KEYBOARDS AND INTERNATIONALIZATION

A Lingoport White Paper
Authored by Olivier Libouban
Globalization Lead at Lingoport
www.lingoport.com

TABLE OF CONTENTS

INTRODUCTION

Congratulations, your software is now shipping throughout the world. Now, what happens to folks who need to enter data? What happens if your clients want to enter Japanese ideograms? Could they do that with a US keyboard? Conversely, what if Japanese users need to type in US text? Is your application able to handle switching from one locale to another and still let the user enter data?

KEYBOARDS BASICS

Keyboards are mostly hardware devices which send codes to an operating system. Those codes correspond to letters.

> "COMPUTER KEYBOARDS INCLUDE CONTROL CIRCUITRY TO CONVERT KEY PRESSES INTO KEY CODES THAT THE COMPUTER'S ELECTRONICS CAN UNDERSTAND."
>
> "THE CHARACTER CODE PRODUCED BY ANY KEY PRESS IS DETERMINED BY THE KEYBOARD DRIVER SOFTWARE. A KEY PRESS GENERATES A _scancode_ WHICH IS INTERPRETED AS AN ALPHANUMERIC CHARACTER OR CONTROL FUNCTION."
>
> The input character code can they be mapped to a character for a given OS. The fact that I click on the 'a' letter on my keyboard does not always mean the letter 'a' will be input in my software. It depends mostly on operating system settings.
>
> Source: http://en.wikipedia.org/wiki/Keyboard_(computing)

KEYBOARD LAYOUTS EXAMPLES

Here are some keyboard layouts for different languages:

FIGURE 1: HEBREW KEYBOARD (GOOGLE TRANSLATE)
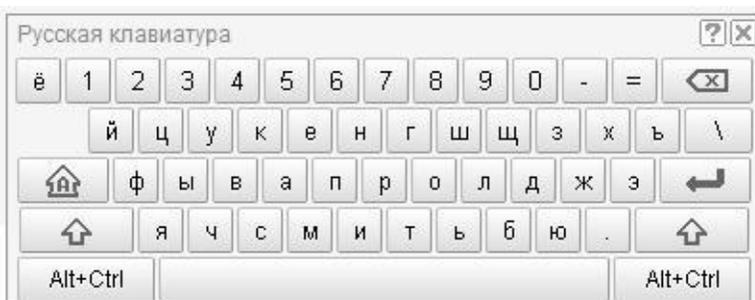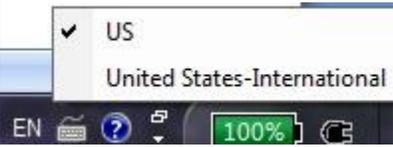

FIGURE 2: RUSSIAN KEYBOARD (GOOGLE TRANSLATE)

Japanese needs to handle hiragana, katakana, and ideographs. A software keyboard (for a touch screen, used more and more for different devices, from hand-held phones to embedded devices) could follow this pattern, where the layout and the keys switch from one to the other and to US English keyboard using control keys:

MAPPING BETWEEN KEYBOARD AND ACTUAL INPUT

Keeping the **same hardware keyboard**, a regular US keyboard, let's change the Microsoft setting for this combination of keys: **"q`az^o".**

Microsoft Setting	Result in MS Word
	q`az^o
	qàzô
	a²qw6o
	たろちつぉら

IME: INPUT METHOD EDITOR

BACKGROUND

Some writing systems have thousands of possible ideographs, like Chinese and Japanese. Some writing systems are based on ideographic stems, like Korean. Let's take **Chinese** to illustrate the major points. Because the set of Chinese characters is so huge, it is highly impractical (and for any practical keyboard, impossible) to try to map each character to a single key. Therefore, inputting Chinese characters resort to schemes involving sequences of key presses to select specific Chinese characters or sequences of characters from the available repertoire supported.

IME

An Input Method Editor is shortened to IME or sometimes called Input Method. It is a scheme to input Chinese or Japanese characters using a reduced set of keys. Depending on what particular method is used to input and select particular characters, IME's have specific names. They may also differ in strategy between inputting Chinese characters for the Chinese language and Chinese characters for the Japanese language (kanji), based on different linguistic expectations of the users and differences in the particular repertoire of characters that needs to be supported.

The ideogram "人員" is obtained by typing the following sequence typed on a US keyboard "**o<Space><Shift>8<space>",** with the following IME setting: "**CH, Chinese Traditional ChangJie**".

Common input methods for Chinese make use of romanization. Others are based on character component and stroke-based methods. There is also direct input of hexadecimal character values.

CHINESE IME TYPES

ROMANIZATION

漢語拼音, Hànyǔ Pīnyīn, or just "**pinyin**" for short is the most common Chinese IME. Pinyin represents each syllable of Beijing Chinese (People's Republic of China Modern Standard) by means of a combination of Latin characters, optionally modified by tone marks. The tone marks consist either of numbers at the end of the syllable or diacritics placed on the main vowel.

A given syllable as **Romanized** in pinyin may correspond to one or — more often — to many particular Chinese characters. The user types in the pinyin syllable as a sequence of Latin characters (and the tone indicators). When the syllable is to be converted to the correct Chinese character for input, the input method presents the user with a choice of ideograms having that pronunciation.

漢語 pinyin

1拼音 2品 3拼 4頻 5聘 6貧 7拚 8姘 9颦 ◀ ▶

The user can select one of them by use of the mouse or by shifting a selection number.

COMPONENT- AND STROKE-BASED IME

Learning to write Chinese by hand involves learning basic strokes. This type of IME is based on the graphological aspect of the characters. Each graphical unit is represented by a basic character component. There are 24 components in all, each mapped to a particular letter key on a standard QWERTY keyboard. The user types some base strokes represented on the keyboard and the IME proposes a set of ideograms which could be derived from the sequence. The user then selects an ideogram either by clicking with a mouse or by a number associated with it.

CHINESE LAYOUT AND IME

OVERVIEW

The codes of three input methods are typically printed on the Chinese (traditional) keyboard:

- Zhuyin, UPPER RIGHT;
- Cangjie, LOWER LEFT; and
- Dayi, LOWER RIGHT.

BOPOMOFO/ZHUYIN FUHAO:

(pinyin: Zhùyīn fúhào; Zhuyin Fuhao: ㄓㄨˋ ㄧㄣ ㄈㄨˊ ㄏㄠˋ; literally "phonetic symbol") or zhuyin

fuxao, often abbreviated as zhuyin and colloquially called bopomofo,[1] was introduced in the 1910s as

the first official phonetic system for transcribing Chinese, especially Mandarin.

Consisting of 37 characters and four tone marks, it transcribes all possible sounds in Mandarin. Despite

being phased out in China in the 1950s, this system is still widely used as an educational tool and in

Taiwan.

CHANGJIE OR CANGJIE:

The ChangJie IME is based on the basic strokes of the ideograms which give it its shape.

DAYI:

Dayi (Chinese: 大易; pinyin: dàyì, literally "big easy") uses a set of 46 character components laid out on a

standard QWERTY keyboard. A Chinese character is built by combining up to four of the 40 of the 46

characters (the other six are provided for typing Taiwanese addresses), using a system similar to that of

Cangjie, but is decomposed in stroke order instead of in geometric shape in Cangjie.

IN-DEPTH LOOK AT REGIONAL AND KEYBOARD SETTINGS

Most operating systems will let you switch between settings for your keyboard. It does not necessarily mean you have to switch the hardware keyboard, even though it may seem natural. You could type US English text with a Russian hardware keyboard for instance; it may be easier to plug in a US English keyboard for those operating system settings.

The following gives an example of settings on Windows XP to indicate which layout the user will be using, i.e. which actual software character/text will be input when clicking on a hardware or software key. Other windows systems will have similar configurations. The target locales are Chinese and Turkish for the purpose of this illustration. Other languages follow the same settings.

1. Add Chinese and Turkish to the language settings of your system. Go to Windows status bar and click "Start | Settings | Control Panel". Double click the icon "Regional and Language". Depending on your version of Windows, it may look like the following:


2. On the "Languages" tab, check the box before "Install files for East Asian Languages" or other languages as needed under section "Supplemental language support."
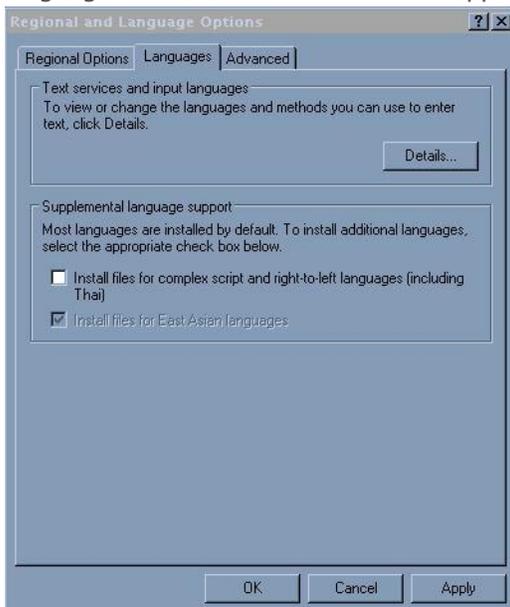

   Note: after this step, you may have to reboot the system.

3. On the Advanced tab, make sure the code pages conversion tables are setup for those programs not using Unicode. Here is the Turkish example. The East Asian languages are at the top of the list.
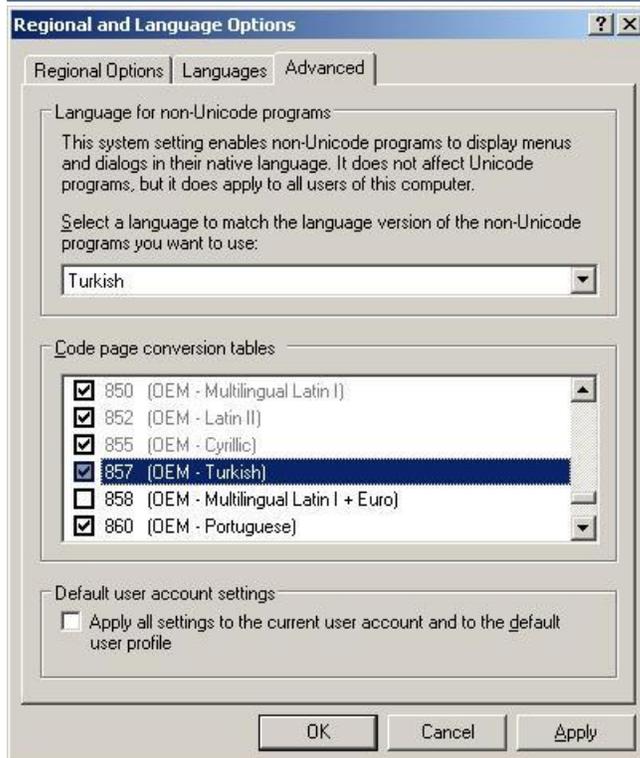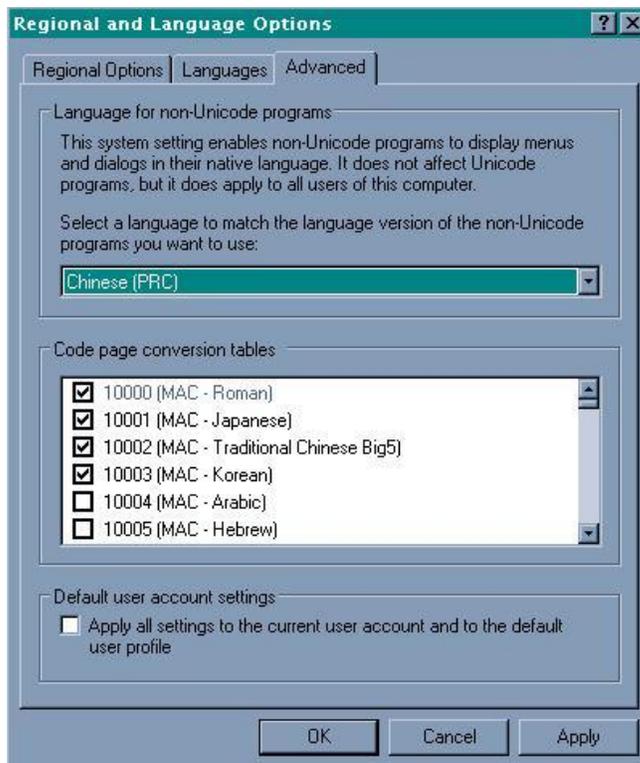
4. Open the Text Services and input Languages dialog by right clicking on the "CH" or "EN" or "TR" language symbol in the bottom right of your screen. Add the keyboard needed for Chinese (depending on the client, it could be for instance Chinese PRC – Microsoft Pinyin) or Turkish (Q for
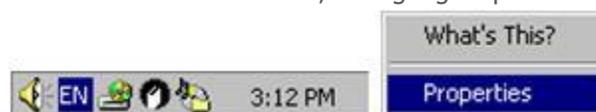
Turkish).
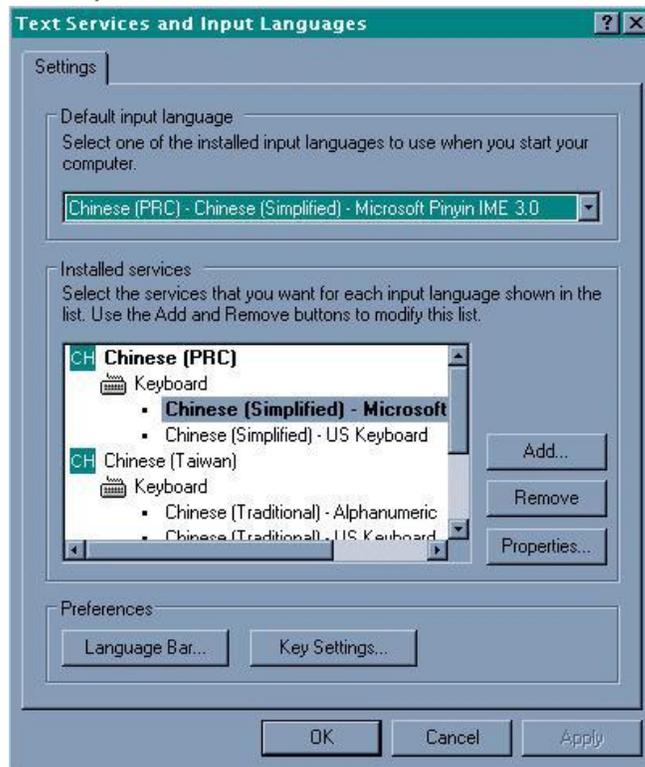
5. After Windows copied some files into your local drive, you will be prompted the following Window. Click the "Yes" button. You can then be prompted to restart Windows. Click Yes button to restart Windows.
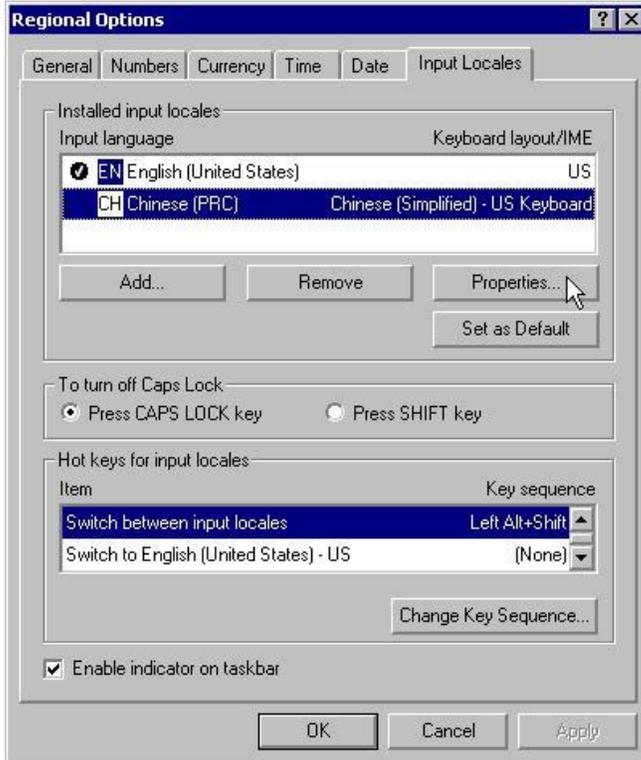

6. After Windows is restarted, a language input icon will be displayed on your Windows status bar.

Move the cursor on top of "**EN**" icon, click the right mouse button, then click the "Properties" button.

7. On the "Input Locales" tab, select "CH Chinese (PRC)   Chinese (Simplifies) – US Keyboard" entry, then click on the "Properties" button.


8. On the "Input Locale Properties" window, select the input method you prefer. If you know Pinyin, you can choose "Chinese (Simplified) – MS-PinYin98" and click OK button.

9. After you come back to the "Regional Options" window, the Chinese input icon will be changed to MS-PinYin98. If you want, you can set this input method as the default by clicking on the "Set as Default" button. Click OK button to close this window.

10. For the same language, you may have installed different keyboards. For instance, Turkish shows two different ones:
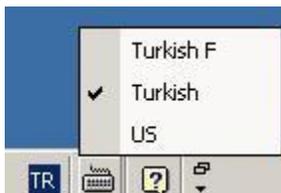
11. At that point, opening an editor like Notepad will let you type text in with a keyboard configured for a language with an input method and font set up to display the text correctly. You may still have an English hardware keyboard, but it will act, for example, like a Turkish keyboard.

ABOUT LINGOPORT

Lingoport offers a wide range of internationalization services combined with the power of Globalyzer, a client/server internationalization software system that works together to scan code and step engineers through internationalization issues in a wide variety of programming languages, to help your company achieve its global software release goals, on time and on budget.

Our teams of architects, project leaders and engineers will help you meet your global release objectives while augmenting your engineering team's product development efforts.

Lingoport's internationalization expertise and services include:

- Internationalization planning, requirements development and assessments
- Detailed code review, architectural direction, project planning
- Internationalization software development
- Globalization testing
- Internationalization training and more.

Learn more about Lingoport at www.lingoport.com and www.globalyzer.com. For a technical discussion, a free quote, or a guided tour and demonstration of Globalyzer, please contact us at info@lingoport.com or by phone at 303.444.8020 (USA).