ZRCola

Peter Weiss

Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU  
Novi trg 2, 1000 Ljubljana  
Peter.Weiss@guest.arnes.si

Vnašalni sistem za dialektološke in za druge jezikoslovne (npr. etimološke in jezikovnozgodovinske) potrebe, imenovan ZRCola, ki deluje v programu word v operacijskem sistemu windows, je skupek pisave, temelječe na standardu unikod in razširjene z dodatnimi latiničnimi in ciriličnimi črkovnimi znaki, ter makrov in dokumentacije. V dialektološki sekciji Inštituta za slovenski jezik Frana Ramovša Znanstvenoraziskovalnega centra Slovenske akademije znanosti in umetnosti (ZRC SAZU) v Ljubljani sem ga razvil za potrebe Slovanskega lingvističnega atlasa (s kratico OLA) – v njem se pišejo dialektološki znaki v latinici, spremljevalna besedila pa tudi v cirilici –, za izvedbo drugih nalog predvsem slovenske in tudi drugih slovanskih dialektologij ter nasploh za vnašanje jezikoslovnih besedil, ki včasih zahtevajo najrazličnejše znake.

ZRCola odpravlja dve zadregi, ki sta doslej oteževali predvsem pisanje in urejanje zapletenih narečnih besedil v pri nas najbolj razširjenem urejevalniku besedil.

(1) V ZRColi se da sestavljene znake, tj. tiste z ločevalnimi (diakritičnimi) sestavinami, vnašati tako, da za osnovnim znakom večinoma s kombinacijami tipk na programske izpopolnjeni tipkovnici vnesemo – zaporedje je obvezno – najprej ločevalne znake, ki se izvorno pišejo nad osnovnim in se, če jih je več, kopičijo navzgor, potem tiste pod osnovnim, ki se kopičijo navzdol, in nazadnje še ločevalne znake ob osnovnem ali zapisane čezenj:  $a\acute{u} \rightarrow \acute{a}$ ,  $a\acute{u}\grave{u} \rightarrow \acute{a}$ ,  $c\check{c} \rightarrow \check{c}$  (za tiste tuje uporabnike, ki na tipkovnici nimajo šumevcev),  $c\grave{c} \rightarrow \grave{c}$ ,  $d\check{d} \rightarrow \check{d}$ ,  $L\grave{L} \rightarrow \grave{L}$ ,  $o\acute{u}\grave{u} \rightarrow \acute{o}$ ,  $a\acute{u}\grave{u} \rightarrow \acute{a}$ ,  $=\# \rightarrow \neq$ . (Na podoben način, le da v drugačnem zaporedju, se da v wordu že zdaj posredno vnesti npr. znaka  $\acute{a}$  ali  $\check{c}$ .) Na kateri koli stopnji vnašanja besedila lahko znake sestavimo (in sestavljene seveda razstavimo) in jih take natisnemo ali npr. oddamo v tisk. S podobnimi sorazmerno preprostimi mehanizmi vnašamo npr. tudi ligature z ločevalnimi znaki vred:  $A\grave{u}E\acute{u} \rightarrow \grave{A}E$ .

Razstavljeni znaki, kot jih med delom vidimo na zaslonu, so koristni zato, ker pri napačnem vnosu predzadnjega znaka v nizu recimo šestih ni treba ponoviti vnašanja celotnega niza, ampak spremenimo le napačni znak. (Primer takega kompleksnega znaka je  $i\grave{u}\grave{u}e\acute{u}\grave{u} \rightarrow \grave{i}\grave{e}$ .) Tudi iskanje po nizih nesestavljenih znakov utegne biti lažje, npr. vseh pojavitev ostrivca zgoraj in pike spodaj za

osnovnim znakom, pri čemer je kdaj celo koristno, ker ni treba poznati osnovnega znaka:  $e\acute{u}\grave{u} \rightarrow \acute{e}$ ,  $o\acute{u}\grave{u} \rightarrow \acute{o}$ ; iskanje po sestavljenih znakih z določenimi značilnostmi, kot v tem primeru, je lahko veliko napornejše. V programu word je privzeto, da se da nekatere črke z ločevalnimi znaki sestaviti v en sam novi znak ali pa na videz, tako, da program npr. sestavljivo piko premakne pod osnovni znak, uporabnik pa mora izbrati eno izmed dobrega ducata pik, ki se razlikujejo le glede na širino znaka, pod katerim naj bi stale – zanesljivo iskanje po tako vnesenem besedilu je nemogoče in tak način vnašanja znakov za resno delo ni primeren. Vnašalni sistem ZRCola ponuja eno samo piko ( $\#$ ), ki nastopa v kombinacijah z drugimi ločevalnimi znaki in se na uporabnikovo zahtevo sestavi z osnovnim znakom a v  $\grave{a}$ ,  $\acute{a}$ ,  $\grave{a}$ ,  $\check{a}$ ,  $\grave{a}$ ,  $\check{a}$ ,  $\grave{a}$  ali  $\grave{a}$ .

(2) V pisavi 00 ZRCola so poleg vseh unikodnih latiničnih znakov upoštevani še vsi znaki iz cirilične, grške, hebrejske in arabske pisave, mednarodna fonetična abeceda ter velika količina simbolov in ločil iz unikoda. Na področju zasebne rabe (angl. private use area) unikoda, kjer je trenutno v navedenem programskem okolju in programu prostora za več kot 6000 znakov, pa je vse manjkajoče s standardiziranega področja unikoda, kar lahko dodajo posamezniki ali skupine za svoje potrebe: uvedeni so dejansko rabljeni dodatni sestavljeni znaki in dodatne latinične in cirilične črke za dialektološke potrebe, poleg njih pa doslej v računalniških naborih manjkajoči sistemsko prikazani slovenski zgodovinski znaki iz bohoričice, dajničice, metelčice in stare prekmurske knjižne pisave, potem slovenski zgodovinski narečni in drugi jezikoslovni znaki, osnovni metrični in še nekateri dodatni tehnični znaki ter posebna ločila. S tem je v ZRColi omogočeno tudi vnašanje starejših slovenskih ne samo jezikoslovnih del, tako da so na zasebnem področju na razpolago tudi različice istopomenskih znakov, npr. v dajničici. To sicer pomeni odmik od same zamisli unikoda, v katerem naj ne bi bilo različic istovsebinskih znakov: v dajničici se recimo pri znakih  $\mathfrak{S}$ ,  $\mathfrak{X}$  in  $\mathfrak{8}$  – za današnji  $\mathfrak{S}$  – ne da reči, kateri je »pravi« in katera dva sta izvedenki iz njega, saj so se zaradi različno opremljenih tiskarn sredi prve polovice 19. stoletja v knjigah uporabljali vsi trije.

Potreba ne samo dialektološke stroke po oblikovanju podatkovnih zbirk in po preprosti izmenljivosti vnesenih podatkov zahteva zanesljive

rešitve, ki jih v računalniških programih in tudi v unikodu še lep čas ne bo, čeprav so obljubljene in jih napovedujejo (npr. samodejno sestavljanje znakov), vendar nanje ne moremo čakati. Od standardnega unikodnega nabora, ki je sicer obeta-ven, tudi ne moremo pričakovati vsega, kot kažejo primeri od U+0200 do U+0217 iz unikoda (Ă ă Â â ... Ę ę Ħ ħ Ũ ù Ū ū), ki so vanj uvrščeni, kot je izrecno navedeno, tudi zaradi slovenščine, v celoti pa pokrivajo le rabo v hrvaščini in srbsščini. V unikodu manjkajo slovenski posebneži e in o s cirkumfleksom zgoraj in piko spodaj in na isti ravni (namreč pri zapisu tonemskega naglasa v slovenskem knjižnem jeziku) vsaj še e in o z ostrivcem zgoraj in piko spodaj ter polglasnik s krativcem zgoraj in z dvojnimi krativcem zgoraj. Marsikateri znak mora biti na razpolago iz zgodovinskih razlogov, recimo za ponatiskovanje starih del in za njihovo upoštevanje v podatkovnih zbirkah.

S sestavljanjem se da v ZRColi trenutno vnašati nad 1800 znakov od skoraj 4000 vseh, kolikor jih je zajetih v njej tako s standardiziranega področja kot s področja zasebne rabe unikoda.

Sistem ZRCola, ki je zastoj (http://ZRCola.zrc-sazu.si), se bo razvijal. Mesta znakov z zasebnega področja se (zaradi združljivosti za nazaj) ne bodo spreminjala, novi znaki pa se bodo dodajali na predlog uporabnikov. Iskali bomo hitrejšo in preprostejšo programske rešitve in v delu so programske prilagoditve

tipkovnic za vse slovanske in še nekatere druge jezike. Pripravlja se tudi popolnejša dokumentacija s potrebnimi navodili in nasveti za vnašanje (recimo o narekovajih, o velikih začetnicah v narečnih zapisih, o izbiri znaka za polglasnik v slovenskih narečnih besedilih ter o pomišljajih in vezaju), v katero bi bilo zaželeno vključiti slovenski prevod vsaj tistega dela unikoda, ki je vključen v ZRColo, z ustreznim opisom podstandarda s področja zasebne rabe.

## Literatura

Primož Peterlin – Aleš Košir – Tomaž Erjavec, Digitalni zapis slovenskih znakov. V: Tomaž Erjavec – Jerneja Gros (ur.), *Jezikovne tehnologije za slovenski jezik – Language Technologies for the Slovene Language: Zbornik konference – Proceedings of the Conference*. Ljubljana: Institut Jožef Stefan, 1998, str. 128–132.

*Slovenski pravopis 2001*. Ljubljana: SAZU – ZRC SAZU (izd.), Založba ZRC (zal.), 2001.

The Unicode Consortium, *The Unicode Standard: Version 4.0*. Boston itd.: Addison-Wesley, 2003.