

Indic to English Transliterator User Manual

Pravin Paratey

July 23, 2008

1 About

This application lets you convert a word from an Indian language¹ to English while retaining the way it sounds. This process is known as **Transliteration**. A web interface is provided with language keyboards for easy operation.

2 Installation and Running

2.1 On Windows

Usage on windows is pretty straight forward. All you have to do is double click the setup file and follow the installation process. If you are viewing this document from the CFILT cd, the setup file will be found in `system/exe` folder.

The transliterator can be uninstalled from the Add/Remove programs area in the control panel. All files created by this application will be removed.

To run the transliterator, double click the icon on the desktop. An icon will appear in the system tray at the lower right corner of your screen. Right click the icon and click start. Transliterator should run in your default web browser.

Note: Transliterator can be run as a server so that people from other computers can connect to this computer. On running Transliterator for the

¹Currently Devanagari and Bengali only

first time, Windows Firewall might prompt you block this application or to allow access.

- If you block this application, only the user on that computer will be able to use the Transliterator.
- If you allow access, then other computers can access Transliterator by entering `http://< ip - address >:8000/` on their computer. This is useful if you want many people to train the same data set.

To exit, right click the icon again and select Exit.

2.2 On Linux and Macs

First, make sure you have python installed. This can be checked by typing

```
$ which python
```

at the command prompt. If a location is returned, it means python is installed on your machine.

Next, navigate to the source directory. If you are viewing this document from the CFILT cd, the source directory will be found at `system/sources`. Then type the following at the command prompt:

```
$ python WebServer.py
```

Navigate your browser to `http://localhost:8000` to view the Transliterator interface.

3 Usage

In this section, we're going to take a look at the running of the Transliterator application and the possible cases one might encounter while running the application.

3.1 User Interface

When the Transliterater is started, an interface like in Figure(1) will be seen. The text box is where one has to enter a word in an Indian language (as of writing this report, only Devangari and Bengali) which will be transliterated into English on hitting the Go button.

A keyboard selector is also present. Clicking on *Switch to Bengali* switches the keyboard to show Bengali characters. Clicking on *Switch to Devanagari* will restore the Devanagari keyboard. If a unicode keyboard is present, the characters can be typed directly into the text box without clicking on the on-screen keyboard.

Finally, after one hits on the Go button, the screen contains a textbox at the bottom where one can enter the correct transliteration if the system was unable to generate the correct transliteration. This will be illustrated shortly.

3.2 Illustrating transliteration when target word exists in the dictionary

The dictionary is used to remove non-words out of the output. This is most useful in a project like the CLIA where one has to return indexed documents. In that case, we'll use the term list generated by the indexer as our dictionary.

Let's illustrate transliteration for the word प्रविन. For this example, the words pravin and parvin were present in the dictionary but praveen wasn't. Figure (2) shows the output. As we can see, pravin has a higher score than parvin and that **praveen** isn't shown even though it is a valid transliteration.

To add *praveen*, enter **praveen** in the Enter correct transliteration box and click Submit. The next time you search for प्रविन, you will see pravin, parvin and praveen as the possible transliterations.

3.3 Illustrating transliteration when target word isn't found in the dictionary

There are two cases to consider here,

1. Target word is in the possible transliterations.
2. Target word is **not among** the possible transliterations.

3.3.1 Target word is among the list of possible transliterations

To illustrate this, let's take the name of one of Mumbai's train stations - चिन्चपोकली. Entering it in the Transliteration box gives us the result as in Figure(3). Since the word isn't found in the dictionary, a list of possible transliterations are output and the user can choose one of them to be the correct transliteration. As the topmost result is the correct one, we select it and click **Submit** to add this word to the dictionary.

3.3.2 Target word is not among the list of possible transliterations

To illustrate this, we pick the word तबलामास्टर. Entering this word in the Transliteration box and hitting Go gives us the results as shown in Figure(4). Since none of the possibilities suggested by the Transliterator is correct, we enter the correct transliteration in the correction box provided.

Note that one can enter the word as `tablamaster` or `tab la mas ter` which is the syllabified version. Though the system adjusts its weights for both inputs, entering a syllabified input helps increase the accuracy of the system for inputs containing similar syllables.

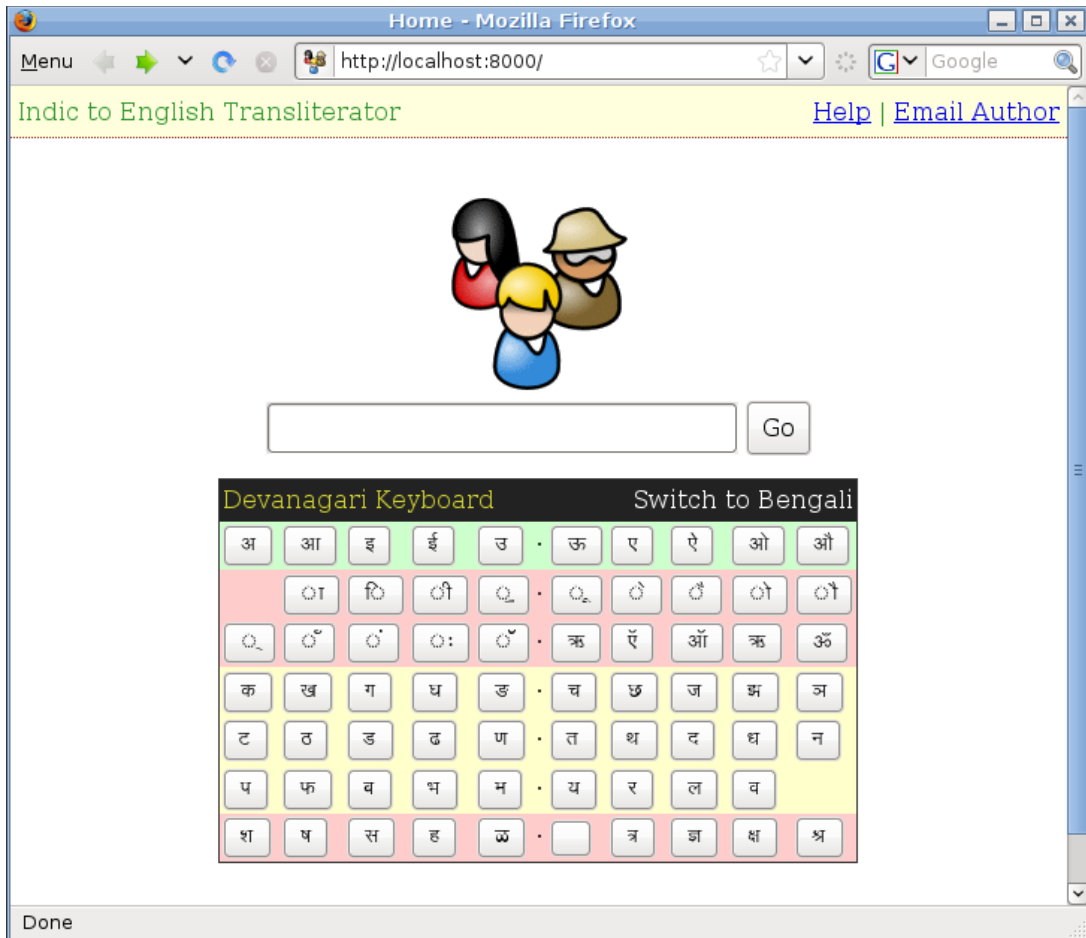


Figure 1: Transliterator interface

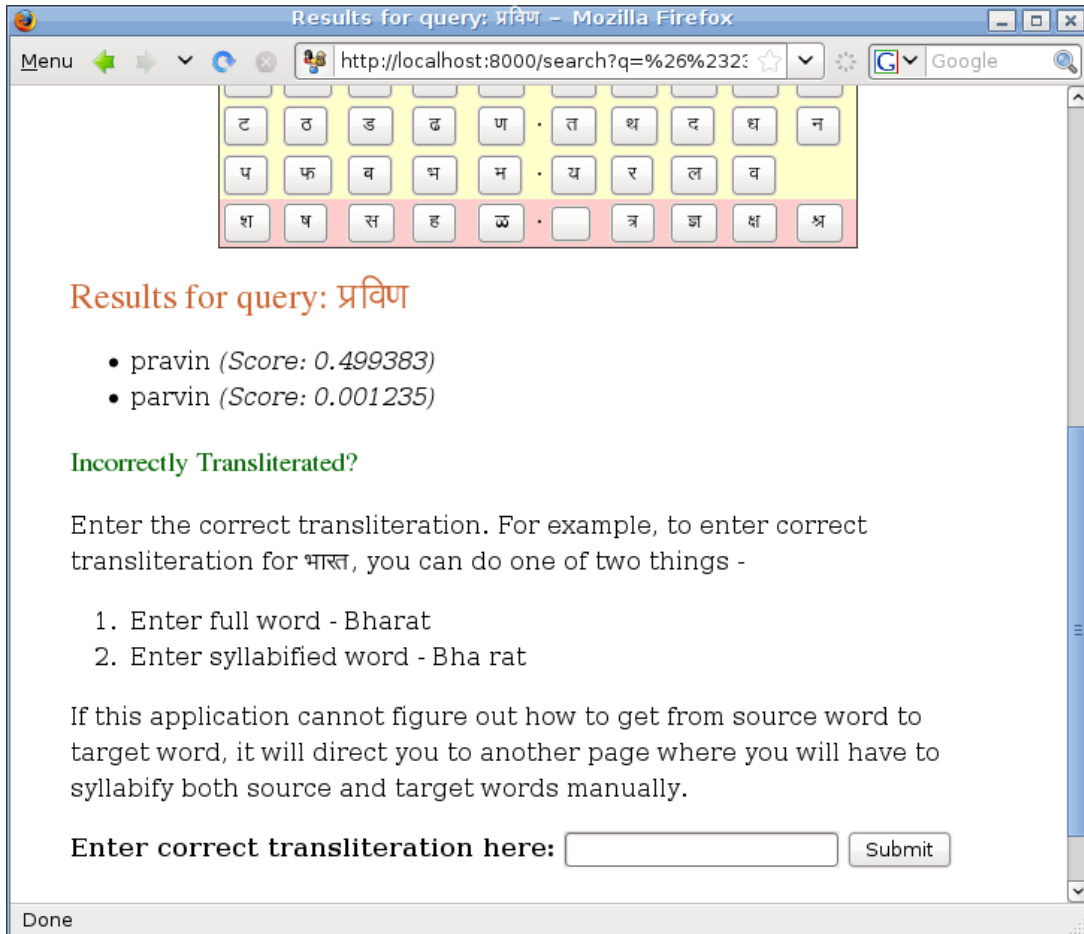


Figure 2: Transliteration of Pravin

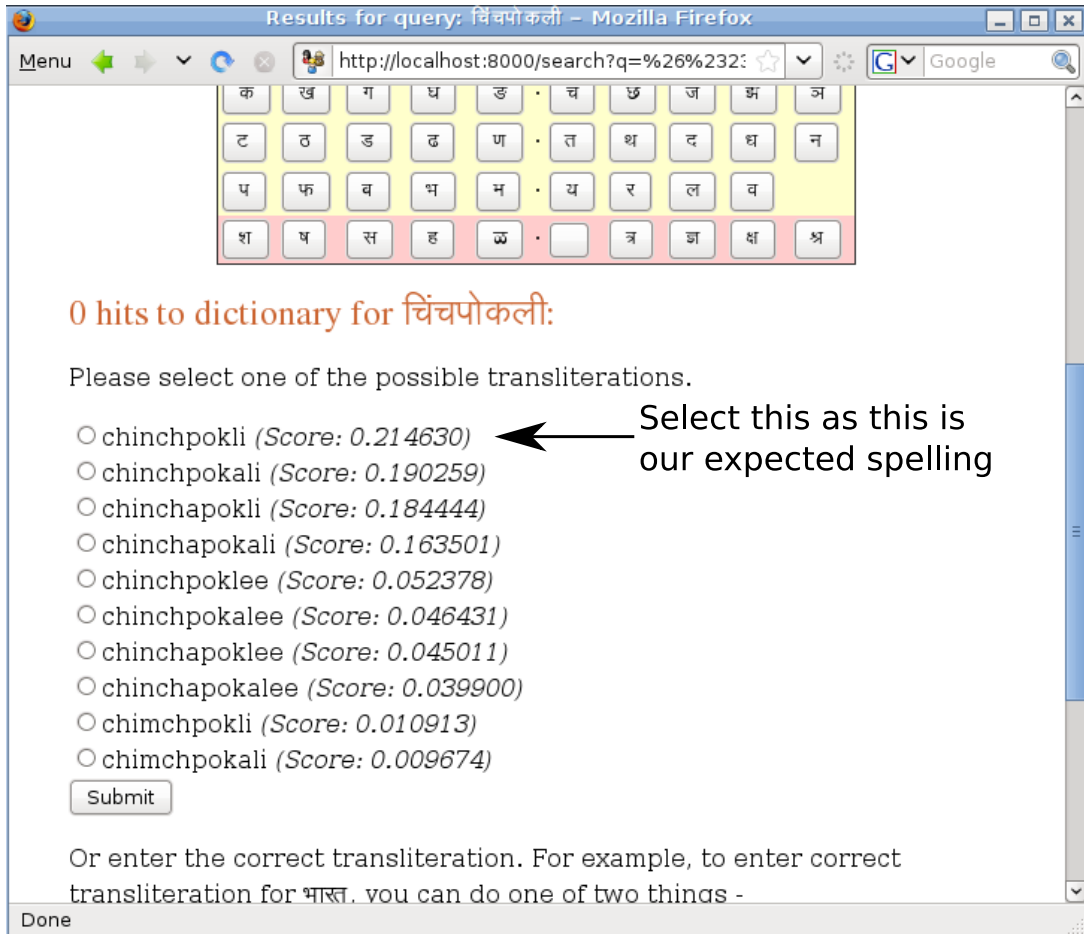


Figure 3: When word isn't found in the dictionary

