

The many ways of converting a string from one language to another

 devblogs.microsoft.com/oldnewthing/20100507-00

May 7, 2010



Raymond Chen

A customer asked, “I’m looking for a way to convert English characters to another language. For example, if the target language is Arabic and the string is the word Hello, I want it to convert to H(Arabic)e(Arabic)l(Arabic)l(Arabic)o(Arabic).” The question is still vague, even with the assistance of the example, since it’s not clear what “H(Arabic)” means. There are a variety of ways of converting a string from one language to another. Here are a few I was able to think of.

- Translation. For example, converting *cat* to German would result in *Katze*. Of course, there are many words with no good direct translation and others which are ambiguous. *Good-bye* is normally *Auf Wiedersehen*, but if you’re saying good-bye to someone on the telephone, then it’s *Auf Wiederhören*. But at least in that case, even if you get it wrong, the reader has some idea of what you meant. Whereas *sort* could be translated as *Art* (as in *What sort of apple is this?*) or as *ordnen* (as in *Sort these alphabetically*), and if you get the wrong one, the reader is completely baffled. (Now, translation is clearly not what was intended here, but I included it for completeness.)
- Transliteration. This is an operation most commonly performed between scripts, such as how the name of the capital of China 北京 becomes *Beijing* in English. The great thing about transliteration systems is that you usually have many to choose from. For example, if you prefer Wade-Giles over Pinyin, then the capital city would be spelled *Pei-Ching*. Transliterating to non-alphabetic scripts can be quite a challenge as well: Everyone is familiar with the story of Coca-Cola in Chinese. ([Learn more about Extended Linguistic Services](#) from [Kieran Snyder, the long-distance linguist](#). I considered linking to specific articles until I realized that I was basically linking to everything, so here’s [the main page](#). Go nuts.)
- Phoneticization. This is similar to transliteration, but spells out the sounds when you cannot assume that the reader is familiar with any particular transliteration system. The capital of China would be phoneticized as *bay-jing*.

As it turns out, the customer wasn't interested in any of these! What the customer wanted was, "Take the word Hello and imagine how you would type it on a US-English keyboard. Now change the keyboard layout to Arabic and then press exactly the same keys. That's what I want." In other words, the customer wanted to see what the result would have been if you had told a blindfolded touch-typist to type *Hello*, but secretly replace the US-English keyboard with an Arabic one. Wow, that's something that hadn't even occurred to me as a possibility.

Fortunately, Michael Kaplan was able to come to the rescue by pointing the customer to the [Keyboard Convert Service](#).

[Raymond Chen](#)

Follow

