kirilmath — Cyrillic and maths

kirilmath version 1.10

Georgi Boshnakov

1 Introduction

This documentation is a draft.

A common and extremely irritating mistake when writing mathematical texts in Bulgarian (and other Cyrillic based languages) is to forget to switch the character encoding when changing from maths (which uses Latin alphabet) to text (Cyrillic alphabet) and vice versa.

In Emacs the alternative encoding is specified by a default input method which is turned on (for text) or off (for mathematics or a Latin based text).

We provide commands that almost eliminate the need to manually turn the default encoding on and off when editing text.

This package is intended to be used mainly in TEX modes in conjunction with AUCTeX. Some of the commands are meaningful in ordinary text modes, too.

These features are best used in connection with AucTeX.

2 Installation

This package can be installed through the packaging system distributed with Emacs 24 and available as add-on for older versions of Emacs.

The package currently resides in my package archive at http://www.maths.manchester.ac.uk/~gb/emacs/packages/. To make the packages provided there known to the package system, add the above site to the list of package archives. This can be done, for example, by putting the following command in your .emacs file (and restarting Emacs).

```
(setq package-archives
      '(("gnu"       . "http://elpa.gnu.org/packages/")
        ("marmalade" . "http://marmalade-repo.org/packages/")
        ("georgi" . "http://www.maths.manchester.ac.uk/~gb/emacs/packages/"))
```

Once this is done, you can run the command 'list-packages' and choose package 'Kirilmath'. The Info version of this documentation should also appear at the top level of the Info help.

You may need to put the following line in your '.emacs' file.

```
(require 'kirilmath)
```

and restart Emacs.

The old method of installation will work as well - download the package from http://www.maths.manchester.ac.uk/~gb/emacs/packages/ (it is a tar file) extract 'kirilmath.el' and place it where Emacs looks for lisp files, e.g. in subdirectory 'site-lisp' of the Emacs installation directory, put the above 'require' command in your '.emacs' file and restart Emacs.

Place the file 'kirilmath.el' at a place where Emacs looks for lisp files, e.g. in subdirectory 'site-lisp' of the Emacs installation directory.

Put the following line in your '.emacs' file

```
(require 'kirilmath)
```

or

```
(load "kirilmath")
```

3 Alternating between Cyrillic and Latin input

The discussion below assumes that there is a “background” keyboard layout, such as US or British English, for entering characters for languages that use (variants of) the Latin alphabet. This layout is usually set at the operating system level and is used to communicate with it, as well. The Cyrillic layout is selected by choosing an input method with a key like *C*-`\` (`toggle-input-method`).

The discussion also assumes that the user changes the keyboard encoding within Emacs, typically using keys bound to relevant commands. OS level commands may also be used but the results may vary between OSes[1].

When a Cyrillic layout is selected, we say that the input method is “on”, otherwise it is “off”.

A very annoying mistake is to input text with the wrong input method. This is particularly common when for mathematical texts. It is often convenient and less error prone to use commands that automatically choose the correct input method depending on context.

The command `kiril-adjust-language` in Kirilmath turns on or off the default language encoding depending on the buffer contents around point. This command can be used in any mode but is most useful in TEX modes. If `kiril-adjust-language` is not able to make a guess, it leaves the encoding as it is, to avoid annoying results. The same happens if the current encoding is the correct one.

`kiril-adjust-language` examines the buffer around point to decide on the state of the input method. In TEX-like buffers it also examines whether point is within a mathematical or text region.

Bind `kiril-adjust-language` to a convenient key and press it whenever you are uncertain if the current state of the input method is correct. This is especially useful when moving around in the process of editing existing material. I bind `kiril-adjust-language` to `CapsLock`, see section ‘FIXME:’ for details and other suggested key bindings. Note that accidental or repeated invocation of `kiril-adjust-language` is harmless[2].

Various commands are available in Emacs’ TEX modes for switching between text and mathematics, especially for users of the add-on package AUCTeX. Package Kirilmath provides further enhancements, see the remaining sections of this manual. Many of them call `kiril-adjust-language` to adjust the input method. The net effect is that explicit commands for changing the input method are rarely needed.

The variable *kiril-texmathp-makes-sense* is set to `t` for TEX-like buffers to indicate that it is appropriate to check for TEX-like mathematical context.

`kiril-adjust-language` [Command]

Adjust language to the context around point. Turn off input method if it is on and point is in a math mode context. If point is in text mode and the ‘closest’ letter

[1] On Windows, changing the language at the OS level, often with key combinations like `Alt-Shift`, is fine but makes it more difficult to use some multi-key commands in Cyrillic mode. Also, if you wish to use features of this package that automatically adjust the input method depending on the surrounding context, it is better to stick to the Emacs way.

[2] The command uses heuristics meant to be helpful and at the same time avoid annoying results. Please inform the maintainer if you encounter an instance of annoying behaviour.

is cyrillic turn on the input method, if it is latin, turn it off. The details are still subject to revision and improvement but currently the action is determined by the first success in the following sequence of tests.

1. If the preceding char is cyrillic, on.
2. If the preceding char is latin, off.
3. If the following char is cyrillic, on.
4. If the following char is latin, off.
5. If texmathp makes sense (i.e. in a TeX major mode)
   a. if in math mode, off.
   b. Otherwise, if point is preceded by punctuation and a cyrillic char, on.
   c. If point is preceded by punctuation and a latin char, off.
6. otherwise do similar checks to the above. FIXME: complete this doc string properly.

The check for text/math mode is with `texmathp` and does not pay attention to commented and verbatim material in the TEX source. It would be better to use `TeX-math-input-method-off` here (for the "then" alternative) but it currently works for CJK languages only. (FIXME: Is this still the case?)

(not implemented (todo:(?) assign math-mode map ??).

4 Inserting math and text

The commands in this section are for TeX modes and complement some features provided by AUCTeX.

The command `kiril-TeX-insert-dollar` inserts the dollar symbol *$*, sets up a math or text context depending on whether the dollar sign starts or ends math mode, and adjusts the encoding by calling `kiril-adjust-language`. This is similar to AUCTeX's `TeX-insert-dollar` but the latter does not check the surrounding context. The command `kiril-TeX-insert-dollar` is usually bound to the key *$*.

Another approach to entering mathematics is to insert the opening and closing dollar signs at once and position the cursor between them. The command `kiril-LaTeX-electric-embed` can be used for this purpose. If point is in text context, `kiril-LaTeX-electric-embed` inserts '`$ $`' and puts point between the dollar signs.

If point is in a math context, `kiril-LaTeX-electric-embed` inserts a '`\text{}`' macro (an AMSLaTeX command) and leaves point between the braces.

In both cases `kiril-LaTeX-electric-embed` adjusts the language.

Todo: make the inserted string customizable.

This approach works well in combination with the commands `kiril-jump-before-math` and `kiril-jump-after-math`. See FIXME:

These commands are similar in spirit to `kiril-adjust-language` in that they decide what environment to embed on the basis of surrounding context. I bind `kiril-LaTeX-electric-embed` to *`C-CapsLock`*.

Set variable `kiril-math-auto-on-off` to `nil` to stop the above and similar functions from changing the encoding by looking at the surrounding context. FIXME: not used currently! If you do this because of annoying behaviour of the functions, please let me know about the problem.

5 Moving out of math

These are convenience functions for moving around when editing TeX buffers.

The command `kiril-jump-before-math` moves point to a position just before the current math environment. Similarly, `kiril-jump-after-math` moves point just after the current math environment.

“Environment” here means any (known) TeX or LaTeX math construct, such as the macro \ensuremath and inline math delimited by *$*. If point in not in math context, nothing happens and a message is shown in the echo area.

`kiril-jump-before-math` () [Command]

> Jump to a position just before the current math environment. `"Environment"` here means any (known) TeX or LaTeX math construct, including macros like `\ensuremath`.

`kiril-jump-after-math` () [Command]

> Jump after the current math environment. `"Environment"` here means any (known) TeX or LaTeX math construct, including macros like `\ensuremath`.

`kiril-find-LaTeX-matching-end` () [Command]

> Go to the end of the current math environment. `"Environment"` here means any (known) TeX or LaTeX math construct, including macros like `\ensuremath`.

`kiril-TeX-find-closing-brace` [Command]

> Find the position of the closing brace of the current TeX group. Like `TeX-find-closing-brace` but interactive.

6 Key bindings

Here are some commands that you may wish to set in your ‘`.emacs`’ file to activate some of the features. You can use any keys you like, the ones given here are suggestions or examples.

It is worth binding the command `kiril-adjust-language` to some always available convenient key. Hitting that key will ensure the correct choice between the Cyrillic and background encoding. For example, this binds globally `kiril-adjust-language` to *`capslock`*.[1]

```
(global-set-key [capslock] 'kiril-adjust-language)
```

I like *`capslock`* for this command since it is conveniently located and I never use it for its original purpose. Also, accidentally hitting it is harmless with this binding.

Use the following lines in your .emacs file to bind the command `kiril-LaTeX-electric-embed` could similarly be bound globally to a key but it is better to bind it only in TEX modes. Here is how to bind it to `C-capslock` for AUCTeX's TEX modes.

```
(add-hook 'TeX-mode-hook '(lambda ()
   (define-key TeX-mode-map [\C-capslock] 'kiril-LaTeX-electric-embed)))
```

The following binds `kiril-TeX-insert-dollar` to *$*.

```
;; To switch  the  input encoding as necessary when entering a $ in TeX modes.
(add-hook 'TeX-mode-hook '(lambda ()
          (define-key TeX-mode-map (kbd "$") 'kiril-TeX-insert-dollar)))
```

Users of AUCTeX may consider making the backslash active with the setting

```
(setq TeX-electric-escape t)
```

Then, pressing the backslash switches off the input method (among other things), which is what we almost always want since TEX commands usually use only ASCII letters and symbols.

Unicode has slots for the lowercase and uppercase Cyrillic letter “i with grave accent”. Some keyboard layouts for Bulgarian language allocate keys to these letters. If you wish these keys to insert the “classic” TEX syntax for these letters, set variable `kiril-cyrIi-grave-TeX` to ‘'’ as in

```
(setq kiril-cyrIi-grave-TeX "'")  ; grave-cyri using \`, in TeX modes.
```

[1] This assumes that you have made `CAPSLOCK` an ordinary key. On Windows this can be accomplished by putting the command `(setq w32-enable-caps-lock nil)` in your .emacs file.

7 Opening and closing quotes

Computer typesetting makes it possible to use proper opening and closing quotes. There are slots for them in the encodings provided here. The neutral double quotes are also available at their usual keyboard place.

When one typesetting TEX documents one may nevertheless prefer to use the syntax for the TEX '`babel`' package.

`kiril-toggle-TeX-quotes ()` [Command]

In TEX modes, toggle the cyrillic quotes between Unicode and LaTeX (babel) format. The default is to use Unicode quotes. For LaTeX, quotation marks may be entered either via their Unicode slots (there are keys for the opening and closing quotation marks in the modern Bulgarian layout) or via ligatures using `"`` and `"'`.

This command changes the bindings for the quotes only for the TEX modes, it does not change the global map. If it detects that Auctex is loaded it changes its TEX map, otherwise it changes the TEX map for the native TEX modes.

This command changes the keys for the Cyrillic opening and closing quotes only. In particular, any bindings to the non-directional quotes, `"`, are not affected.

`kiril-open-quotes-TeX` [Command]

Print cyrillic opening quotes in LaTeX (babel) format

`kiril-close-quotes-TeX` [Command]

Print cyrillic closing quotes in LaTeX (babel) format

8 Cyrillic-i-grave (TeX variants)

This section may be of interest only for users writing in Bulgarian. The material should become obsolete as the Unicode support by TeX engines becomes complete.

i-grave is used relatively rarely in Bulgarian and it was not available on typewriters in the past. Its main purpose is to facilitate the reader in relatively rare cases when a word might be interpreted wrongly initially. Since the main purpose of good typography is precisely to help the reader, it is a good style to put the accent when needed.

For LaTex users putting an accent over most letters is not a problem but for others it may be. Sure, i-grave has a slot in Unicode but it is relatively recent and the glyph is not always available.

Alternatively, the letter can be obtained by combining an ordinary cyrillic-i with a "combining" grave accent \(this is also a feature of Unicode). At the time of writing, the support for this is also patchy. Emacs 23 deals with this correctly (if you do not mind the somewhat bizarre look of the combined character). Emacs 22 displays the combination as an cyrillic-i followed by a grave accent (this is what MS word does on my system, as well). LaTeX renders the combination correctly but requires that it is enclosed in a command.

For the TeX modes a third alternative is offered - to output the accent using `` \` `` directive. This makes the file more portable, too.

`kiril-cyrIi-grave-TeX` [Variable]

Specify which variant of cyrillic-i-grave to use in TeX modes. If *kiril-cyrIi-grave-TeX* is set to `'` use “classic” TeX. If *kiril-cyrIi-grave-TeX* is set to `combine` use the Unicode TeX combining accent. If *kiril-cyrIi-grave-TeX* is set to any other value, retain the setting of the global map.

The default (i.e., if *kiril-cyrIi-grave-TeX* is not set) is “classic” TeX.

Variants of cyri-grave for TeX modes

cyri with Unicode combination accent: needs to be enclosed in a special way for smooth LaTeX-ing. Todo: this variant is LaTeX specific! Maybe allow the user to choose the enclosure.

`kiril-cyri-grave-TeX-combine` [Command]

Insert cyri-grave as Unicode combined accent using LaTeX syntax.

`kiril-cyrI-grave-TeX-combine` [Command]

Insert cyrI-grave as Unicode combined accent using LaTeX syntax.

cyri for "classic" TeX.

`kiril-cyri-grave-TeX` [Command]

Insert cyri-grave using classic TeX syntax.

`kiril-cyrI-grave-TeX` [Command]

Insert cyrI-grave using classic TeX syntax.

`kiril-cyrIi-grave-TeX` [Command]

Set the TeX syntax of lowercase and uppercase cyrillic i-grave. This is controlled by the variable *kiril-cyrIi-grave-TeX*. `'` stands for "classic" TeX, *`combine`* stands for Unicode TeX combining accent, any other value retains the setting of the global map. The default (i.e., if *kiril-cyrIi-grave-TeX* is not set) is classic TeX.

9 Miscelaneous variables and functions

`kiril-LaTeX-electric-embed` [Command]

If in math mode insert a `\text` macro. If in text mode insert $ $ and set point between the dollars. Adjust language in both cases. To do: make the inserted string customizable.

`kiril-TeX-insert-dollar` [Command]

Insert dollar sign. Calls 'TeX-insert-dollar' and then adjust the encoding according to the math/text mode.

`kiril-texmathp-makes-sense` [Variable]

A non-nil in a buffer specifies that 'texmathp' makes sense for that buffer. This variable is usually set in TeX modes and allows lisp functions to use 'texmathp' to determine if point is in text or math context. This variable is buffer local.

`kiril-math-auto-on-off` [Variable]

It `t` allow automatic adjustment of the math according to context. This variable is buffer local.

FIXME: not used currently!

`kiril-insert-tilde` [Command]

Inserts the tilde, '`~`' character. This character is used for non-breakable space in TeX but standard Bulgarian keyboard layouts do not have this character. This command can be attached to a key to make the tilde available.

`kiril-insert-sharp` [Command]

Inserts the '`#`' character.

Index

(Index is nonexistent)

Table of Contents