Very late remarks on the original Chinese dictionary series

devblogs.microsoft.com/oldnewthing/20060303-13

March 3, 2006

Raymond Chen

I have not forgotten about the Chinese/English dictionary series, but I simply haven't had the motivation to sit down and write up descriptions and discussion for the code that I wrote along the way, so instead of adding to the program, I'm going to answer some questions that were asked back when I started the series but which I didn't respond to at the time since I was out of town. More than one commenter suggested using `v.reserve()` to pre-allocate the vector memory. First of all, the cost of vector reallocation really didn't factor into the performance after the first few rounds of optimization, so adding a reservation step ended up being unnecessary. Furthermore, getting the correct value to pass to `v.reserve()` would mean making two passes over the dictionary, one to get the number of entries in the dictionary and set the vector reservation size, and another to fill the dictionary itself. The alternative would have been to make a guess as to the number of entries in the dictionary based on the total file size and the average length of each entry. Fortunately, it never came to that. Another commenter suggested preprocessing the file. That is also a valid technique, but I intentionally avoided it partly for expository purposes (it would have removed much of the challenge), and partly because I wanted to be able to update the dictionary by merely replacing the `dict.b5` file.

Commenter CornedBee suggested using the wcsrchr function as an alternative to the missing std::rfind method. Note, however, that the `DirctionaryEntry::Parse` method takes a string in the form of a start and end; it is not a null-terminated string. Passing this to `wcsrchr` would have resulted in quite undesirable behavior.

Raymond Chen

Follow