

Chinese Character-Recognition Methods for Computer Data Entry

*C*hina is the world's most populous country, and in the last decade its economy has begun to mushroom. Because modern economies rely heavily on computers, the potential market for computers in China has grown along with its economy.

COMPOSITE PERFORMANCE SCORE

(Based on a four star rating.)

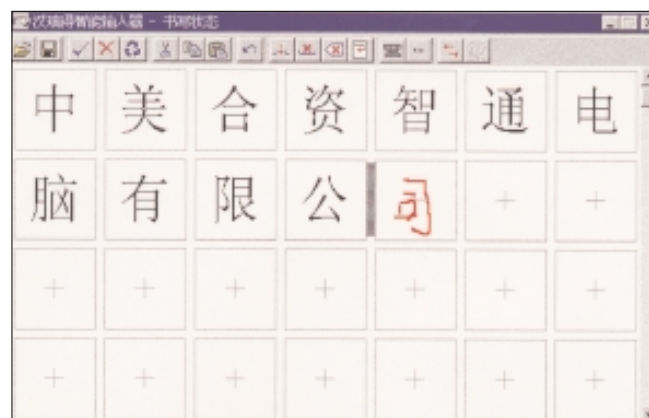


Accessing China's Giant Computer Market Potential

A major technical problem, however, impedes the widespread use of computers in China: the Chinese language is ideographic, using symbols to form characters representing things or ideas rather than letters to form words. Written Chinese employs thousands of symbols, as opposed to the 26 letters used in written English. Some keyboard methods exist for entering Chinese characters into a computer, but they are laborious. This technical barrier means that the large potential Chinese market is not readily accessible to U.S. computer businesses.

*... the large potential Chinese market
is not readily accessible to U.S.
computer businesses.*

This ATP project enabled Communication Intelligence Corporation (CIC), a small California company, to develop a stylus-and-tablet method for writing Chinese directly into a computer. CIC is a spin-off from SRI International (formerly Stanford Research Institute) and was founded in 1984 to commercialize English handwriting recognition technology. In its first ATP project, CIC developed technology for a digitized stylus-and-pad system that can be used to enter cursive handwriting in English into a computer. In this second ATP project, CIC



A screenful of Chinese characters, with one in the process of being composed. They were entered into the computer after being written on a pad using a stylus.

applied several techniques from its earlier work: using a tablet and stylus to record pen strokes, getting tablet sensory data into the computer, and using algorithms to convert graphics signals to digital form. In addition, the company created a way to recognize handwritten Chinese characters.

A System That Recognizes Nonalphabetic Writing

For the foundation of its system, CIC developed a high-quality database of about 750,000 characters penned by 2,800 Chinese writers. It also developed an algorithm that recognizes 6,763 Guojia Biao zhun characters, the standard set of characters determined by the Chinese government to be used by schools, publishers, and other institutions.

PROJECT HIGHLIGHTS

PROJECT:

To develop a Chinese character-recognition system to be used in place of a keyboard for computer entry of information in Chinese, opening Chinese markets to U.S. computer products.

Duration: 12/20/1993 — 3/19/1996

ATP Number: 93-01-0211

FUNDING (in thousands):

ATP	\$1,480	62%
Company	911	38%
Total	\$2,391	

ACCOMPLISHMENTS:

CIC fulfilled its goals by developing a recognition system for Chinese characters. The company's progress is indicated by the fact that it:

- collected a high-quality database of about 750,000 Chinese characters penned by 2,800 Chinese writers;
- developed a recognition algorithm that supports 6,763 Guojia Biaozhun characters, the standard set of characters determined by the Chinese government to be used by printers, schools, and so forth;
- entered into a joint venture with the Ministry of Electronic Industries of Jiangsu Province, China, to perform system integration services and to market the company's pen-based business computer systems to Chinese businesses and government agencies;
- released the first major product version of its character-recognition software in September 1997; and

- entered into discussions with several major U.S. computer companies about incorporating the CIC character recognition technology into their computers for sale in China.

COMMERCIALIZATION STATUS:

Commercialization is in progress. CIC recently closed its first major deal with a Chinese company to incorporate the ATP-funded technology into its products. CIC is also in discussions with major U.S. computer companies to incorporate the technology into their products for China.

OUTLOOK:

Benefits from this project are expected to accrue to U.S. companies through U.S. leadership in China's computer market, the development of computer standards in China based on U.S. technology, and large direct sales of U.S. computer components into China's markets.

Composite Performance Score: ★ ★

COMPANY:

Communication Intelligence Corporation (CIC)
275 Shoreline Drive, Suite 520
Redwood Shores, CA 94065-1413

Contact: Russ Davis

Phone: (650) 802-7757

Number of employees: 66 at project start, 93 at the end of 1997

The technology can be applied to personal computers in the People's Republic of China, the Republic of China, Taiwan and countries such as Japan and Korea, where Chinese characters are part of the written language.

The technology will also be useful in applications for other languages that use nonalphabetic writing. Most important is Japanese, which uses symbols to represent the syllables of words and employs two different syllable sets — hiragana (made with more-flowing strokes) and katakana (made with more-angular strokes). Application to handwritten Japanese is also complicated by the interspersing of Chinese characters and English words in Japanese writing.

Entering the Chinese Market

The company has entered into a joint venture, which is called CICC and has 50 employees, with the Ministry of Electronic Industries of Jiangsu (the coastal province that includes Shanghai). Under the agreement, CIC will perform system integration services and market its pen-based business computer systems (incorporating the ATP-funded technology) to Chinese business and government users. The goal of the venture is to develop and market a "Chinese computer" designed specifically to meet Chinese business requirements.

Part of the agreement specifies that the company will package U.S. hardware and office automation software as part of the Chinese computer. To implement this agreement, CIC is in discussions with several major U.S. computer companies about installing the CIC character-recognition software in their products before selling them in the Chinese market.

... entered into discussions with several major U.S. computer companies about incorporating the CIC character recognition technology into their computers for sale in China.

The sale of its products in the Chinese market will open a huge opportunity for CIC, as well as many other U.S. sellers of personal computer hardware and software in China. For a country with a population of about one billion (few of whom now use computers), the potential market is vast. But solving the technical barrier to entering

data in Chinese was a necessary step in actualizing the market and making it accessible to U.S. producers of computers and computer products.

ATP Accelerates Technology Development

CIC officials say the company was able to accomplish this technology development 18 to 24 months sooner than it could have without the ATP funds. Moreover, the ATP award helped the company develop licensing agreements and secure a joint-venture partner.

*. . . the ATP award helped the company
develop licensing agreements and
secure a joint-venture partner.*
