

Updated SCIM Input Method

Basanta Krishna Shrestha
basanta@mpp.org.np
Madan Puraskar Pustakalaya, Nepal

Abstract

The document is an updated report on the SCIM Input Method. The updated version of the input method has been integrated in the recently released latest version of NepaLinux, i.e. NepaLinux 1.1. This updated version addresses the zero-width joiner and zero-width non joiner issues which had not been addressed in the earlier version of the SCIM Input Method included in NepaLinux 1.0. The document provides a brief introduction of the SCIM Input Method, steps for the installation of the SCIM Input Method and finally also provides guidelines for using the input method.

1. Introduction

Smart Common Input Method (SCIM), is an input method (IM) for POSIX-style operating systems such as Linux, Unix, BSD [5]. Input methods are needed to enter data. Apart from providing a user-friendly input method user interface, it also can be used for producing new input methods [5].

SCIM is highly modular. It consists of 4 basic modules, each provided by different packages. Listed below are the package names and their brief module descriptions.

scim : scim is the core package, which provides the fundamental routines and data types. This package contains the main binary "scim" and other support programs. it provides a common platform for various modules to be plugged in. It also includes a set of programs and modules of its own.

scim-gtk2-immodule: This employs GTK+2 input method module with scim serving as backend [10]. This means, if GTK_IM_MODULE is set to use scim, this package is responsible for making gtk+ application(eg. gedit, firefox) to use scim as input

module by default. This input method should be compatible with all GTK+ 2.x platforms, and this includes gtk-x11 [5].

scim-modules-socket: This module is responsible for ending communication between the GTK IM module and SCIM back. The communication is mainly accomplished by the socket-IM-engine and socket-front- end modules available in the package [5].

scim-tables-additional: This module provides input method data tables. Character mapping is done in data tables. This package contains data tables for several languages.

2. Steps for SCIM installation in Debian GNU/Linux

2.1. Install SCIM core packages

```
#apt-get update  
#apt-get install scim  
#apt-get install scim-gtk2-immodule
```

(Note: This will also install package scim-modules-socket)

```
#apt-get install scim-modules-tables
```

(Note: The above package provides "scim-make-table" binary which we will use in the example below to create a new input method data table)

2.2. Creating a new Input Module Data Table

Create a file named myinput.txt. The file is divided into two sections, namely the TABLE DEFINITION and the TABLE DATA with other information as mentioned below [2, 3,4, 6, 7,8]:

- TABLE DEFINITION: Entries like table

name, language code, locale & author name are defined.

- TABLE DATA : Character mapping is done in this section

```
-----  
### File header must not be modified  
### This file must be encoded into UTF-8.  
### This files tries to implement the Traditional  
### keyboard layout modified by MPP for PAN  
Project  
SCIM_Generic_Table_Phase_Library_TEXT  
VERSION_1_0
```

```
### Begin Table definition.  
BEGIN_DEFINITION
```

```
### An unique id to distinguish this table among  
others.  
### Use uuidgen to generate this kind of id.  
UUID = 16f49d28-677b-4ac7-a93c-9f714b070a5a
```

```
### A unique number indicates the version of this file.  
### For example the last modified date of this file.  
### This number must be less than 2^32.  
SERIAL_NUMBER = 20051103
```

```
ICON = /usr/share/scim/icons/Nepali.png
```

```
### The default name of this table  
NAME = Traditional
```

```
### The local names of this table  
NAME.ne_NP = ट्रेडिस्नल
```

```
### Supported languages of this table  
LANGUAGES = ne_NP
```

```
### The author of this table  
AUTHOR = Harkhe <harkhe@gmail.com>
```

```
### Prompt string to be displayed in the status area.  
STATUS_PROMPT = NP
```

```
### If true then the first candidate phrase  
### will be selected automatically during inputing.  
AUTO_SELECT = TRUE
```

```
### If true then a multi wildcard will be appended  
### at the end of inputing string automatically.  
AUTO_WILDCARD = FALSE
```

```
### If true then the result string will be committed to  
client automatically.
```

```
### This should be used with AUTO_SELECT =  
TRUE.  
AUTO_COMMIT = TRUE
```

```
### If true then the inputed string will be automatically  
splitted during inputing.  
AUTO_SPLIT = FALSE
```

```
### If true then the phrases' frequencies will be  
adjusted dynamically.  
DYNAMIC_ADJUST = FALSE
```

```
### If true then the preedit area will be filled up by the  
current candidate phrase automatically.  
AUTO_FILL = FALSE
```

```
### If true then the lookup table will always be shown  
if there is any candidate phrase.  
### Otherwise the lookup table won't be shown unless  
the user requires it by moving the pre-edit caret left.  
ALWAYS_SHOW_LOOKUP = FALSE
```

```
### Enable full width punctuation property  
USE_FULL_WIDTH_PUNCT = FALSE
```

```
### Use full width punctuation by default  
DEF_FULL_WIDTH_PUNCT = FALSE
```

```
### Enable full width letter property  
USE_FULL_WIDTH_LETTER = FALSE
```

```
### Use full width letter by default  
DEF_FULL_WIDTH_LETTER = FALSE
```

```
### The maximum length of a key.  
MAX_KEY_LENGTH = 1
```

```
### Valid input chars.  
VALID_INPUT_CHARS =  
abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMN  
OPQRSTUVWXYZ1234567890!@#%&*&*()_-  
+=\~`.;'";{}|<,>.:?/
```

```
### Single wildcard char, can have multiple chars.  
SINGLE_WILDCARD_CHAR = ?
```

```
### Multi wildcard char.  
MULTI_WILDCARD_CHAR = *
```

```
### The key strokes to split input string.  
SPLIT_KEYS = quoteright
```

The key strokes to commit the convert result to client.

COMMIT_KEYS = space

The key strokes to forward the inputted string to client.

FORWARD_KEYS = Return

The key strokes to select candidate phrases.

SELECT_KEYS = 1,2,3,4,5,6,7,8,9

The key strokes to page up the lookup table.

PAGE_UP_KEYS = Page_Up

The key strokes to page down the lookup table.

PAGE_DOWN_KEYS = Page_Down

END_DEFINITION

Begin Table data.

BEGIN_TABLE

"	्
#	घ
%	छ
^	ट
&	ठ
*	ड
'	डु
(ढ
)	ण
+	ं
=	0x200c
,	s
-	औ
.	।
/	र
0	०
1	१
2	२
3	३
4	४

5	५
6	६
7	७
8	८
9	९
;	स
<	ड
?	रु
@	ई
A	आ
B	ौ
C	ऋ
E	ऐ
F	ँ
H	झ
J	ो
K	फ
L	ी
O	इ
P	ए
U	ऊ
V	उँ
\	्
]	े
_	ओ
`	ज
a	ब
b	द
c	अ
d	म
e	भ
f	ा
g	न
h	ज

```

i      ष
j      व
k      प
l      ि
m      ः
n      ल
o      य
p      उ
r      च
s      क
t      त
u      ग
v      ख
w      ध
x      ह
y      थ
z      श
{      ृ
|      0x200d
}      ै
~      ॥
### PHRASES
!      ज
$      ङ
:      ट्ठ
>      श्र
D      इग
G      द्
I      क्ष
M      इड
N      य
Q      त्त
R      द्व
S      इक
T      ट्ट

```

```

W      इढ
X      ह्य
Y      ठ्ठ
Z      क्क
[      र्
q      त्र
END_TABLE

```

2.3. Converting the table into binary format and installing

```

# scim-make-table myinput.txt -b -o myinput.bin
#mkdir -p /usr/share/scim/tables
#cp myinput.bin /usr/share/scim/tables/

```

Please refer to following page for more information on creating new table data.

http://www.scim-im.org/development/contribute/how_to_create_a_new_ime_in_about_15_minutes_with_scim_and_scim_tables

2.4. Installing already available Input Method data tables:

If the input method data table for your language has already been uploaded into scim upstream and is available in debian, you can simply install the package as:

```
#apt-get install scim-tables-additional
```

This package contains IM data tables for non CJK languages. Currently it supports Arabic, Nepali, Russian, Thai, Vietnamese, Bengali, Gujrati, Hindi etc.

UsingSCIM

Create a file /etc/X11/Xsession.d/95scim_start and export following environment.

```

#touch /etc/X11/Xsession.d/95scim_start
#vi /etc/X11/Xsession.d/95scim_start

```

```

-----
export XMODIFIERS=@im=SCIM
export GTK_IM_MODULE=xim
/usr/bin/scim -d

```

Execute any program and hit cntrl+space to activate scim input.

3. Conclusion

The updated SCIM Input Method provides efficient input facilities for the Nepali language in the Debian Linux. This report documents the whole process of customizing and using the input software and is believed to be useful for anybody interested to develop a SCIM Input Method for their respective languages.

4. References:

- [1] "SCIM" <http://scim-im.org>
- [2] <http://www.xychen.org/zhdesktop/scim/CangJie3.txt>
- [3] <http://www.xychen.org/zhdesktop/scim/CangJie5JT.txt>
- [4] <http://www.cygron.com/Downloads/FinAnal.dpr>
- [5] <http://bozu.sytes.net>
- [6] <http://linux.com.np/modules.php?name=Forums&file=viewtopic&p=73>
- [7] <http://www.xychen.org/zhdesktop/scim/CantonHK.txt>
- [8] <http://www.northwalesleisure.com/download/office/PFILES/MSOFFICE/OFFICE/1033/XLMAIN9.AW>
- [9] <http://microsoftxiao.cnblogs.com/>
- [10] <http://klik.atekon.de/?section=utils/>