CLIAC Paper 10/99
(Updated)

An Implementation Strategy for the Adoption of ISO 10646 in Hong Kong SAR

Introduction

This paper seeks Members' views on the proposed implementation strategy for the adoption of ISO 10646 in Hong Kong Special Administrative Region (HKSAR).

Background

2. In August 1999, the Chinese Language Interface Advisory Committee (CLIAC) has finalised the Hong Kong Supplementary Character Set (HKSCS). This is one of the major steps towards the adoption of the ISO 10646 in HKSAR.

Current Situation

3. In parallel with the development of the HKSCS, the Information Technology Services Department (ITSD) of the HKSAR Government has recently taken steps to ascertain the current situation on the adoption of ISO 10646. The findings are summarised below :

(a) There are positive supports to the direction of adopting ISO 10646 from a number of the local subsidiaries of the major hardware and software vendors as well as the local small and medium enterprises (SMEs) in the IT industry. Most of them have already provided products supporting ISO 10646 (using Unicode). To support new releases of ISO 10646, they estimated that about 3 to 6 months are normally needed to enhance their products. They also considered that the Government should take a leading role, and that a clear and formal announcement would help local subsidiaries to persuade their headquarters to develop products using ISO 10646.

(b) There are only a few projects in local universities that support ISO 10646 or address the need for data interchange between different coding standards.

(c) ISO 10646/Unicode is getting more popular and is supported by products in areas like office automation, Internet browser and applications, and telecommunications devices etc.

Implementation Strategy For Adoption of ISO 10646 in HKSAR

4. To facilitate the adoption of ISO 10646 in HKSAR, a strategy has been formulated to address various issues before and during the transition to ISO 10646. The proposed strategy is detailed in the following sections.

Coordinated Use of Private-use-area for Supplementary Characters

5. The primary issues we face in using Big-5 or ISO 10646 are the uncoordinated use of the limited code space and the need for the continuous inclusion of supplementary characters in the private-use-area of these coding schemes. Users will need additional supplementary characters continuously for their specific requirements, e.g. names of persons and new scientific terminology.

6. It is recommended that a central mechanism comprising policy, organisation and procedures be established to better coordinate and manage requests for inclusion of additional supplementary characters from the community and for the allocation of code points in the private-use-area. The proposed main functions of this central body are as follows :

(a) To manage applications for inclusion of new characters and appeal cases;

(b) To define validation and inclusion criteria for new characters;

(c) To recommend areas of code space in the private-use-areas of Big-5 and ISO 10646 for individual use by the public and for HKSCS, and to allocate code points to new characters in the HKSCS;

(d) To prepare and maintain mapping tables including those of HKSCS between different coding schemes, say, Big-5 and various extensions of ISO 10646;

(e) To publish and announce updated HKSCS to the public; and

(f) To advise on the submission and validation of characters for the further inclusion into the ISO 10646 standard.

7. Currently, the CLIAC is performing some of the above functions. In the longer term, a new central body should be set up to discharge these functions.

8. To alleviate the contention for code space in the private-use-area, we would explore the feasibility of securing a dedicated area outside the current private-use-area of ISO 10646 for the purpose of temporary storage of new characters of individual territories (e.g. HKSCS) between releases of extensions. Should this dedicated area be allocated, the entire private-use-area can then be released to the general public. It may however take a long time to secure the acceptance of the members of IRG and other working groups of ISO.

Legal Measures for Better Management of Supplementary Characters

9. The community will require additional supplementary characters to cater for their needs on a continuous basis. While the central body will perform the ongoing management of the HKSCS, we shall consider the need for legislative measures to empower this central body to carry out the functions and to give a statutory status to the supplementary character set managed under its jurisdiction. A working group under the CLIAC has already started some preliminary studies on this issue.

10. Initial findings indicate that there are no such laws or regulations in the Mainland and in Taiwan for the management of new characters to be created. Further reference needs to be made to ascertain whether there are laws or regulations facilitating the management of such elsewhere.

Measures for promoting the adoption of ISO 10646

11. To promote the adoption of ISO 10646, we have proposed the following measures:

(a) Adoption of ISO standard as a mandatory requirement of government tenders at the earliest opportunity, say, three months after the formal release of ISO 10646 version 2 (which contains characters of Extension A), to set an example to encourage the adoption of ISO 10646 by the IT industry and the community;

(b) Discontinuation of the enhancement of the HKSCS for Big-5 to encourage the adoption to ISO 10646 from a certain date, say, after the release of ISO 10646 Extension B; and

(c) Seminars, road shows and exhibitions to promote the adoption of ISO 10646 in the community and in the industry (e.g. for content hosting and for mobile computing).

Measure to Facilitate Effective Data Interchange and Conversion

12. To handle the different coding schemes and backward compatibility issues, there is a need for technical solutions to ensure correct round-trip conversion (from Big 5 to ISO 10646 and back) in data interchange and to enhance interoperability. From our recent market research, it is found that several technical solutions (e.g. the use of mapping tables, portable document formats, extensible markup language (XML)) are available for this purpose. We shall closely liaise with the IT industry for more solutions in this area.

13. To ensure the timely adoption to ISO 10646 and its extensions when the standards are formally announced by ISO, we will encourage the IT industry and academia to undertake more research and development on software supporting data interchange and conversion under the ISO 10646 environment so that users will not be unduly delayed in their adoption of the standard.

14. In this regard, the Government has been providing support under the Industrial Support Fund (ISF) for the development of migration and document interchange tools to facilitate the migration from the Big-5 standard to ISO 10646. The Government shall continue to render its

support when new research and development projects conducive to the adoption of ISO 10646 in HKSAR are available.

15. Developers and academia would be encouraged to give priority to the areas of research and development described below :

(a) *Conversion Tools using XML*
Projects for studying and developing conversion solutions for handling Chinese characters in XML files, using the customised language tags.

(b) *Internet Email*
Email is very popular today for communication over Internet. It is a common practice of many users to keep the content of original mails in their reply, forming a chain of emails which may be of different coding standards. However, existing email products do not cater for an email with different Chinese coding schemes. A solution is highly desirable to address this problem.

(c) *Web Browsing*
Currently, Chinese web sites may be using different standards in encoding Chinese characters. While solutions exist for browsers to display Chinese web pages in different codings, not all of them support the display of characters in the HKSCS. Developers of browsers would be encouraged to implement the mapping of the HKSCS between ISO 10646 and other coding schemes in their products. Local academia and developers would also be encouraged to identify alternative solutions.

(d) *Office Automation Systems*
Some common OA applications already support both Big-5 and ISO 10646 but not HKSCS in their native mode. Developers would be encouraged to implement the mapping of the HKSCS between ISO 10646 and other coding schemes in their OA products. Local academia and developers would also be encouraged to identify alternative solutions.

(e) *Input / Output of Chinese Characters*
Products supporting ISO 10646 and HKSCS, e.g., input

methods, and fonts for display and printing.

(f) *Other Research Areas*
Search engine, translation, voice recognition, optical character recognition and linguistics corpus.

Road Map for the Adoption of ISO 10646 in HKSAR

16. A road map for the adoption of ISO 10646 in HKSAR is at the Appendix. It contains various paths for users to adopt or migrate to ISO 10646.

(a) *Stage 1*
With the announcement of the HKSCS, users are encouraged to use this character set as part of the common Chinese language interface for the local users in transition to ISO 10646.

(b) *Stage 2*
When ISO 10646 version 2 is formally released in around year 2000, new systems should adopt the ISO 10646 standard where possible. This is also an opportunity for the existing Big-5 systems to migrate to the ISO 10646 enabled platforms.

(c) *Stage 3*
When ISO 10646 Extension B is available, the existing Big-5 systems are strongly encouraged to take this very good opportunity to migrate to the ISO 10646 enabled platforms.

(d) *Stage 4*
Enhancement to HKSCS for Big-5 will be discontinued. Only HKSCS for ISO 10646 will be further maintained.

Advice Sought

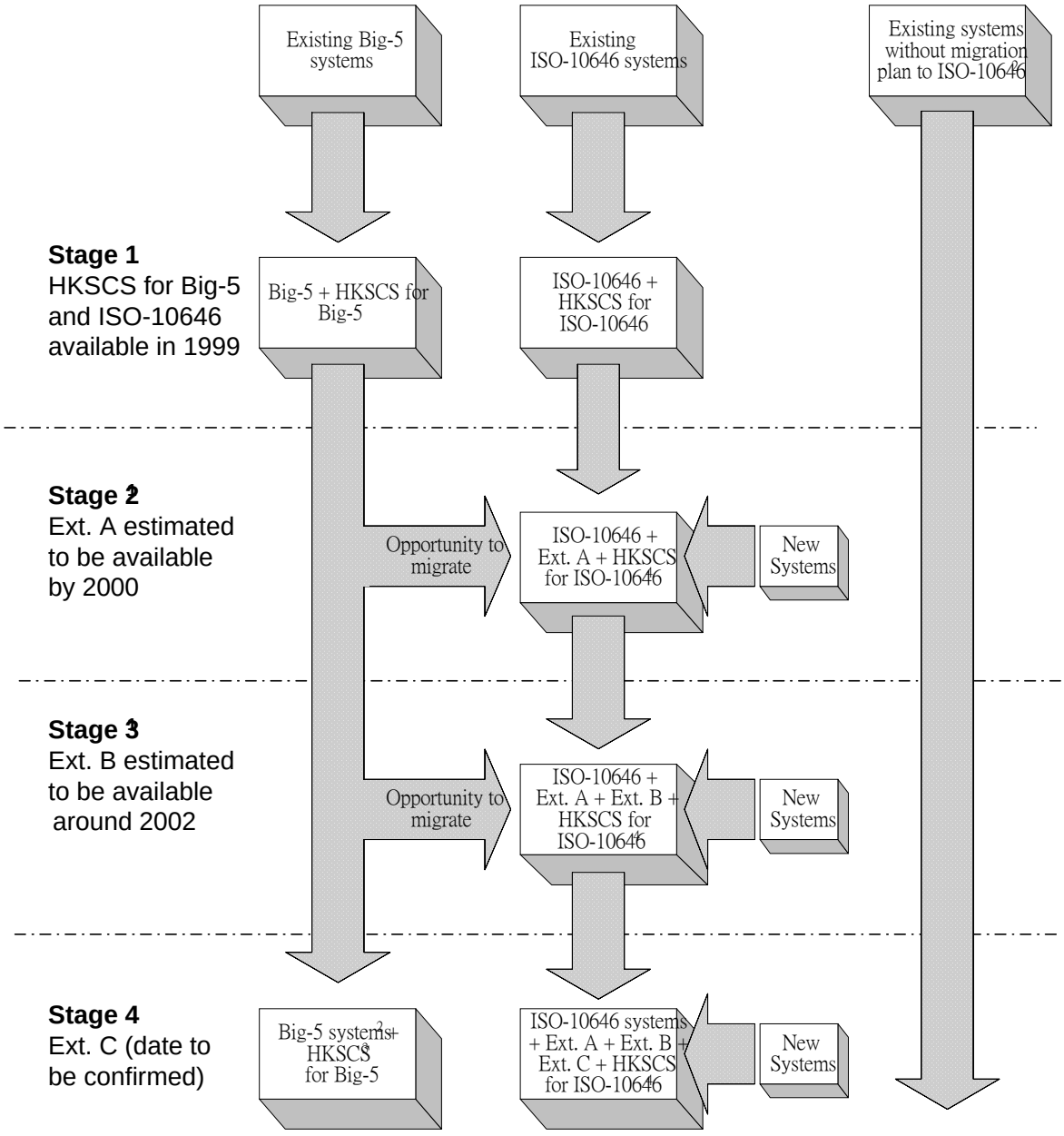
17. We welcome Members' views on the implementation strategy for the adoption of ISO 10646 in HKSAR.

Information Technology Services Department

September 1999

Appendix

Roadmap for the Adoption of ISO 10646 in HKSAR



1. During Stage 2 and 3, HKSCS for Big5 would be continually enhanced to include additional supplementary characters collected from users.
2. These systems have to maintain the prevailing mapping tables or conversion tools to communicate with those ISO 10646 systems.
3. HKSCS for Big5 would not be further enhanced.
4. HKSCS for ISO 10646 would be further enhanced.