

Hong Kong Supplementary Character Set - 2001

**Information Technology Services Department & Official Languages Agency
Government of the Hong Kong Special Administrative Region
December 2001**

Table of Contents

Preface

Section 1 Overview

Section 2 Code Allocation of the HKSCS-2001 in Big-5

Section 3 Code Allocation of the HKSCS-2001 in ISO 10646

Annex I Lists of Characters in the GCCS Not Included in the HKSCS-2001

**Annex II List of Compatibility Points and Their Corresponding Characters in
ISO/IEC 10646-1:1993**

**Annex III List of Compatibility Points and Their Corresponding Characters in
ISO/IEC 10646-1:2000**

**Annex IV List of Compatibility Points and Their Corresponding Characters in
ISO/IEC 10646-2:2001**

Preface

Information stored in a computer or transmitted in electronic communication is coded according to a pre-defined coding scheme. For information in Chinese, there are different coding schemes including Big-5, GB (Guo Biao) and the ISO 10646 standard. These coding schemes do not cover all characters in use in the Hong Kong Special Administrative Region (HKSAR). Among the characters not included in these schemes, many are commonly used in the HKSAR. Some of them are used in names of persons and places while some are used in Cantonese expressions. To be able to use such characters in a computer system, users need to assign internal codes for these characters in the user-defined area by themselves. This works well in stand-alone computers, but when computers are connected with each other, these user-defined characters may give rise to problems in communication and data exchange.

To facilitate electronic communication within the Government, a set of Chinese characters, the Government Common Character Set (GCCS), was developed in 1995. The release of the GCCS marked the first step in coordinating the use of user-defined Chinese characters and was well received by the public as a supplement to the standard character set of Big-5. This supplementary character set was later enhanced by the Government in collaboration with the Chinese Language Interface Advisory Committee (CLIAC) which comprises representatives from academia, language and linguistics associations, the information technology industry and the publishing industry. The enhanced character set included characters collected from various sectors in the HKSAR and represented a common set for the community. It was named the Hong Kong Supplementary Character Set (HKSCS) and was published in September 1999. There are two code allocation schemes for the HKSCS, one for Big-5 and the other for ISO 10646. The HKSCS for Big-5 is in effect the updated version of the GCCS.

Recognizing the need of the public and government departments to include new characters into the HKSCS from time to time, the Government worked together with the CLIAC and published the procedures and principles for inclusion of new characters into the HKSCS in April 2000. The CLIAC meets regularly to consider applications for inclusion of characters into the HKSCS. Once approved, the new HKSCS characters will be submitted to the Ideographic Rapporteur Group, a working group under the International Organization for Standardization (ISO), for inclusion in future releases of the ISO 10646 standard. So far,

CLIAC has accepted 116 new characters after processing applications from the public and government departments.

One of the initiatives under Government's "Digital 21" Strategy for IT development is to develop an open and common Chinese language interface in the HKSAR for users who prefer to communicate electronically in Chinese. A pivotal element of this initiative is the adoption of the ISO 10646 standard and the HKSCS. ISO 10646 is an international coding standard being developed under the aegis of the ISO to embrace characters used in all major languages in the world, including traditional and simplified Chinese characters.

The ISO released the ISO/IEC 10646-1:1993 in 1993 and the ISO/IEC 10646-1:2000 in 2000. ISO/IEC 10646-1:2000 contains the 20,902 ideographic characters from ISO/IEC 10646-1:1993 plus 6,582 additional ideographic characters (a total of 27,484 ideographic characters). The ISO/IEC 10646-2:2001 was released in November 2001 with 42,711 new ideographic characters, making a total of more than 70,000 ideographic characters in the ISO 10646 standard.

The ISO/IEC 10646-2:2001 includes 1,622 additional characters from the HKSCS. Currently out of the 4,702 characters in the HKSCS, 4,677 are included in the ISO 10646 standard.

To align with the ISO/IEC 10646-2:2001 published by the ISO, and to include the above mentioned 116 new characters into the HKSCS, the Information Technology Services Department and the Official Languages Agency herewith promulgate the Hong Kong Supplementary Character Set – 2001 (HKSCS-2001) to replace the HKSCS that was published in September 1999. The HKSCS-2001 contains a total of 4,818 characters. In respect of those characters of the HKSCS-2001 that are included in ISO/IEC 10646-2:2001, their ISO 10646 code allocations have been revised accordingly with reference to the code points in the ISO/IEC 10646-2:2001. Details are described in this specification.

Acknowledgement

This specification was completed with invaluable assistance from CLIAC. We would like to thank the following committee members for their contribution - current members : Dr. CHAN Chun Leung, Mr. CHAN Chung Ling, Ms. Yolanda CHAN, Mr. Dominic CHENG, Dr. CHEUNG Kwan Hin, Ms. Cordelia CHUNG, Mr. HO Lap Yiu, Dr. KEUNG Wing Ching, Ms. Annie LOK, Dr. LU Qin, Dr. LUKE Kang Kwong, Mr. Caesar LUN, Mr. Lawrence MO, Mr. Arics POON, Dr. Elizabeth QUAT and Prof. Benjamin TSOU, and past members: Mr. Raymond CHENG, Mr. Francis LEE, Mr. Fan LOOK, Mr. Tony TAI, Mr. H.T. TSANG and Mr. T.H. YAO.

We would also like to thank the two working groups under CLIAC for their great efforts: the Working Group on Characters for Chinese Computer Systems chaired by Dr. CHEUNG Kwan Hin and the Working Group on Chinese Information Processing chaired by Dr. LU Qin. Members of the two working groups consist of some CLIAC members and Mr. FAN Kwok, Prof. LEE Kin Hong, Mr. Terence LEE and Mr. TSE Ka Ho.