

Cahiers **GUT** *enberg*

© RUSSIAN T_EX

© Basil MALYSHEV, Alexander SAMARIN, Dimitri VULIS

Cahiers GUTenberg, n° 10-11 (1991), p. 1-6.

<http://cahiers.gutenberg.eu.org/fitem?id=CG_1991__10-11_1_0>

© Association GUTenberg, 1991, tous droits réservés.

L'accès aux articles des *Cahiers GUTenberg*

(<http://cahiers.gutenberg.eu.org/>),

implique l'accord avec les conditions générales

d'utilisation (<http://cahiers.gutenberg.eu.org/legal.html>).

Toute utilisation commerciale ou impression systématique

est constitutive d'une infraction pénale. Toute copie ou impression

de ce fichier doit contenir la présente mention de copyright.

Russian T_EX

Basil MALYSHEV[†], Alexander SAMARIN[†], and Dimitri VULIS^{*}

[†] *Institute for High Energy Physics, 142284, Protvino, USSR*

samarin@m9.ihep.su, malyshev@m9.ihep.su

^{*} *CUNY Graduate Center, NY, USA*

dlv@cynuvms1.bitnet

Abstract. This article presents the T_EX extension for processing the russian texts. Russian T_EX is based on version 3.0 and virtual fonts. The different coding schemes for russian characters are allowed.

Résumé. *Cet article présente une version étendue de T_EX permettant de traiter des textes russes. Russian T_EX est construit sur une base T_EX version 3.0 et utilise le mécanisme de fontes virtuelles. Divers schémas de codage des caractères cyrilliques sont disponibles.*

Key words: russian language, russian, virtual fonts.

For russian texts¹ processing by T_EX one should adjust T_EX to:

- russian language hyphenation,
- coding of the russian characters,
- and fonts with the cyrillic symbols.

T_EX 3.0 can be adjusted without changes!

The hyphenation patterns have been described in [Vulis89] are used for russian language. Actually T_EX is bilingual – the russian and english hyphenation patterns are loaded by the following file:

```
\language=0          % English
\lefthyphenmin=2
\righthyphenmin=3
\input ehyphen.tex  % patterns for English language
\language=1          % Russian
```

¹The first usage of cyrillic characters into T_EX have been described in [Beeton85].

```
\lefthyphenmin=2
\righthyphenmin=2
\input cyrdef.tex
\input rhyphen.tex % patterns for Russian language
\language=0        % English as default
```

In the file `cyrdef.tex` proper `catcode`, `uccode`, `lccode` and `mathcode` are set for cyrillic characters.

The switching between russian and english hyphenation is performing by primitive `language`: the setting `\language=0` means english, the `\language=1` means russian. English words are not hyphenated if the russian hyphenations are active and reversely. Another possibility is to merge the english and russian hyphenation patterns as a single language.

\TeX can use any 8-bits coding scheme for russian characters – “alternative”², KOI-8³, ISO 8859-5⁴ etc. The hyphenation patterns and `.tfm` files should correspond to coding scheme being used. Russian \TeX works with virtual fonts, each of them consists of an original Computer Modern font (below 128-th code) and a font with cyrillic characters (above 128-th code). For \TeX the cyrillic characters are completely equal in “rights” with latin characters. One could define new commands as russian words! The simultaneous usage of cyrillic and latin characters does not require any additional commands for switching or separating of them.

For creating `.tfm` file of a virtual font the program `TFMerge` was designed. It merges the `.tfm` for Computer Modern font and `.tfm` for cyrillic font into a virtual font `.tfm` and `.vf` in accordance with the local coding scheme of the cyrillic characters.

The managing of the program `TFMerge` is performed by the file which contains the table of correspondence between the position of a character in a real font and the position of the same character in a virtual font. E.g. the correspondence between alternative coding into virtual fonts and phonetic-like coding (with swapped lower/upper case) into cyrillic fonts[Glonti90] is expressed by following triples:

```
64 238 1 % 0 - > yu
```

²Used mainly into IBM PC, alphabetically ordered, almost identical to Microsoft's codepage 866.

³Used into some UNIX-like systems, based on “phonetic” correspondence between latin and cyrillic characters.

⁴Used into VAX/VMS, alphabetically ordered.

```

65 160 1   % A - > a
66 161 1   % B - > be
67 230 1   % C - > tse
68 164 1   % D - > de
69 165 1   % E - > ie
70 228 1   % F - > ef
71 163 1   % G - > ghe
72 229 1   % H - > ha

```

The first decimal digit is the position into a real font, the next is the position into a virtual font and the last is the counter which defines how many characters must be copied from real to virtual fonts by increasing the previous numbers.

If Computer Modern font is copied fully to a virtual font into the same positions, then only one triple is enough:

```
0 0 128   % all 128 characters
```

The calling sequence of TFMerge program is:

```
TFMerge  xcmr10  cmr10/t=all128  cmcyr10/t=cyralt
```

where the first parameter is the name of the virtual font and the other parameters are the names of the “real” fonts. The virtual font `xcmr10` is constructed from Computer Modern font `cmr10` and cyrillic font `cmcyr10`. Switch `/t` defines the file which contains the correspondence table for given real font.

Note that such merging is correct because cyrillic fonts are created by METAFONT with the same setup files, like `cmr10.mf`. The font parameters are identical for latin and cyrillic fonts.

The files are merged by following couples:

```

cmbx*      cmcbx*      xcmbx*
cmbxsl10   cmcbxsl10   xcmbxsl10
cmbxti10   cmcbxti10   xcmbxti10
cmbxsl10   cmcbxsl10   xcmbxsl10
cmbxti10   cmcbxti10   xcmbxti10
cmmi5      cmcyr5      xcmmi5

```

cmmi6	cmcyr6	xcmmi6
cmmi*	cmcti*	xcmmi*
cmmib10	cmcbx10	xcmmib10
cmr*	cmcyr*	xcmr*
cmsl*	cmcsl*	xcmsl*
cmslt10	cmcslt10	xcmslt10
cmss*	cmcss*	xcmss*
cmssbx10	cmcssbx10	xcmssbx10
cmssdc10	cmcssdc10	xcmssdc10
cmssi*	cmcssi*	xcmssi*
cmti*	cmcti*	xcmti*
cmtt*	cmctt*	xcmtt*

Mathematical Italic fonts `cmmi*` are merged with Cyrillic Text Italic fonts and proper `mathcode`'s have been set. One can use Russian letters in math.

For use another realization of the cyrillic fonts one should create the table of the correspondence for `tfmerge` and select the another couples for merging.

File `plain.tex` for \TeX and file `lfonts.tex` for \LaTeX should be changed to substitute the references to the latin font been merged by the references to the proper virtual font. Note that, file `fontdef.tex` from "New Font Family Selection" by Frank Mittelbach and Rainer Schöpf is more convenient for such changes.

For VAX/VMS realization, the creating `.fmt` files is required only the setting a parameter `trie_size` to 16000. For SB30TEX (on MS-DOS) some `.tfm` files are not preloaded, because the size of the `.tfm` files is increased. For $\text{Em}\TeX$ [3a] the options `-i -o -8 -mt:12700` have been set.

The main problem with virtual fonts is that not all `.dvi` drivers can handle the virtual fonts. To avoid that problem the program `PosTeX` has been designed. It reads a `.dvi` file, expands the virtual fonts and writes new `.dvi` file, which is accepted by any `.dvi` driver.

There are some `.dvi` drivers which already accept the virtual fonts, e.g. from fine collection by Eberhard Mattes. One can use such `.dvi` drivers without `PosTeX`. Another reason to use the program `PosTeX` is the portability of `.dvi` files. Our virtual fonts refer to local coding of russian characters and immediately after \TeX ing a `.dvi` file is not portable. Being transforming by `PosTeX` a `.dvi` file becomes portable – it refers only to real fonts.

It's also possible to enter Russian text using pure ASCII, for people who don't do much Russian T_EXing, but need to set an occasional citation. In this case, control sequences can be used, and it is necessary to specify the boundary between Russian and non-Russian text to switch the hyphenation patterns. E.g. for printing this article in *TUGboat* the `WNcyr` realization of the cyrillic characters by Thomas Ridgeway has been used.

An integration of Russian into "International L^AT_EX" by Joachim Schrod is done by following `russian.sty` file:

```
\def\contentsname{Содержание}
\def\listfigurename{Список рисунков}
\def\listtablename{Список таблиц}
\def\abstractname{Аннотация}
\def\partname{Часть}
\def\chaptername{Глава}
\def\appendixname{Приложение}
\def\refname{Литература}
\def\bibname{Библиография}
\def\indexname{Алфавитный указатель}
\def\figurename{Рис.}
\def\tablename{Таб.}
\def\enclname{Вложение}
\def\ssname{Копия}
\def\headtoname{К:}
\def\pagename{Страница}
% переопределение команды вывода текущей даты
\def\today{\number\day\space
\ifcase\month\or
  января\or февраля\or марта\or апреля\or мая\or июня\or
  июля\or августа\or сентября\or октября\or ноября\or
  декабря\fi\space\number\year}
\language=1% переключение на русский язык
```

The macros `\Alph` and `\alph` are redefined too.

For IBM PC compatible computers the coding scheme presented on tab. 1. is used. The russian letter "yo" (absent in KOI-8 coding scheme) and double quotes (like "<< >>") are added. For easy input the quotes occupy the positions of the characters "<>", but the symbols "i" and "j" are moved to free positions.

	00	10	20	30	40	50	60	70	80	90	A0	B0	C0	D0	E0	F0
0	Г	1	-	0	@	P	'	p	А	Р	а				р	Е
1	Δ	j	!	1	A	Q	a	q	Б	С	б				с	ē
2	Θ	'	"	2	B	R	b	r	В	Т	в				т	№
3	Λ	'	#	3	C	S	c	s	Г	У	г				у	
4	Ξ	'	\$	4	D	T	d	t	Д	Ф	д				ф	
5	Π	'	%	5	E	U	e	u	Е	Х	е				х	
6	Σ	'	&	6	F	V	f	v	Ж	Ц	ж				ц	
7	Υ	'	'	7	G	W	g	w	З	Ч	з				ч	
8	Φ	'	(8	H	X	h	x	И	Ш	и				ш	
9	Ψ	ß)	9	I	Y	i	y	Й	Щ	й				щ	
A	Ω	æ	*	:	J	Z	j	z	К	Ъ	к				ъ	
B	ff	œ	+	;	K	[k	-	Л	Ы	л				ы	
C	fi	ø	,	«	L	"	l	—	М	Ь	м				ь	
D	fl	Æ	-	=	M]	m	"	Н	Э	н				э	
E	ff	Œ	.	»	N	'	n	'	О	Ю	о				ю	ı
F	ff	Ø	/	?	O	'	o	'	П	Я	п				я	ı

Table 1. Coding scheme for T_EX at IBM PC

References

- [Beeton85] Barbara BEETON, "Mathematical symbols and cyrillic fonts ready for distribution (revised)", TUGBOAT 6, (1985), no. 3, pp. 124-128.
- [Vulis89] Dimitri VULIS, "Notes on Russian T_EX", TUGBOAT 10, (1989), no. 3, pp. 332-336.
- [Glonti90] Nana L. GLONTI et al, "Metaproject of the cyrillic alphabet for high quality printers", IHEP Preprint, 90-66, Protvino, 1990.