# 14

# Chinese Codes – the IT Cornerstone that Leads Taiwan to the World

With a cellphone, notebook computer, or 3C gadget, people in Taiwan can travel anywhere in the world. To be able to access all the information apparatuses overseas, the development of self-created Chinese codes has apparently played a pivotal role. Gratitude should be given to many anonymous heroes whose efforts have brought about unified Chinese codes, making Chinese a means of communication and message delivery in worldwide computers.

The magnitude of a country's information readiness manifests itself in the level of digitization implementation and usage, which is exercised by the government, industries and individuals. More importantly, it offers a crucial reference to measuring a country's national competiveness. Networked Readiness Index (NRI) of World Economic Forum (WEF) grades the outcome of its rating on four major facets, that is, environment, readiness, usage and impact. Beneath them are ten sub-items and 50 more detail indicators. They are used to measure the information readiness of more than 130 countries around the world. So far it has been regarded as the most representative international rating system.

During the last decade, Taiwan's ranking in WEF/NRI rating remained steady among the top 20. In 2011, it even got its best-ever performance of a No. 6 ranking in the world. Taiwan was outstanding in sub-items such as "Business and Innovative Environment", "Infrastructure", "Affordability", "Enterprise Usage", "Economic Impact" and "Social Influence". To look for the motivating factor that makes possible Taiwan's excellent outcome in the rating, there is no denying that a highly-promoted information society plays a key role.

Speaking of information promotion in Taiwan, what comes first is the Chinese dada processing technology that involves inputting, computing,

saving and outputting of data in Chinese. The fundamental element is the Chinese coding standard, which lays the very foundation for all Chinese data processing. As Chinese is an ideograph language with thousands of word symbols in its word sets, coding standards for phonogram languages, such as ISO646 (ASCII) and EBCDIC, cannot be applied to it. It is also the problem confronting several other ideograph-using Asian countries.

Around 1970, what predominantly engaged the academia and research institutes in Taiwan was the study of the inputting and outputting of Chinese data. The study included using telecom codes as a means of inputting, Chinese phoneme and font type recognition, Chinese natural language processing (NLP) and others. In 1983, as far as the industrial and commercial circles were concerned, the majority of Chinese computers put much more emphasis on the inputting and outputting of Chinese data. However, most of the products had their own coding standards, which made interchanging of the data and cross-system processing impossible among different companies. It was as in impediment to the promotion of Chinese information system.

In 1984, after the launch of IBM's first Chinese computer, IBM 5550, what appeared as a great shock to people in Taiwan was its exceptional quality in Chinese data processing. It was at this time that the Institute for Information Industry (III), the authority taking charge of the IBM 5550 project, started releasing IBM 5550's coding specifications (Big 5) and font types to the outside world. What soon followed was great improvement in the technology and quality of the Chinese computers made in Taiwan.

Ho I-tzu, then deputy Minister of National Science Council and CEO of Institute for Information Industry, instructed the III to take the Chinese Word Table promulgated by the Ministry of Education as reference. The contents of the table comprised both standard Chinese typefaces and other variant ones. As a result, the Institute for Information Industry came up with 13,051 Chinese character sets and 684 symbols. In 1986 the Table of Word Code was pronounced by the Executive Yuan as Chinese Standard Interchange Code (CNS11643) for general Chinese characters. It not only

provided the Chinese computers with basic word sets but also laid a solid footstone for the development of Chinese Information System.

To meet the needs of large computers that operated Household Registration and Conscription systems in Taiwan, the III assisted Central Bureau of Standards, MOEA, in editing the coding structures of CNS 11643. All the 48,027 Chinese characters promulgated by the Ministry of Education were coded according to CNS 11643 specifications. They formed the coding standards for various large information systems and made up the main bases for data exchanges among different systems.

Following CNS 11643 standards, the Executive Yuan went on to push forward interconnections between large information systems, such as those of the Department of Household Registration & Conscription, banks, and so forth.

To meet the requirements set for ISO 10646, the Institute for Information Industry and Taipei Computer Association co-worked with Chinese Foundation For Digitization Technology, Acer, MiTAC, Arphic, DynaLab, IBM-Taiwan, and National Language Committee of the Ministry of Education to cooperatively participate in editing the standards of CJK/JRG and ISO/IEC JTC1/SC2/WG2/IRG. Having recoded the Chinese word sets of CNS 116443 to meet the standards of ISO 10646, the information system of Taiwan can readily get connected with global markets. It also speeds up the development of Taiwan's information society and enhances the power of Taiwan's information readiness.

## 14 中文碼通國際 奠定台灣資訊化基石

帶著手機、筆電或 3C 裝置，台灣人行遍天下，能夠通行海內外所有資訊設備，**無中生有**的**中文碼**制訂扮演關鍵任務，感謝許多無名英雄努力，中文碼統一，教電腦可以用中文溝通、傳遞訊息。

資訊國力強弱可展現於政府、產業、個人對於**數位化應用**的**落實**

**(implementation)**與**活用**程度，更是衡量國家競爭力的重要依據。世界經濟論壇(World Economic Forum, WEF)的網路整備度(Networked Readiness Index, NRI)評比，以環境、**整備(readiness)**、使用和影響等 4 大**構面**、10 個**分項**、超過 50 個**細部指標，**衡量全球超過 130 個國家資訊國力表現，是當今全球最具代表性的國際**評比**。

近 10 年來，台灣在 WEF/NRI 評比穩定維持在前 20 名的領先群，2011 年更創下第 6 名的最佳成績，台灣在「商業及創新環境」、「基礎建設」、「負擔能力」、「企業使用」、「經濟影響」，以及「社會影響」等分項表現突出。探究台灣**資訊國力**亮眼表現背後的驅動因素，資訊化社會的推動扮演重要關鍵角色。

談到台灣資訊化社會的推動，最主要的就是**中文資訊處理技術**，也就是中文資訊的輸入、運算、儲存、及輸出，這裡面最基本的元素就是中文字的**編碼標準**，它是中文資訊處理的一切基礎。由於台灣使用**表意文字**，**字集**達上萬個**字符**，因此無法使用當時用於**拼音文字**的 ISO 646(ASCII)及 EBCDIC 等標準，這也是亞洲幾個使用表意文字國家所碰到的問題。

1970 年左右，當時台灣的**學術界**及研究單位主要多從事於中文輸出入的研究，例如用**電信明碼**來輸入中文資料、中文語音辨識、中文字形辨識、中文自然語言處理等相關研究。1983 年時代，在**工商界**方面，大部分的中文電腦產品也大都偏重於中文資料的輸出入處理。但大部分的產品均有各自的編碼標準，造成資料無法交換，也無法進行**跨系統**的處理，非常不利於中文資訊系統的推展。

1984 年 IBM 在台灣推出 IBM 5550 中文電腦，其在中文資訊處理方面的**優異品質**，帶給國內工商業界相當大的震撼，承接 IBM 5550 專案的資策會把此案**編碼規範**(Big 5)、**字形**開放給外界使用，國內廠商的中文電腦產品技術及品質隨之而有相當大的改進。

資策會也在當時**國科會副主委**兼資策會**執行長何宜慈(Ho I-tzu)**指示下，參考**教育部**所公布的標準**字體**暨**異體**國字**字表**，整理出 13,051 個中文**字集**、684 個**符號**，這套**字碼表**在 1986 年由行政院公布為**通**

**用漢字標準交換碼** (CNS 11643)，為中文電腦普遍採用的基本字集，並為中文資訊系統發展奠定良好的基石。

滿足台灣**戶役政**等大型電腦中文資訊系統需求，資策會進一步協助經濟部**中央標準局**修訂了 CNS 11643 **編碼架構**，並將教育部所公布的 48,027 個中文字依據 CNS 11643 的規範編碼，成為各大型資訊系統**編定字碼標準**及不同系統間資料交換的主要依據。

行政院也依據 CNS 11643 的標準陸續推動**戶役政**、銀行連線等大型資訊系統，對台灣邁向資訊化影響深遠。

因應**國際標準組織**制訂 ISO 10646 需求，資策會、**台北市電腦公會**結合**中推會**、**宏碁**、**神通**、**文鼎**、**華康**、台灣 IBM 等業者，**教育部國語會委員**參與 CJK/JRG 及 ISO/IEC JTC1/SC2/WG2/IRG 的**標準編訂**活動，將 CNS 11643 的中文字集放入了 ISO 10646 標準內，將台灣的需求融入到國際標準內，促進台灣資訊系統與國際市場的連結，並帶動台灣資訊化社會的發展及資訊國力的提升。