

How to create a new IME on Linux in about 15 minutes with SCIM and m17n

Author: JanuszPrusaczyk<JanuszPrusaczyk@unknown.com>

With SCIM, you can easily create your own input methods, and you do not necessarily have to be a programmer to do this, though you can achieve more spectacular results if you are one. In this article I am proposing to extend m17n database.

SCIM already supports many languages. It is easy to use, you just click on it's icon on the taskbar, select the one you want to use and start writing. Your input appears in your web browser's input fields, mail program or general purpose editor. But what happens if SCIM does not have an input method that you need? The reasons for this can be different - maybe your language is just not supported (ex. you are using some exotic African language), maybe you are a scholar and want to use some strange transcription, or maybe you are just a creative person and want to have something tailored for your needs (ex. you need a special input method for the language of your own).

If you are using scim-m17n, your extension will be read in when SCIM restarts and you will be able to use your own input method in all the programs supported by SCIM.

This is not the only way to achieve this result, I will describe alternative solutions in future articles.

Specification of the problem

I am creating a webpage about languages. I would like to input transcription of foreign words in International Phonetic Alphabet. The problem is that I do not have IME for this. Quick search shows, that such IME exists in UIM, but I do not use it with SCIM. Searching the web I found information about SAMPA and X-SAMPA. Looking further I came to this webpage: <http://www.theiling.de/ipa/cxs.html> It describes a variant of X-Sampa used on Conlang Mailing List. I decided this is what I needed. I would like to write in Sampa, and get IPA in the document.

The approach to the solution

As we look at the problem, we see that we want some characters be substituted by other

characters. For example, "&" should input "Ã~Â¿Â½". But this is not all. Sometimes other characters following the given one can modify it. For example, "@" should be substituted by "Ã%â,ç", but if we write "@", we want to get "Ã%Ëœ", and "@`" should be substituted by "Ã%Å¡". Such modifying strings can be longer.

The similar problem is solved in Vietnamese VIQR and latin-post input methods, available in scim-m17n. Scim-m17n is using m17n library as the backend. If we make m17n support our new input method, SCIM will use it as well.

Technical documentation

If you want to understand what you will be doing, you should read the documentation of m17n. This is not necessary in the very beginning, and I suggest that you first create your first simple input method, then read more.

This is the main page of the m17n-lib and m17n-db: <http://www.m17n.org/m17n-lib/>

This page describes the format of input method data: <http://www.m17n.org/m17n-lib/m17n-docs/m17nDBFormat.html#mdbIM>

This page provides short explanations for each input method: <http://www.m17n.org/m17n-lib/m17n-docs/m17nDBData.html#mim-list>

OK, let us do it!

When we create a new XXX.mim file (our input method), we put it under the directory where the m17n-db is installed (usually, /usr/share/m17n or /usr/local/share/m17n). We have to edit the file mdb.dir in the same directory and add this line:

```
(input-method LL NAME "XXX.mim")
```

LL is a two letter language code of ISO639-1 (e.g. "vi" is for Vietnamese), NAME identifies the input method.

Let us open mdb.dir file and register new table (the last entry is the new one): . . .

```
(input-method vi telenx "vi-telex.mim")
(input-method vi viqr "vi-viqr.mim")
(input-method zh py "zh-py.mim")
(input-method t latin-post "latin-post.mim")
(input-method t unicode "unicode.mim")
(input-method t x-sampa "x-sampa.mim")
...
```

As you can see, I decided to name my input table x-sampa. I followed the example of latin-post, as it's purpose is very similar to what I try to create. We have to create the *.mim file: touch /usr/share/m17n/x-sampa.mim

```
A quick look inside latin-post.mim shows what do we have to enter in x-sampa.mim ;;; <li>
x-sampa.mim
;;;
;;; Modified X-Sampa Used on the Conlang Mailing List.

(title "x-sampa")

(map
  (trans

    ("_0" 0x030A) ("_0" 0x0325)
    ("&" 0x00E6)
    ("|" 0x01C0)
    ("||" 0x01C1)

    ...

    ("z" 0x0291)
    ("Z" 0x0292)
    ("Z_j" 0x0293)

  ))

(state
  (init
    (trans)))
```

That is all. We just have to check whether our IME is working. The easiest way is to start medit, select our new input method and make sure it works as expected.

Final step

The only thing we still have to do is to restart SCIM (I just reboot my computer, but restarting X Windows should be enough). That's it!

Thanks

I would like to thank Mr Kenichi Handa from m17n project for his kind help during my own struggle with this simple problem.