Abstract

The main aim of this project “Simplify Chinese Even Further” is to work out a set of simplification principles and word segmentation algorithms for written Chinese language, and then implement the algorithms to create a system that accepts ordinary Chinese texts and transforms them into segmented word combinations of simplified characters based on the prebuilt tables.

In this project, I determined a set of simplification principles to simplify ordinary Chinese characters, built a table of 1850 entries for the simplification process based on the principles. Meanwhile, I came up with a new segmentation algorithm for Chinese characters that took the use of two-dimension array (virtual table), rather than traditional FMM/RMM methods. The algorithm enables the system to find out not only whether there is ambiguity within a given sentence, but also which kind of ambiguity it is (either combinational or segmenting ambiguity). Besides, I developed a transformation system for Chinese characters, and implemented the substitution and segmentation algorithm in it. The friendly user interface also enables the user to input the original text written in ordinary Chinese characters and acquires the transformed simplified characters with spaces between words. The statistical data obtained during the testing process proves the effectiveness of the system. In addition, the analyze of the questionnaires collected from the subjects also shows that the recognizability of the transformed characters can be ensured among Chinese people, which indicates that this kind of transformation is possibly a practical way to simplify Chinese characters.