出國報告(出國類別：出席國際會議)

出席ISO/IEC JTC1/SC2/WG2/IRG 第25次會議

服務機關：經濟部標準檢驗局
姓名職稱：莊組長素琴、陳技正星光
派赴國家：美國舊金山
出國期間：94 年 11 月 27 日至 12 月 4 日
報告日期：95 年 2 月 3 日

摘　　要

本次會議由Unicode協會主辦，共計32人出席，主要目的在續審各會員體所申請擴編ISO 10646之CJK_C1漢字集、討論古漢字與筆畫之整理及確認IRG第26、27次會議舉辦地點與時間。

依會議決議，IRG首席編輯將從CJK_C1選出一組穩定小字集審查，而各會員體須提交該等小字集的IDS屬性給IRG首席編輯。另古漢字的整理工作，各會員體先就《說文解字》前180個部首進行甲骨文的蒐集和整理工作，以便建置甲骨文資料庫。

ISO 10646系列標準可說是少數我國能夠積極參與制訂與維護的國際標準之一，後續IRG工作組將再進行有關漢字集的字集、中文字的屬性及古漢字的整理等多項工作，爰我國應能長期參與該工作組活動，並積極貢獻，才能建立及維持我國在該領域的影響力。

目　　次

壹、會議概要及目的 - 1

貳、會議過程 - 4

參、會議議題及決議 - 7

肆、我國應配合辦理之工作 - - - - - - - - - - - - - - - - 9

伍、心得與建議 - 10

陸、附件 - 11

壹、會議概要及目的

一、會議時間

94年11月27日至12月4日。

二、會議地點

美國加州柏克萊大學Dwinelle大樓370室。

三、主辦單位

Unicode協會(The Unicode Consortium)。

四、出席人員

本次會議共計32人出席會議，分屬台灣、中國大陸、香港、日本、南韓、越南及Unicode協會，與會人員名單如下：

姓　　名	會 員 體	服　務　單　位
陸勤(Lu, Qin)	IRG主席	香港理工大學電子計算學系
莊素琴(Chuang, Suh-chyin)	台灣(TCA)	經濟部標準檢驗局
曾士熊(Tseng, Shih-shyeng)	台灣(TCA)	中央研究院計算中心
魏林梅(Wei, Lin-mei)	台灣(TCA)	中文數位化技術推廣基金會
許學仁(Hsu, Hsueh-Jen)	台灣(TCA)	花蓮師範學院
陳星光(Chen, Hsin-kuang)	台灣(TCA)	經濟部標準檢驗局
陳信誠(Chen, Hsin-cheng)	台灣(TCA)	主計處電子資料處理中心
許其清(Hsu, Chi-ching)	台灣(TCA)	台灣IBM公司
陳文男(Chen, Wen-nan)	台灣(TCA)	中文數位化技術推廣基金會
鍾祥仁(Chung, Hsiang-jen)	台灣(TCA)	資訊工業策進會
陳壯(Chen, Zhuang)	中國大陸	中國電子技術標準化研究所
代紅(Dai, Hong)	中國大陸	中國電子技術標準化研究所
尹江紅(Yin, Jianghong)	中國大陸	北京方正電子公司
吳健(Wu, Jian)	中國大陸	中國科學院軟件研究所
李國英(Li, Guoying)	中國大陸	北京師範大學
張力偉(Zhang, Liwei)	中國大陸	北京中華書局

姓　　名	會員體	服　務　單　位
史建橋(Shi, Jianqiao)	中國大陸	商務印書館
張德劭(Zhang, Deshao)	中國大陸	華東師大中文研究與應用中心
秦少雄(Chuen, S.-h. Francis)	香港	香港政府資訊科技署
鄭偉康(Cheng, Wai-hong, Peter)	香港	香港政府法定語文事務處
關口正裕(Sekiguchi, Masahiro)	日本	富士通公司
小林龍生(Tatsuo, Kobayashi)	日本	Justsystem數位文化研究中心
阿南康宏(Anan, Yasuhiro)	日本	微軟產品開發公司
川幡太一(Taichi, Kawabata)	日本	日本電訊公司
山本知(Yamamoto, Satoshi)	日本	日立公司
李承宰(Lee, Seungiae)	南韓	國立韓國語言研究院
鄭雨峰(Chung, Woo Bong)	南韓	高麗大學
朴鍾寓(Park, Jong-woo)	南韓	高麗大學
吳中越(Ngo, Trung Viet)	越南	越南資訊技術研究所
阮光紅(Nguyen, Quang Hong)	越南	社會科學與人文中心漢喃研究所
Cook, Rechard	Unicode協會	柏克萊大學
畢夏普(Bishop, Tom)	Unicode協會	Wenlin Institute, Inc.

五、會議目的

爲容納全球各種語言的字元和符號，ISO國際標準組織的會員國於1984年組成ISO/IEC JTC1/SC2/WG2（簡稱WG2），負責制定新的國際字元集編碼標準“Universal Multiple-Octet Coded Character Set”（簡稱UCS），編號則訂爲ISO/IEC 10646（簡稱ISO 10646）；其中「中日韓認同表意文字區」即爲收容亞洲各國所使用之漢字。因漢字集的規模龐大，且爲多個國家（地區）共同使用，WG2爲此再設「表意文字小組」(Ideograph Rapporteur Group,簡稱IRG)，專責蒐集、比對各國（地區）之中文字集，再向WG2提出彙整後漢字集；我國自73年起即由台北市電腦公會代表參與該標準草案之編訂，嗣後由本局專案委託財團法人資訊工業策進會、中央研究院、財團法人中文數位化技術推廣基金會等辦理將我國CNS 11643「中文標準交換碼」資訊用字納入ISO 10646國際標準中，並代表出席相關會議。CNS 11643「中文標準交換碼」分爲16字面，其中第1～7字面列有教育部所公布的常用、次常用、罕用及異體等字集，第8～11字面預留空白，而12～16爲戶政等系統之使用者加字區，多年來經各相關單位，CNS 11643第1～7字面的中文字集，已全數編入ISO 10646，包括ISO 10646-1「廣用多八位元編碼字元集（UCS）－基本多語文字面」計23,989個字及ISO 10646-2「廣用多八位元編碼字元集（UCS）－輔助字面」計29,950個字。

本次會議，本次會議將續審各會員體所申請擴編ISO 10646之CJK_C1漢字集、討論古漢字與筆畫之整理及確認IRG第26、27次會議舉辦地點與時間。我國已在第19次會議提出戶政用資訊用字約1萬字，申請編入ISO 10646之CJK_C1漢字集；另依前次會議（即IRG第25次）決議，IRG第27次會議將由台灣主辦，時間爲95年11月27日～12月1日。

貳、會議過程

◆ 11月27日（星期日）：去程

◆ 11月28日（星期一）上午：

1、主辦單位至歡迎詞。

2、主席宣布開會並說明本次會議重要待議事項。

3、與會人員逐一自我介紹(roll call)。

4、確認議程。

5、主席指派大會秘書及決議起草小組（各會員體派一名代表）。

6、各會員體代表報告各自相關活動，我國的活動報告請詳附件IRG-N1146。

7、首席編輯報告自IRG第24次會議後CJK_C1審查的情況。

8、討論IRG所提重新檢討CJK_C1審查程序與時程議題。

9、主席將出席人員分配成CJK_C1、古漢字及筆畫等三個工作組。

◆ 11月28日（星期一）下午：

CJK_C1工作組：

（1）日本代表川幡太一說明他所提出的IDS（表意文字描述序列，Ideographic Description Sequences）拆字原理。

（2）各會員體討論IDS的可行性。

（3）決定採用川幡太一的IDS拆字原理做爲拆解CJK_C1字集的基礎，明天上午以兩個小時用該拆字原理嘗試拆解CJK_C1的漢字，藉以發現該拆字原理的問題和需改進之處。

◆ 11月29日（星期二）上午：

CJK_C1工作組：

就首席編輯事先從CJK_C1字集中隨意挑選的100個字，利用川幡太一的IDS拆字原理嘗試進行拆解，看看兩個小時內可以拆解多少個字。

◆ 11月29日（星期二）下午：

1、CJK_C1工作組：繼續審查CJK_C1-D42字集。

（1）討論首席編輯所提，用以記錄CJK、CJK_A及CJK_B錯誤（包括該認同而未認同以及不該認同而認同者）的紀錄格式。

（2）討論CJK_C1後續審查工作如何進行。

2、古漢字工作組：

（1）溝通對古漢字「隸定」的觀點與作法。

（2）確認進行古漢字整理工作的步驟。

◆ 11月30日（星期三）上午：

1、CJK_C1工作組：

（1）討論昨天下午對「CJK、CJK_A及CJK_B錯誤紀錄格式」議題的初步結論。

（2）討論日本代表對改進CJK_C1審查程序的建議案。

（3）討論是否把CJK_C1字集中有緊急編碼需求的漢字挑出，先行完成審查、編碼程序。

2、古漢字工作組：

（1）討論古漢字字集提交原則，包含隸定、處理、編輯。

（2）討論古漢字字集對應現代漢字字集原則。

（3）訂定古漢字（甲骨文）字集提交日期、進度及格式。

◆ 11月30日（星期三）下午：

1、大會：

柏克萊大學研究員Debora Anderson和Unicode協會副總裁Rick McGowan介紹Unicode的文字編碼工作，以及WG2和UTC(Unicode Technical Committee)的概況。

2、CJK_C1工作組：

（1）繼續討論日本代表對改進CJK_C1審查程序的建議案。

（2）討論新的CJK_C1審查程序。

（3）起草本組報告第1部（IRG-N1167）。

3、古漢字工作組：與Rick McGowan討論古漢字整裡的構想和作法。

4、筆畫工作組：討論筆畫認同。

◆ 11月30日（星期三）晚上：大會晚宴。

◆ 12月1日（星期四）上午：

1、CJK_C1工作組：

（1）日本代表川幡太一解說日本、香港和南韓編輯嘗試依據IRG N1153拆解50個CJK_C1漢字的結果。

（2）討論我國與Unicode協會聯名提案的IRG N1173及日本回應的IRG N1175，做成不強制遵守IRG N1153以及允許一字多個IDS的結論。

（3）起草本組報告第2部（IRG-N1179）。

2、筆畫工作組：重新檢查與確認所有筆畫。

◆ 12月1日（星期四）下午：

1、決議起草小組：草擬本次會議決議。

2、古漢字工作組：起草本組報告（IRG-N1182）。

3、筆畫工作組：起草本組報告（IRG-N1174）。

4、大會：

（1）張德劭報告古漢字工作組討論成果（IRG-N1182）。

（2）阿南康宏報告筆畫工作組討論成果（IRG-N1174）。

（3）討論並通過本次會議決議（IRG-N1158）。

◆ 12月2日（星期五）：參訪活動。

◆ 12月3日（星期六）、12月4日（星期日）：回程

參、會議議題及決議

一、關於CJK_C1字集的整理工作（IRG-N1158的M25.3及M25.4）：

1、暫停目前的CJK_C1字集審查工作。

2、就CJK_C1_v52字集中，挑選未曾被質疑有錯獲必須被認同者，藉助IDS檢查是否與CJK、CJK_A、CJK_B或CJK_C1有重複字。

3、CJK_C1字集整理工作的會後時程如下：

完成日期	工作項目
95_01_06	IRG 首席編輯從 CJK_C1_v52 選出一組穩定的漢字（以下簡稱小字集）分送給各會員體首席編輯。
95_02_10	會員體首席編輯提交 CJK_C1 小字集的 IDS 屬性給 IRG 首席編輯。
95_03_03	IRG 首席編輯分送 CJK_C1_v60 小字集（含字樣）給會員體首席編輯審查。
95_04_07	會員體首席編輯提交對 CJK_C1_v60 小字集的審查意見給 IRG 首席編輯。
95_04_21	IRG 首席編輯分送彙整的審查意見給會員體首席編輯，俾便確認。
95_04_28	會員體首席編輯提交確認結果給 IRG 首席編輯。
95_05_19	IRG 首席編輯分送 CJK_C1_v61 小字集給會員體首席編輯審查。
IRG 第 26 次會議前	會員體首席編輯審查完成 CJK_C1_v61。

二、關於IDS規則（IRG-N1153、IRG-N1173、IRG-N1158的M25.2）：

1、CJK_C1工作組同意採用日本代表川幡太一所提出的IDS規則（IRG-N1153）分析CJK_C1小字集的IDS，以便和已建置好的CJK、CJK_A和CJK_B字集的IDS比對，找出可能的重複字，以提高CJK_C1的品質。

2、我國代表和Unicode協會代表都覺得IRG-N1153的IDS規則太多限制，將造成各會員體進行IDS分析時的困擾，因此提出放鬆限制的建議（IRG-N1173）。經討論後，達成IRG-N1153僅供參考而非強制的決議（IRG-N1158的M25.2）。

三、關於CJK、CJK_A及CJK_B字集的錯誤報告（IRG-N1158的M25.6）：

由IRG首席編輯就WG2-N3002的Annex I、IRG-N1164、IRG-N1172及N1178研擬用於報告CJK、CJK_A及CJK_B字集錯誤的報告格式及填表規定。

四、關於古漢字的整理工作（IRG-N1182及IRG-N1158的M25.5）：

1、採取先建立再分工的工作模式。

2、先就《說文解字》前180個部首進行甲骨文的蒐集和整理工作，於95年5月8日前提交，以便建置甲骨文資料庫。提交資料的格式詳見IRG-N1182第1項，資料庫的資料檔記錄格式內容詳見IRG-N1182第2項及第4.(3)項。

五、關於筆畫的整理工作（IRG-N1174及IRG-N1158的M25.7）：

1、延續IRG第24次會議未完成的議題，繼續討論筆畫認同。

2、原44個筆畫經認同10個筆畫以及新增3個筆畫之後縮減爲37個，我國所提的筆畫之中，長頓點及撇挑兩筆畫被認同。

3、將有21個新筆畫（扣除WG2已編碼的16個筆畫）提交WG2第48次會議。

六、未來議程

1、IRG第26次會議：由越南主辦。

地點：越南順化。

日期：95年6月5～9日。

2、IRG第27次會議：暫訂由台灣主辦。

地點：台北市。

日期：暫訂爲95年11月27日～12月1日。

3、IRG第28次會議：暫訂由中國主辦。

地點：杭州。

日期：96年5、6月間。

4、IRG第29次會議：暫訂由美國主辦（待確認及授權）。

肆、我國應配合辦理之工作

一、關於CJK_C1字集的整理工作：

1、95年1月6日前IRG首席編輯從CJK_C1_v52選出一組穩定的漢字（以下簡稱CJK_C1小字集）分送給各會員體首席編輯，而各會員體首席編輯須於95年2月10日前提交CJK_C1小字集的IDS屬性給IRG首席編輯。

2、95年3月3日前IRG首席編輯分送CJK_C1_v60小字集（含字樣）給會員體首席編輯審查，而各會員體首席編輯須於95年4月7日前提交對CJK_C1_v60小字集的審查意見 給IRG首席編輯。

3、95年4月21日前IRG首席編輯分送彙整的審查意見給會員體首席編輯，俾便確認，而各會員體首席編輯須於95年4月28日前提交確認結果給IRG首席編輯。

4、95年5月19日前IRG首席編輯分送CJK_C1_v61小字集給會員體首席編輯審查，而各會員體首席編輯須於IRG第26次會議前審查完成。

二、關於古漢字的整理工作

各會員體先就《說文解字》前180個部首進行甲骨文的蒐集和整理工作，於95年5月8日前提交，以便建置甲骨文資料庫。提交資料的格式詳見IRG-N1182第1項，資料庫的資料檔記錄格式內容詳見IRG-N1182第2項及第4.(3)項。

三、關於舉辦IRG第27次會議

IRG第27次會議將由台灣主辦，地點爲台北市，時間暫訂爲95年11月27日～12月1日。

上述應配合辦理之工作，除第一項第1點「各會員體首席編輯須於95年2月10日前提交CJK_C1小字集的IDS屬性給IRG首席編輯」工作因時間較爲緊迫，爰已商請本局94年度「中文資訊及資通術語標準之維護及推廣」委辦計畫之受託單位中文數位化技術推廣基金會協助幫忙處理外，其餘部分將促請本局95年度「中文資訊標準之維護及推廣」委辦計畫之受託單位配合辦理。

伍、心得與建議

一、依據IRG決議，CJK_C1字集將分批審查，第1批將挑選具有來源事實（例如字典、文獻、政府文書等的影本）之5,000個漢字提交WG2，爰爲使我國所提1萬字戶政系統用字能順利納入ISO 10646 CJK_C1字集，建議應重新將該等字集與教育部異體字典比對，並予以檢視，將無來源事實的字暫時移出。

二、依據IRG決議，95年1月6日前IRG首席編輯從CJK_C1_v52選出一組穩定的漢字（以下簡稱CJK_C1小字集）分送給各會員體首席編輯，而各會員體首席編輯須於95年2月10日前提交CJK_C1小字集的IDS屬性給IRG首席編輯，由於本項工作時間較爲緊迫，已商請本局94年度「中文資訊及資通術語標準之維護及推廣」委辦計畫之受託單位中文數位化技術推廣基金會協助幫忙處理。另其他有關CJK_C1字集的整理工作，將促請本局95年度「中文資訊標準之維護及推廣」委辦計畫之受託單位配合辦理。

三、有關古漢字的整理工作，依據IRG決議，各會員體先就《說文解字》前180個部首進行甲骨文的蒐集和整理工作，於95年5月8日前提交，以便建置甲骨文資料庫，提交資料的格式詳見IRG-N1182第1項，資料庫的資料檔記錄格式內容詳見IRG-N1182第2項及第4.(3)項，本項工作將促請本局95年度「中文資訊標準之維護及推廣」委辦計畫之受託單位配合辦理。

四、IRG第27次會議將由台灣主辦，地點爲台北市，時間暫訂爲95年11月27日～12月1日，此對非屬ISO會員國之我國，誠屬不易，爰爲能順利召開此次會議，本局於94年11月2日先行邀集行政院主計處及教育部國語推行委員會等相關單位，討論相關分工事宜，初步達成由行政院主計處、教育部國語推行委員會與標準檢驗局共籌經費辦理該次會議，惟依以往經驗，IRG第27次會議是否順利於我國召開，應考量IRG首席編輯者及相關工作組之編輯者（皆由大陸人士擔任）能否赴台出席會議之因素，爰本局將密切與IRG保持聯繫，並促請如何任何變數，應儘速告知，俾便我能及早應。

五、ISO 10646可說是少數我國能夠積極參與制訂與維護的國際標準之一，後續IRG工作組將再進行有關漢字集的字集、中文字的屬性及古漢字的整理等多項工作，爰我國應能長期參與該工作組活動，並積極貢獻，才能建立及維持我國在該領域的影響力。

陸、附件

1、IRG_N1146：我國活動報告。

2、IRG_N1153：IDS（Ideograph Description Sequence，表意文字描述序列）規則。

3、IRG_N1158：本次會議決議。

4、IRG_N1165：我國對古漢字平台及工作程序的建議。

5、IRG_N1167及N1179：CJK_C1工作組報告。

6、IRG_N1172：CJK、CJK_A及CJK_B字集錯誤的報告格式。

7、IRG_N1173R：我國及Unicode協會對IRG-N1153的意見。

8、IRG_N1174：筆畫工作組報告。

9、IRG_N1178：CJK、CJK_A 及 CJK_B 字集錯誤的填表規定。

10、IRG_N1182：古漢字工作組報告。

附件 1

IRG_N1146：我國活動報告

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1146

Date: 2005-11-26

Doc. Type:	Member body contribution
Title:	Activity Report form TCA
Source:	TCA
Status:	Input to IRG
Action:	For Your Information
Distribution:	IRG Members and Ideographic Experts
Reference:	
No. of pages:	2
Medium:	Electronic

1. New standard on the structure of Chinese characters:
   Based on the project from BSMI (Bureau of Standards, Metrology and Inspection, the organization for standardization in Taiwan), CMEX (Chinese Foundation for Digitization Technology) proposed a standard drafts, named "CNS 11643-3: Basic components and component attributes for Chinese characters". It specified 517 basic components which could be used to describe the structure of a Chinese character. Based on those basic components, the standard specified a component attribute to each Chinese character in the planes 1 & 2 of CNS 11643. The component attribute is a sequence of basic components for describing the structure of a Chinese character. The component attribute could play the role similar to IDS. In addition, CNS 11643-3 also contained the conformance clause that specifies the guidance for checking whether a new Chinese character could be accepted as a new one in CNS 11643.

2. Promotion works for CNS 11643 serial standards and CNS 14649 (i.e., the Chinese version of ISO/IEC 10646):
   On June, BSMI hosted 3 seminars, respectively in Taipei, Taichung and Kaohsiung, for introducing those standards to government staffs and IT engineers.
   From May to September, we host 24 seminars for college staffs and students, and 9 seminars for public. More than 1,000 people attended those 31 seminars.

3. Review works on CJK-C1 v5 sets:
   We assigned dedicate people to perform the review work assigned by the IRG Chief Editor. We also host the relative meeting once a month for discussing the review work.

4. For the old Hanzi encoding:
   On October 26-28, we host an ad hoc meeting for the old Hanzi encoding. Experts from Taiwan and mainland China made some conclusions in the meeting. All experts agreed that the work of old Hanzi encoding is very important and they'll contribute more effort on cooperating work. Next ad hoc meeting will be host by mainland China.

5. New standard fonts for Chinese characters:
   The MOE (Ministry of Education) in Taiwan had published 2 new standard fonts for the Chinese characters in the planes 1 & 2 in CNS 11643. They were the true type fonts in Song style and Kai style respectively. The true type fonts in Song style for the Chinese characters in the planes 3-7 of CNS 11643, including the ideographs in CJK-B, are reviewed by the ideographic experts right now. They'll be published in next year by the MOE.

附件 2

IRG_N1153：IDS（Ideograph Description Sequence，表意文字描述序列）規則

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1153
Date: 2005-11-21

Source:	Japan
Title:	Guidelines on IDS Decomposition.
Status :	
Actions required	Review by IRG Editors for discussion at IRG meeting No. 25.
Distribution:	IRG Members and Ideographic Experts
Medium :	Electronic

1. Backgrounds

The authors believe that the use of IDS greatly helps the standardization works of CJK UNIFIED IDEOGRAPHS family of characters, especially during the review process. With IDS, we can find ***similar ideographs*** much more easily than ever, helped by a small program.

IDS database for already standardized ideographs, i.e., those for all ideographs in URO, Ext-A, and Ext-B already exist, although it may contain some errors. However, such errors make no serious problem. (We anyway need eye-to-eye review). Just consider the IDS-based report the program generates a suggestion or candidates. Even if the accuracy of the IDS based review is 90%, it greatly reduces the required workloads to look up duplicates from nearly a hundred thousands of ideographs.

An ideograph often can be divided into two or more different IDSs. A program to find duplicate makes its best effort to find same or similar ideographs even if they have different IDS. (See the document IRG N1154 for the current algorithm the program uses and how smartly it can recognize different IDSs represent same ideograph.) However, some simple guideline about the way of dividing will gain the accuracy.

2. Principles

The principles behind the guideline are summarized as follows:

2.1. Minimal division.

We should not divide too much. If we need further division, a program can easily generate such deep division forms, because we only use existing (already standardized) ideographs with their own IDS division. On the contrary, if we starts with the maximum division, its is not easy to algorithmically re-structuring the original shapes.

2.2. Concentration on visual shapes.

We should not stick to the ideographs meaning, origin, or the traditional classification/separation of components. Remember that our purpose of use of IDS is only to review the proposed ideographs. If we rely on, for example, the knowledge about the radical, IDS division by a person who doesn't know the correct radical may make a ***wrong*** IDS division.

By ignoring the detailed knowledge on ideograph's meaning, origin, etc., there are more chance that the IDS assigned by a person is same to those by another, regardless of the difference of knowledge on that particular ideograph.

2.3. Giving up early.

Some ideograph have a unique shape and/or structure and not easy to find an IDS. That's OK. Let them leave alone. We don't need a complete collection.

Again, we are just reviewing. We are not compiling a dictionary. As long as a number of such exceptional cases are relatively small, they have no repercussion with the entire review process.

2.4. Restricted use of ***surrounding*** and ***overlapping*** IDCs.

The use of surrounding or overlapping IDCs is sometimes ambiguous and may fail to detect the duplicate character algorithmically. This principle is to remove this difficulty.

2.5. Generousness on minor differences

Don't try to represent details of the shapes of an ideographs. Ignore minor differences. We have a set of unification rules and if the difference is important (for the unification rules), we can consider so through the eye-to-eye review after the IDS based matching. On the other hand, if the IDS is constructed under a draconian policy, two shapes to be unified may have a totally different IDS and we may fail to find them duplicate.

3. Definitions

IDC (Ideographic description character): One of 12 UCS characters whose code points are in range 2FF0 to 2FFB. See Annex F.3 of ISO/IEC 10646 for details.

CDC (Character description component): A UCS character that is included either in CJK UNIFIED IDEOGRAPHS, in CJK UNIFIED IDEOGRAPHS EXTENSION A, in

CJK UNIFIED IDEOGRAPHS EXTENSION B, in KANGXI RADICALS, in CJK RADICALS SUPPLEMENT, or in CJK COMPATIBILITY IDEOGRAPHS. In other words, CDC is a DC that consists of just one UCS character.

SDC (Sequence description component): An IDS that is used as a DC in other IDSs. In other words, SDC is a DC that consists of a sequence of an IDC and following DCs.

DC: either CDC or SDC.

4. The procedure for Constructing IDC

[1] See if the ideograph has a structure that two same components ***pinch*** another components. If so, take the division. i.e.,

[1-1] If the ideograph can be divided into three parts using 2FF2 (⿲),where the left-most and right-most components are same CDC, divide so. (The middle may be CDC or SDC in this case.)

Example:

嬲 → ⿲女男女 (rather than ⿰娚女)

弼 → ⿲弓百弓 (rather than ⿰弼弓)

[1-2] Otherwise, if an ideograph can be divided into three parts using 2FF3(⿳), where the top and bottom components are same CDC, divide so.(The middle DC may be CDC or SDC in this case.)

Example:

器 → ⿳口犬口 (rather than ⿳哭口)

[2] If the **[1]** above doesn't apply, see if the given ideograph is divided into two parts, and both parts are coded ideographs (CDCs). i.e.,

[2-1] If an ideograph can be divided into two parts using 2FF0 (⿰), where the both left and right components are (not necessarily same) CDCs, divide so.

Examples:

雖 → ⿰虽隹 (not ⿰唯虫)

[2-2] Otherwise, if an ideograph can be divided into two parts using 2FF1(⿱), where the both top and bottom components are (not necessarily same) CDCs, divide so.

Examples:

笈 → ⿱竹及

[2-3] Otherwise, if an ideograph can be divided into two parts using 2FF4(⿴), where the both outer and inner components are (not necessarily same) CDCs, divide so.

Examples:

園 → ⿴口袁

[2-4] Otherwise, if an ideograph can be divided into two parts using 2FF5(⿵), where the both outer and inner components are (not necessarily same) CDCs, divide so.

Examples:

間 → ⿵門日

[2-5] Otherwise, if an ideograph can be divided into two parts using 2FF6(⿶), where the both outer and inner components are (not necessarily same) CDCs, divide so.

Examples:

凶 → ⿶凵乂

[2-6] Otherwise, if an ideograph can be divided into two parts using 2FF7(⿷), where the both outer and inner components are (not necessarily same) CDCs, divide so.

Examples:

匣 → ⿷匚甲

[2-7] Otherwise, if an ideograph can be divided into two parts using 2FF8(⿸), where the both outer and inner components are (not necessarily same) CDCs, divide so.

Examples:

厘 → ⿸厂里

[2-8] Otherwise, if an ideograph can be divided into two parts using 2FF9(⿹), where the both outer and inner components are (not necessarily same) CDCs, divide so.

Examples:

勾 → ⿹勹厶

[2-9] Otherwise, if an ideograph can be divided into two parts using 2FFA (⿺), where the both outer and inner components are (not necessarily same) CDCs, divide so.

Examples:

赶 → ⿺走干

[2-10] Otherwise, if an ideograph can be divided into two parts using 2FFB(⿻), where the both outer and inner components are (not necessarily same) CDCs, divide so.

Examples:

幽 → ⿻山幺

Note the explicitly given priority of IDCs. If an ideograph can be divided into two parts either horizontally or vertically, we always divide it hirozontally (even if the division contradicts the ideographs origin!)

Examples:

灸 → ⿱从从 (rather than ⿰夂夂)

[3] If the [1] and [2] above still don't apply, see if the given ideograph is divided into three parts, and all parts are coded ideographs (CDCs), take it. i.e.,

[3-1] If the ideograph can be divided into three parts using 2FF2, where all left, middle, and right components are CDCs,divide so.

Examples:

徹 → ⿲彳育攵

[3-2] Otherwise, if an ideograph can be divided into three parts using 2FF3, where the both top and bottom components are CDCs, divide so.

Examples:

享 → ⿱亠口子

[4] If the [1], [2], and [3] don't apply, we try to divide the ideograph using two IDCs at the same time. During this rule [4], we assume an SDC is an IDS for a component of the ideograph under consideration that if the component was an ideograph and applying the rules [1] through [3] above for it would cause the IDS.

[4-1] If an ideograph can be divided into two parts using 2FF0 (⿰), where the left component is a CDC, and the right component is an SDC, divide so.

Examples:

幹 → ⿰卓𠂉人干

[4-2] Otherwise, if an ideograph can be divided into two parts using 2FF0 (⿰), where the right component is a CDC, and the left component is an SDC, divide so.

Examples:

穎 → ⿰禾𠂉匕禾頁

[4-3] Otherwise, if an ideograph can be divided into two parts using 2FF1(⿱), where the top component is a CDC, and the bottom component is an SDC, divide so.

Examples:

薛 → ⿱艹⿰𠂉自辛

[4-4] Otherwise, if an ideograph can be divided into two parts using 2FF1(⿱), where the bottom component is a CDC, and the top component is an SDC, divide so.

Examples:

憩 → ⿱舌⿰𠂉自心

[4-5] Otherwise, if an ideograph can be divided into two parts using 2FF4(⿴), where the outer component is a CDC, and the inner component is an SDC, divide so.

Examples:

圀 → ⿴囗⿱罒方

[4-6] Otherwise, if an ideograph can be divided into two parts using 2FF5(⿵), where the outer component is a CDC, and the inner component is an SDC, divide so.

Examples:

岡 → ⿵冂⿱罒山

[4-7] Otherwise, if an ideograph can be divided into two parts using 2FF6(⿶), where the outer component is a CDC, and the inner component is an SDC, divide so.

Examples:

㠯 → ⿴凵⿱人二

[4-8] Otherwise, if an ideograph can be divided into two parts using 2FF7(⿲), where the outer component is a CDC, and the inner component is an SDC, divide so.

Examples:

匰 → ⿲匚山王

[4-9] Otherwise, if an ideograph can be divided into two parts using 2FF8(⿳), where the outer component is a CDC, and the inner component is an SDC, divide so.

Examples:

厚 → ⿳厂日子

[4-10] Otherwise, if an ideograph can be divided into two parts using 2FF9 (⿴), where the outer component is a CDC, and the inner component is an SDC, divide so.

Examples:

貳→ ⿴弋二貝

[4-11] Otherwise, if an ideograph can be divided into two parts using 2FFA (⿵), where the outer component is a CDC, and the inner component is an SDC, divide so.

Examples:

邃 → ⿵辶穴豕

Note again on the priority. Also note that, except for the cases for 2FF0 and 2FF1, *we don't allow SDC as the first DC to the IDC.*

[5] If the [1], [2], [3] and [4] don't apply, we now try IDS with three IDCs. Just repeat the [4] with consideration that the SDCs explained in [4] can now be the IDS. (Exact conditions [5-1] to [5-11] are omitted, since they are exactly the same sentences as [4-1] to [4-11].)

[6] If the ideograph is still not divided into an IDS, give up.

Examples.

Examples:

勝 → ⿰月券 （not ⿰朕力）
桂 → ⿰木圭 (prefer to ⿰木⿱土土)
圭 → ⿱土丶 (not ⿻土丶)
圭 → ⿱土丶 (not ⿻土丶)
傘 → ⿱人⿻十灬
傾 → ⿰亻頃（prefer to ⿰化頁）
謄 → ⿰月誊 （not ⿰朕言）
京 → ⿳亠口小
雥 → ⿳隹隹隹
纞 → ⿳糸言糸 (prefer to ⿰糸誶)
厳 → ⿱𠂉⿰厂敢 (not ⿰⿱𠂉厂敢)
彦 → ⿱文⿰厂彡 (not⿰⿱文厂彡)
県 → ⿱⿺乚目小 (not ⿺ ⿱乚小目)
虎 → ⿱卜⿰厂斤 (not ⿰⿱卜厂斤)

4. Sample file.

The sample IDS data (ids.txt) attached with this document covers most of BMP and SIP characters. They might be useful on constituting the IDS of any character, as the most efficient way to constitute the IDS is to refer to the IDS of the similar character and copy (the part of) them.

If you can't find the appropriate DC of the target character, think of any other character which shares the common DC part, then search and see how that character is constituted in the sample file.

附件 3

IRG_N1158：本次會議決議

JTC1/SC2/WG2/IRG N1158
Date: 2005-12-2

ISO/IEC/JTC 1/SC 2/WG 2/IRG
Ideographic Rapporteur Group (IRG)

Resolutions of IRG Meeting #25

Source : IRG Meeting #25
Place: Berkeley,USA
Date: 2005-11-28 ~ 12-2

IRG Meeting #25, attended by delegates from China, Hong Kong SAR, Japan, Republic of Korea, Taipei Computer Association, Vietnam and Unicode (liaison member), has made the following resolutions:

Resolution IRG M25.1: Future meeting schedule

Unanimous

The IRG resolves to adopt the following meeting schedule:

IRG #26 Hue, Vietnam, 2006-6-5 ~ 9
IRG #27 Taipei, Taiwan (TCA), 2006-11-27 ~ 12-1 (date to be confirmed)
IRG #28 Hangzhou, China, May/June 2007 (place and time to be confirmed)
IRG #29 China /Macao SAR/ROK /Japan/Unicode will consider hosting

Resolution IRG M25.2: Ideographic Description Sequence (IDS) scheme (ref.: N1153, N1154, N1173, N1175)

Unanimous

The IRG accepts the documents from Japan (N1153, N1154) including the decomposition mechanism suggested. The IRG further decides that the IDS scheme shall be adopted for review the occurrences of duplicated ideographs in the CJK_C1 and future versions. The IRG agreed that the guidelines in N1153 is for reference only and not mandatory.

Resolution IRG M25.3: CJK Ext C1 Editorial Group Report (ref.: N1156, N1161, N1162, N1163, N1164, N1167, N1171, N1172, N1177, N1178 and N1179)

Unanimous

The IRG accepts the report from the IRG Chief Editor (N1163) and the report from CJK_C1 Editorial Group (N1167).

The IRG also accepts the document from Japan (N1161) and the recommendation from the CJK_C1 Editorial Group to implement suitable quality assurance mechanism for further improvement for the CJK_C1 and future work.

The IRG decides to adopt in accordance with WG2 N3002 and in addition to IDS mentioned in M25.2 above, more effective and comprehensive quality assurance measures. The suggested measures are in document N1178 and N1179.

Resolution IRG M25.4: CJK Ext C1 Work Schedule (ref.: N1179)

Unanimous

The IRG resolves to postpone the CJK_C1 deadline and start working on stable CJK_C1 characters first. The IRG accepts the proposal that the finalization date of the CJK_C1 is set at the IRG Meeting #27. A summary schedule for the remaining rounds of review is set below.

Completion Date	Action item
2006-01-06	IRG Chief Editor assigns a set of stable characters selected from CJK_C1_v52 to each member chief editor.
2006-02-10	Member chief editors submit IDS attributes of their set of stable characters to IRG Chief Editor
2006-03-03	IRG Chief Editor distributes CJK_C1_v60 with glyphs to member chief editors
2006-04-07	Member chief editors send comments on CJK_C1_v60 to IRG Chief Editor
2006-04-21	IRG Chief Editor distributes consolidated comments to member chief editors for confirmation
2006-04-28	Member chief editors send confirmation to IRG Chief Editor
2006-05-19	IRG Chief Editor distributes CJK_C1_v61 to member chief editors for comment
Before IRG Meeting #26	Member chief editors review the CJK_C1_v61
Before	Two more rounds of review and comment

Resolution IRG M25.5: Old Hanzi (pre-Qin) Encoding (ref.: N1165, N1166, N1168)

Unanimous

The IRG accepted the contributions from TCA, USA and China with old Hanzi samples and proposals (N1165, N1166 and N1168). The IRG also accepts the report from Old Hanzi Expert Group (N11xx) with the proposed work plan, the recommendation to focus on the repertoire development of Oracle Bone inscriptions first, the principles of old Hanzi selection.

Resolution IRG M25.6: Errata Report (ref.: WG2N3002 Annex I, N1164, N1172, N1178)

Unanimous

The IRG accepts the contributions from the Chief Editor (N1164). The IRG instructs the IRG Chief Editor to implement recording structure for defects in the published ISO 10646 releases reported by member bodies (N1178). Member bodies are requested to continue to submit errata data, if found, to the IRG Chief Editor for consolidation and discussion in future IRG meetings. In accordance with WG2N3002 Annex I, the results will be reported to WG2 upon IRG's instruction.

Resolution IRG M25.7: CJK Strokes (ref.: N1150, N1174, N1180, N1181)

Unanimous

The IRG accepts the contribution from Dr Richard COOK (N1150). The IRG further accepts the report from CJK Strokes Ad Hoc Group (N1174). The IRG resolves to submit the 'Proposed Addition to the CJK Strokes Block of the UCS' (N1180) and the summary form (N1181) to WG2 for acceptance and the possible inclusion in the Amendment 3.

Resolution IRG M25.8: Report to WG2 on improvement of IRG processes

Unanimous

The IRG instructs the IRG Rapporteur to report on improvement of IRG processes to the WG2.

Resolution IRG M25.8: Appreciation to the Host

By Acclamation

The IRG expresses its sincere appreciation to the host of IRG Meeting #25, the Unicode Consortium, the UC Berkeley Townsend Center for the Humanities, the Institute of East Asian Studies and the East Asian Library at UC Berkeley. The IRG also expresses its appreciation to the organizer, Department of Linguistics, UC Berkeley. The IRG would also like to thank Dr. Deborah Anderson, Dr. Richard Cook and Prof. Jim Matisoff for their excellent logistics, arrangements and hospitality.

The IRG also expresses its sincere appreciation to the Dr. Deborah Anderson and Mr Rick McGowan for their informative presentation on Standardization Process and Cuneiform Encoding Project.

The IRG also expresses its sincere appreciation to the Founder Group for their generous support in making available a complete Extension B font for the CJK_C1 editorial work.

附件 4

IRG_N1165：我國對古漢字平台及工作程序的建議

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1165

Date: 2005-11-29

Doc. Type:	Member body contribution
Title:	Summary of Old Hanzi Forum Held in Taipei 2005.10.26-28
Source:	TCA
Status:	Input to IRG
Action:	To be discuss
Distribution:	IRG Members and Ideographic Experts
Reference:	
No. of pages:	2
Medium:	Electronic

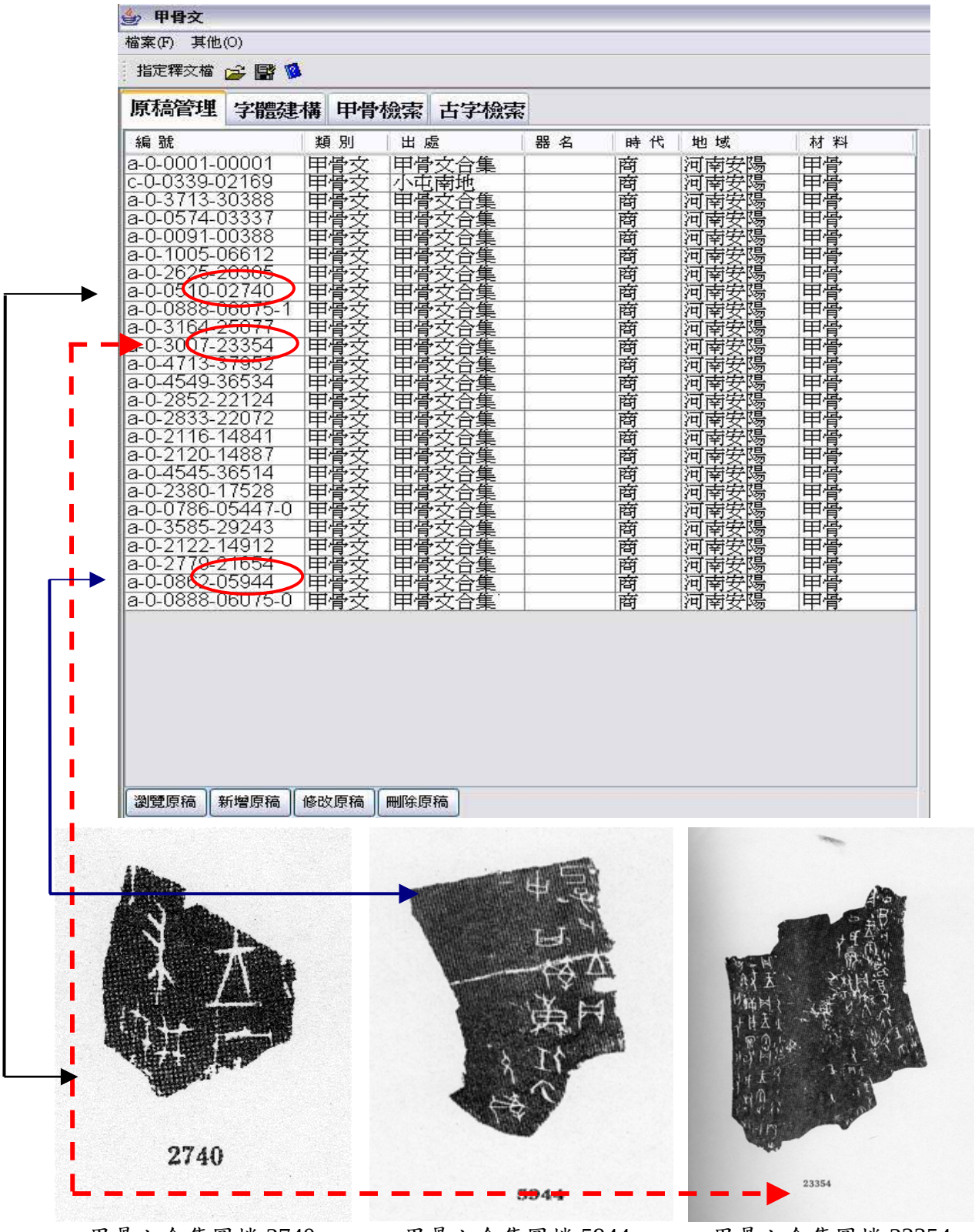
On October 26-28, TCA hosted an old Hanzi forum. Experts from Taiwan and mainland China made some conclusions in the meeting. All experts agreed that a common consensus should be obtained between Taiwan and Mainland China before processing the realignment work separately.

Based on the conclusion of the forum, TCA proposes the arrangement of Old Hanzi attributes plan as follow:

1. Start with building up (construction) an old Hanzi inscription database, including:
   (1) Original image (figure) database,
   (2) The database of glyph determination (隸定), and
   (3) Attribute database of Old Hanzi inscription.

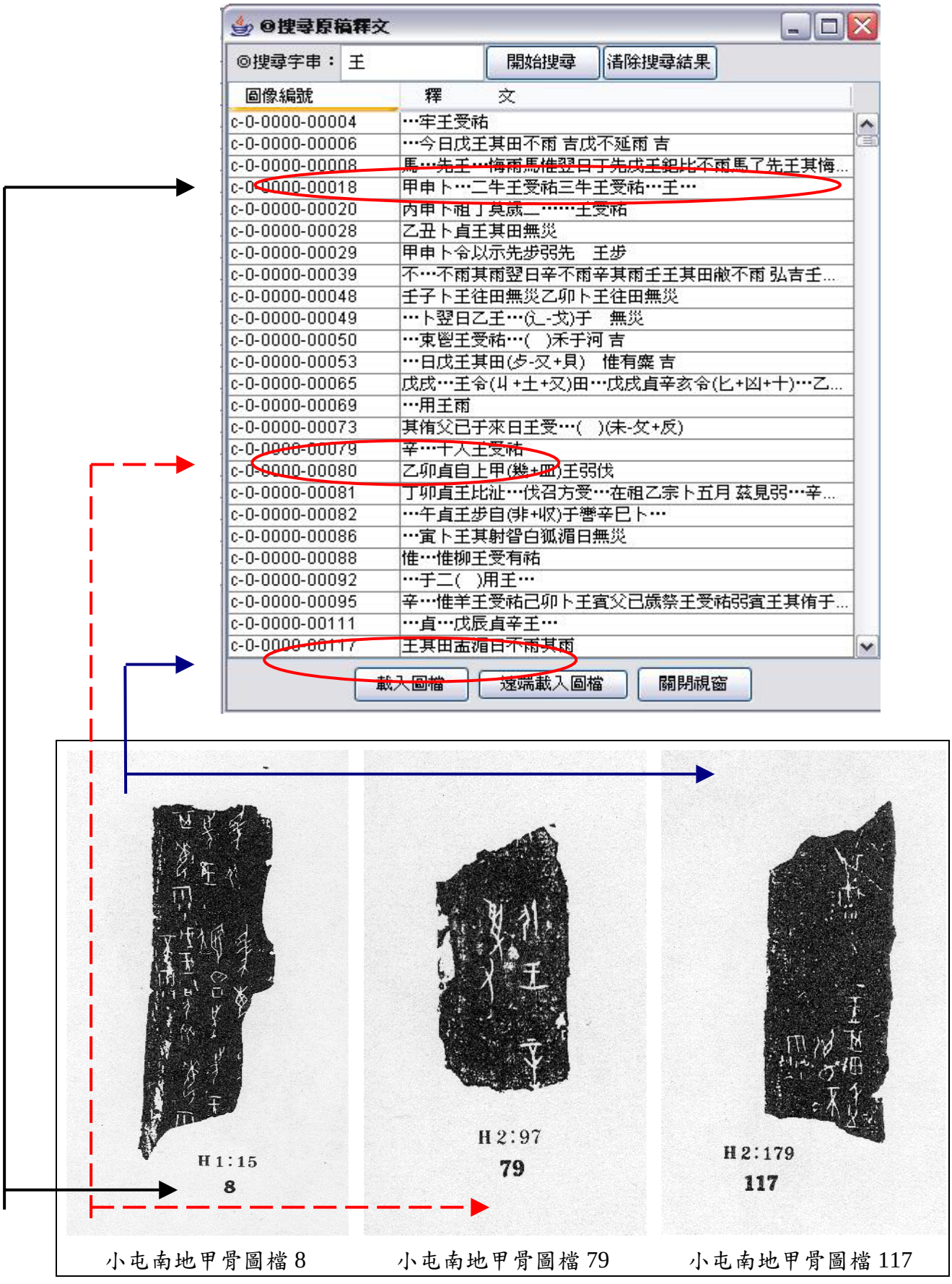
2. This Old Hanzi platform includes
   (1) The image (figure) database: Store original images of the Old Hanzi and the explanation archives. Users can refer to the original image when needed. (refer to Fig. 1 & Fig. 2)
   (2) The database of glyph determination: Store glyph determination and Corresponding Modern Character. (refer to Fig. 3)
   (3) Attributes Database: Stores the attribute data of collected Old Hanzi inscriptions. (refer to Fig. 4)

Fig. 1　the image database frame



甲骨文合集圖檔 2740　　甲骨文合集圖檔 5944　　甲骨文合集圖檔 23354

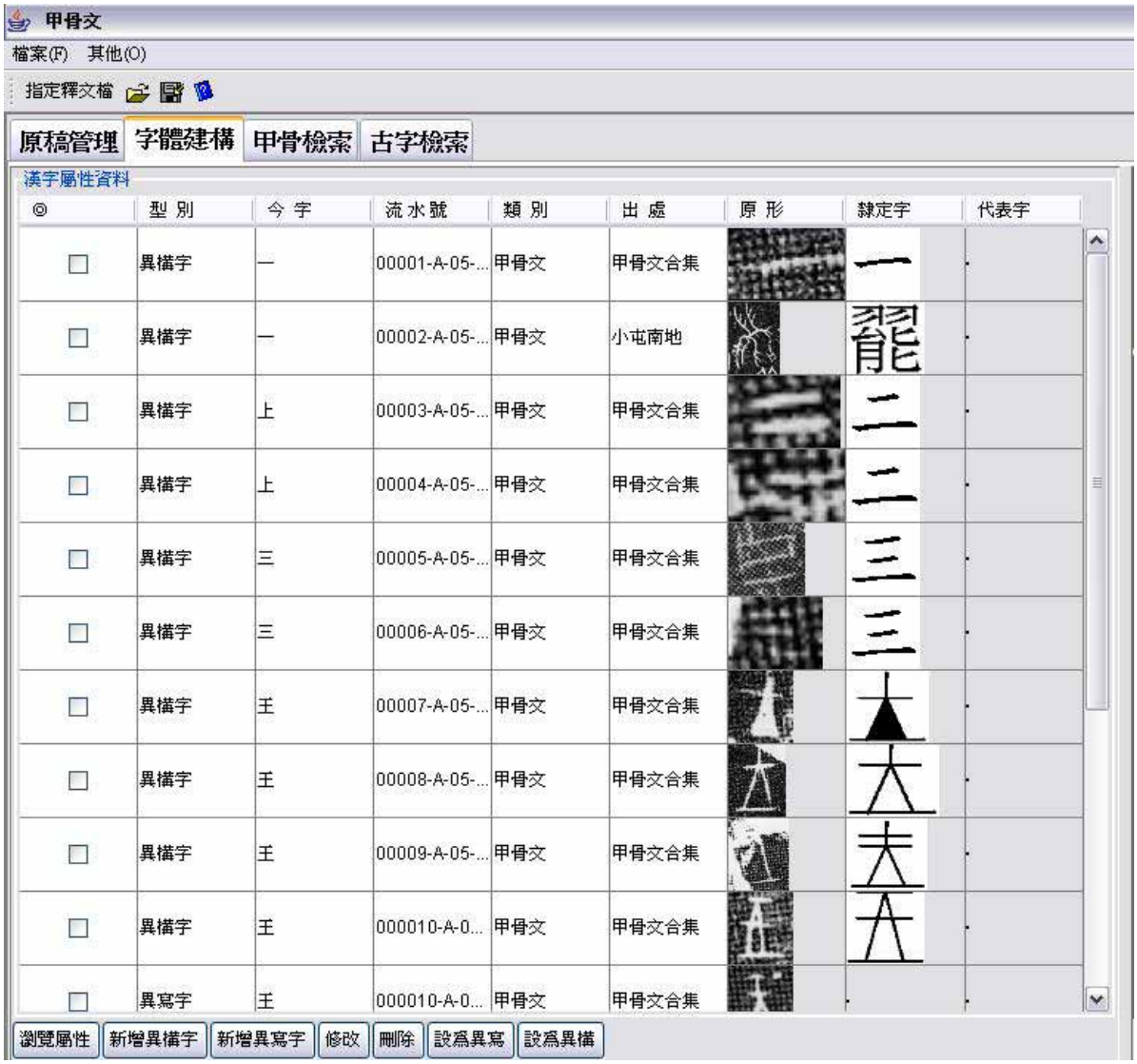
Fig. 2　the frame of using explanation archives search



小屯南地甲骨圖檔 8　　小屯南地甲骨圖檔 79　　小屯南地甲骨圖檔 117

Fig. 3　the database of glyph determination

Fig. 4　attributes database



3. The arrangement work procedure are:

(1) Build up image database: scan all oracle inscription images, and input all the explanation archives (translations) as text files, thus forming an image database.

(2) Choose glyphs: classify and choose words based on the constructed image database.

(3) Glyph Determination: arrange attributes of the chosen glyphs and precede glyph determination, data will be stored in database. The procedure of glyph determination are as follows:

a) Proceed glyph determination to the inscriptions of the Oracle Bones refer to the comments of experts

b) Compare and analysis the structure of glyphs with the same glyph determination result, judge between construct-variance and writing-variance according to the

conclusion made in the Old Hanzi Encoding and Modern Application Conference 2005.

c) choose the most representing (complete construct, clear strokes) construct-variance and writing variance for the glyph, and store it into the database.

(4) Counting representatives: there are many words for oracle bone inscription, every word has its own construct-variance and writing-variance, the representative is the word that represents the whole group of glyphs. Hence after the attributes arrangement, counting the number of representative is necessary.

(5) Edit a wordlist: edit an oracle bone inscription wordlist, for WG2 as reference for encoding

附件 5

IRG_N1167 及 N1179：CJK_C1 工作組報告

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1167

Date: 2005-11-29

Doc. Type:	Member body contribution
Title:	CJK_C1 report 1
Source:	CJK_C1 editorial group
Status:	Input to IRG
Action:	To be reported to the plenary
Distribution:	IRG Members and Ideographic Experts
Attached files:	59 Samples
No. of pages:	
Medium:	Electronic

The CJK_C1 editorial group met on Nov. 28 and 29, 2005, at IRG#25. Editors are from China, Japan, R.O.Korea, Vietnam, USA, Hong Kong SAR, TCA and Unicode Consortium.

The group agreed to postpone the CJK_C1 deadline again (the new deadline is not decided yet) and agreed to use IDS methods to improve the quality and to speed up the review.

The group reviewed IRG N1153 ***Guidelines on IDS Decomposition*** prepared by Japan and decomposed 59 characters selected from CJK_C1_v52 following the guidelines. The CJK C1 ad hoc group gained better understanding on how the guidelines work. Since these 59 characters are specially selected therefore some of the characters are not easy to be decomposed. Hands-on exercise was proposed for the editors to try decomposing 50 more characters selected on random basis, and to be finished by Nov.30 morning. The 50 characters are from CJK_C1_v52, they are:

00001
00401
00801
01201
… …
19601
(first one of every 400)

End of document.

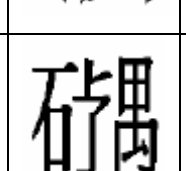
Attachment: **Samples from CJK_C1_v52 for IDS test**

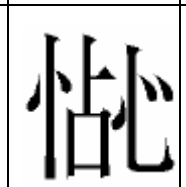
CJK_ C1 No.	Glyph	UCS 1	UCS 2	UCS 3	UCS 4	UCS 5	UCS 6	UCS 7
19487		L R	言	塑	謆		月肉	Unif
19488		LR	言	團				
22946		LR	馬	爨				

19503		LR	言	廉				
21670		LR	钅	氄				
07129		LR	女	⿸	厂	⿱	吅	敢
12620		⿰	水	⿳	亠	由	⿹	勹灬
22281		⿰	假	面				
12772		⿰	水	⿵	門	⿰	湇	

13258		⿰	火	⿳	巛	囟	？	
23703		⿺	孔	鳥				
15285		⿺	礼	多				
23282		⿰	魚	关				
24210		⿰	业	⺌	⿵	冂	巾	兮
21027								

23178				弓	□米	巳	弓	鬲
22618								
22675			食		业	八		
07034				臼	目	女		
22882				?				

21392			𠂉	𠂤		月	金	
22354				艹	本	衣	韋	
21419			金		叚	師		
20915			虎	虎	卩			
15142			卜	了				
01660			備		𠂉	丨		

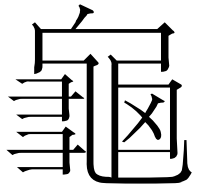
05345		□	吕	夤	吕			
20367		□	⿱	𠂤幺又				
09023		□						
20326		□	車	□	手	手		
22052				幾				
18630		⿱	艹	鬘				

05461	圖	□	?					
02061	與	?				彐		
02556	由	由						Unif
02567	甴	甴	一	申	凵	?		
02548	酉		一	酉	凶			
03141	匹	□	匚	凵				

03142	医	□	匚	矢				
03172	匾	□		□				
03663	爾	□	丿	爾				
03609	殿	□	厂	臼	丩	殳		
03577	扈	□	厂	乩				
00369	虒	□	厂	?	允			

03525	厖	□	厂	尨				
19848	貳	⿴	貳	丿				
23689	鳶	⿴	一	弋	鳥			
20628	逸	□	辶	□	免	厶		
22568	匈	□	⿴	几	二	幻		
23225	魚	□	曰	由	儿	魚		

00371	龜	⿱	𠂉	⿻	口	电		
24356	龜	?						
10168	𣏌	?						
00004	艹	?						
00005	工	⿱	丁	→				
00213	⊢	?						

00220		?						
18355		龕						
18547		□	艹					

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1179

Date: 2005-11-30

Doc. Type:	Member body contribution
Title:	CJK_C1 report 2
Source:	CJK_C1 editorial group
Status:	Input to IRG
Action:	To be reported to the plenary
Distribution:	IRG Members and Ideographic Experts
Attached files:	
No. of pages:	
Reference:	IRG N1171 and IRG N1172
Medium:	Electronic

The CJK_C1 editorial group met on Nov. 30, 2005, at IRG#25. Editors are from China, Japan, R.O.Korea, Vietnam, USA, Hong Kong SAR, TCA and Unicode Consortium.

The editors reviewed IRG N1171 ***Consideration on the Decision Recording for Ext. C1 Development*** and IRG N1172 ***Typical Rationale for Error Report against Published Standard*** and discussed on suggestions proposed by the 2 documents. It is principally accepted that these suggestions such as recording CJK_C1 discussion in detail in an updating memo and using more methods and rationale for discussion of CJK_C1, etc., should be followed in the future review of CJK_C1. The 2 documents may be updated in the future.

The editors agreed to postpone the finalization of CJK_C1 to IRG#27, and:

- With help of the IRG technical editor, the IRG chief editor should select some stable characters (tentative target is 5,000 characters) from CJK_C1_v52 under the minimum criteria: that the characters must have been reviewed for at least three times and never been questioned. Priority will be given to those characters undergone more reviews.
- These characters will be checked with IDS method and manual checking, editors are also encouraged to use other tools.
- Characters with no question after review will be submitted to WG2 as CJK_C1.
- Evidences of submitted characters required by WG2 (WG2 N3002) should be ready before finalizing CJK_C1. The criteria of evidences

in detail will be discussed at IRG#26.

The editors also agreed that characters in CJK_C1_v52 (including E, D and other sets of v52) excluded from the above process will be considered as a part of CJK_C2.

Time table of future review of CJK_C1:

- 2005-12-30: Mr. Kawabata Taichi distributes updated IDS guidelines to all member's chief editors and put them to IRG web site.
- 2006-01-06: IRG chief editor distributes characters selected from CJK_C1_v52, the task of making IDS attributes should be set to member's chief editors at the same time.
- 2006-02-10: IRG member's chief editors send IDS attributes to Mr. Kawabata Taichi for further process.
- 2006-02-17: Mr. Kawabata Taichi sends processing results to IRG chief editor and IRG technical editor.
- 2006-03-03: IRG chief editor distributes new CJK_C1 character list (CJK_C1_v60) with glyphs (generated by IRG technical editor) to IRG member's chief editors.

- 2006-04-07: IRG member's chief editors send comments to IRG chief editor.
- 2006-04-21: IRG chief editor distributes the consolidated comments to IRG member's chief editors for confirmation.
- 2006-04-24: WG2#48
- 2006-04-28: IRG member's chief editors send confirmation to IRG chief editor.
- 2006-05-12: IRG chief editor sends the fixed consolidated comments to IRG technical editor.
- 2006-05-19: IRG chief editor distributes CJK_C1_v61 to IRG member's chief editors.
- 2006-06-12: IRG#26

  Note: IRG editors should review CJK_C1_v61 although they are not required to report their review results before or at IRG#26, it is assumed that they should report their review results some days after IRG#26.

Tentative time table for CJK_C1 after IRG#26

- 2006-06-26, IRG member's chief editors send comments to IRG chief editor.
- 2 rounds of review of CJK_C1
- 2006-11-27: IRG#27

附件 6

IRG_N1172：CJK、CJK_A 及 CJK_B

字集錯誤的報告格式

JTC 1/SC2/WG2/IRG N 1172
Date: 2005-11-30

ISO/IEC JTC 1/SC2/WG2/
Ideographic Rapporteur Group

Source / Contribution Identifier: MS
Meeting: IRG#25 @ Berkeley, CA
Title: Typical Rationale for Error Report against Published Standard
Keywords: (none)
Status: Personal Contribution
Short Description: to be discussed on Editorial Group Meeting, IRG #25
Proposed Conclusion / Requested Action: to be discussed at IRG #25
Arguments / Text of Contribution: See attachments.

Typical Rationale for Error Report against Published Standard

2005-11-30
Masahiro Sekiguchi

When an editor submit an error report against ideographs that are already published in ISO/IEC 10646, the Rationale column in the submission form should show one from the list below.

1. Possible rationale for unification

If your report is of type "Two existing codepoints should have been unified", use rationale from the following list:

(a) Identical glyph shape.
Supply which glyph and which glyph are considered identical; if the said code point shows multiple glyphs, e.g. "Glyph of 2XXXX and K column of YYYY."

(b) Direct appliation of Annex S.1.5 example.
Supply which parts in the said ideographs are to be unified based on which example, e.g., "Upper right components of 2XXXX and J column of YYYY, based on the second example of Annex S.1.5 a)."

(c) Appliation of Annex S.1.5 rule with no applicable example.
Supply which parts in the said ideographs are to be unified based on which rule, and an explanation why you considered the difference in question falls into the said rule, e.g., "Lower right components of 2XXXX and 2YYYY, based on the Annex S.1.5 b), The difference here is just a same type of difference as in the sixth and seventh example to the rule."

(d) Regressive use of source separation example in first half of Annex S.3
Supply which example is regressively applicable to which parts of the glyphs, e.g., "Phonetics of 2XXXX and 2YYYY, the third example of S.3"

(e) Use of multi-column codechart in the standard:
Supply which columns of which codepoint is applied to which parts, e.g., "Upper components of 2XXXX and 2YYYY, J and K columns of ZZZZ."

(f) Referencing past decision
Supply which decision is applicable to which part. Unfortunately, we have no good way to refer to the past decision. IRG needs more discussion...

(g) Other reason:
Explain freely in detail.

2. Possible rationale for disunification

If your report is of type "Two glyphs allocated to a single codepoint should have been disunified", use rationale from the following list (with some additional information similar to the cases in 1.):

(a) Completely different glyph (A mistake...)

(b) Direct application of Annex S.1.4 example

(c) Application of Annex S.1.4 rule with no exactly applicable example

(d) Regressive use of non-source code separation example in the last half of Annex S.3

(e) Use of existing codepoints

NOTE: Since we have source code separation cases other than those listed in Annex S.3, or Extension B still has some error cases, use of exsiting codepoints requires detailed evaluation.

(f) Referencing past decision

(g) Other reason

(END OF DOCUMENT)

附件 7

IRG_N1173R：我國及 Unicode 協會

對 IRG-N1153 的意見

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1173R

Date: 2005-12-01

Doc. Type:	Member body contribution
Title:	Comments on IRG N1153
Source:	TCA and Unicode consortium
Status:	Input to IRG
Action:	To be discussed
Distribution:	IRG Members and Ideographic Experts
Reference:	
No. of pages:	
Medium:	Electronic

TCA and Unicode believe that the use of IDS may greatly help CJK standardization work. But we also believe that the method may cause some unexpected problems. The following are some of our concerns:

1. Is component position and ordering information essential? Is it really helpful to the process of identifying duplicates?

   (1). **High training cost**. To use this method to decompose characters, people need to spend much time learning and getting used to the rules. As is well known, there is variation in the writing order of characters and in identification of components, and it is not easy to force users of IDS to change their habits, especially since people will be working quickly, in order to accomplish the work in a reasonable amount of time.

   (2). **Human error**. Even following the guidelines, people may decompose the same character in different ways. As was seen in yesterday's review, even a well trained person can easily decompose a character incorrectly. More complicated rules mean more human effort and more human error.

2. The main purpose of character decomposition is to identify possible duplicate characters. Why then do we not we just decompose the characters in the most simple way possible? The following are some suggestions and comments on the document IRGN1153.

   (1) **DO NOT restrict decomposition order/direction**. Multiple decompositions might be permitted. For the purpose of character comparison, relative component position and order do not provide sufficient additional information, to compensate for the added complications which they introduce. Normally two different characters contain different

components. We can simply compare two characters without any radical position or order information. Even if the comparison system points out two different characters as possible duplicates, human verification is still required, and it is easier for the human eye to compare simplified descriptions than more complicated ones. The savings in time and effort would be significant.

(2) **DO NOT restrict the depth of component analysis**. For example, the character "彬"might be decomposed in any one of the three following ways: "林+彡", "木+杉" or "木+木+彡". If "彬" is decomposed as "林+彡" then the comparison system should automatically decompose "林+彡" into an undecomposable component level (in this case is "木+木+彡") prior to doing the comparison.

3. **Conclusion**:

The purpose of using character decomposition is to reduce the workload and also the human mistakes. But if the decomposition rules are overly complicated, the learning process will be too long, the results themselves will contain too many errors, and the purpose of using the system will be defeated.

Appendix.

The following examples show how easy it is to make "mistakes".

1-2  →□ 𠙵犬𠙵

WHY NOT?□ 哭𠙵

4-7]  →□ 凵□ 人二

WHY NOT?□ 凵 

4-7 凶→□ 凵乂

WHY NOT?□ 乂凵

4-9]厚→□ 厂□ 日子

WHY NOT?□ 厂孚

4-10]貳→□ 弋□ 二貝

WHY NOT?□ 弍貝

附件 8

IRG_N1174：筆畫工作組報告

JTC1/SC2/WG2/IRG N 1174
Date : 2005.12.1

ISO/IEC JTC1/SC2/WG2/IRG

Ideographic Rapporteur Group

(IRG)

In the IRG#25 meeting at Berkeley, the CJK Strokes ad-hoc group finalized development of the CJK Stroke repertoire augmenting the existing set of 16 strokes with 21 new strokes, for a total of 37 strokes in the block. The current document provides information on each of the proposed 18 strokes including unique IRG ID, proposed name, and examples of usage.

The group went over each pending issue summarized in the previous report (IRG N1137A) and addressed additional comments supplied by US.

References

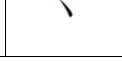
IRGN 1086A (“Stroke classification for Chinese characters”, TCA)
IRGN 987 (“CDL Specification”, Cook), See also IRGN 1096, IRGN 1097
IRGN 1081 (“146 ideographic components for ISO/IEC 10646”, China)
IRGN 1138 (“Suggestions of CJK Strokes for republic of Korea”, ROK)

The CJK Strokes ad-hoc group resolved to accept the following unifications:

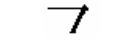
1. IRG ID2 (HT) is unified with IRG ID9 (TCA, U+31C0).

2		HT	地、坊、墻、釘、牡、翔、羚、玩、珪、提	TCA	

2. IRG ID7 (LD) is unified with IRG ID6 (D).

7		LD	TCA:奇、示、奏、父、米、奥、小、卜	TCA	

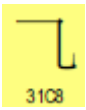
3. IRG ID12 (HXZ) is equivalent to either IRG ID10 (HZ) or IRG ID13(HP).

12		HXZ	今、敢	China	

4. IRG ID24 (PT) is unified with IRG ID23 (PZ)*.

24		PT	TCA:弘、么、公、翁 Cook:厶	TCA、Cook	(*)Name of IRG ID23 is changed from PH to PZ

5. IRG ID29 (HXG) is unified with IRG ID29A (HZWG, U+31C8)

29		HXG	TCA:飛、風、瘋、九、几、氣 China:虱	TCA、Cook China	
29A			Cook:飛、氣 Ad Hoc group：九、几、亢	Cook	Report to WG2 for change of glyph to that of IRG ID 29

6. IRG ID31 (SZXZ) is unified with IRG ID30 (SZZ).

31		SZXZ	TCA:吳 China:吳、吳 Cook:专、吳	TCA、Cook China	

7. IRG ID37 (WP) is unified with IRG ID4 (P) ~~5 (SP)~~.

37		WP	厂	Cook	

8. IRG ID38 (DN) is unified with IRG ID8 (N).

38		DN	入、仒	Cook	

9. IRG ID 39 (PN) is unified with IRG ID8 (N).

39		PN	是、走	Cook	

10. IRG ID40 (TPN) is unified with IRG ID8 (N).

40		TPN	之	Cook	

11. IRG ID 44, 45, 46 are added from ISO/IEC 10646:2003/Amd.1 to complete the stroke set.

44		HZT			(U+31CA)	
45		HZZP			(U+31CB)	
46		HPWG			(U+31CC)	

In addition, the group requests the IRG to make the following suggestions to WG2.

1. The current glyph for U+31C8 in the standard represents a very specific form out of unifiable stroke shapes.

   The group suggests changing the current glyph to  indicated as the representative shape of IRG ID29 (HXG) which is more generic for the stroke type of this kind.
2. The group also suggests adding an annotation "(hxg)" to the name "CJK STROKE HZWG" to indicate U+31C8 represents generic form of the HXG (Heng-Xie-Gou) type of CJK Stroke including HZWG (Heng-Zhe-Wan-Gou) type as a special case.

The group also resolved to instruct the editor of the Stroke ad-hoc group to generate the following two files for the IRG rapporteur to submit the proposal to WG2 after IRG review:

- IRG N1180 ("Proposed additions to the CJK Strokes block of the UCS")
- IRG N1181 ("Summary for Stroke submission")

Schedule:

24th of December 2005 Draft ready

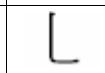
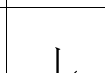
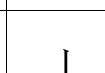
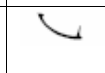
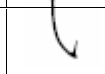
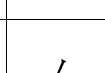
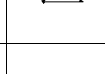
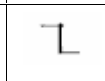
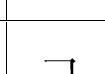
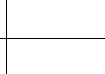
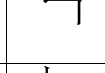
16th of January 2006 Feedback from member bodies

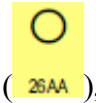
30th of January 2006 Revised submission copy is ready.

CJK Strokes review material (attachment of IRG N1174)

Final proposed CJK Stroke repertoire

IRG ID	Stroke	Proposed name	Usage examples	Submitters	Proposed UCS	mappings to existing UCS
1	一	H	TCA:一、三、丁、丞、丈、世、不、上 Cook: 十、卅、七	TCA、Cook		CJKU: U+4E00 (一)
3	丨	S	TCA: 丩、中、串、訌、乍 Cook:中、上、五、丑	TCA、Cook		CJKU: U+4E28 (丨)
4	丿	P	TCA:乂、爻、禾、毛、乏、乖、采、衣 Cook:八、行	TCA、Cook		CJKU: U+4E3F (丿)
5	丿	SP	TCA:乃、月、用、齊、几、~~六~~、人、班 Cook:大	TCA、Cook		CJKU: U+4E3F (丿)
6	丶	D	TCA:丸、义、永、冰、凡、丹、主、求 Cook:火、主	TCA、Cook		CJKU: U+4E36 (丶)
8	乀	N	TCA :大、人、天、入、走、邊、廷、尺、 Cook: 人	TCA、Cook	(U+31CF)	
9	乁	T	TCA:冰、淋、冶、冽、暴、氾、錄 Cook:地、虫	TCA、Cook	(U+31C0)	
10	𠃌	HZ	TCA:口、囗、田、品、呂、申、甲、圓 China:𠃌 Cook:口	TCA、China、Cook		CJKU: U+200CD (𠃌)
11	𠂉	HG	TCA:疋、了、予、矛、子、字、疏、~~今~~ Cook:、寫、𠂉	TCA、Cook		CJKU: U+4E5B (𠂉)
13	𠂆	HP	TCA:又、双、叒、今 Cook: 又、今	TCA、Cook	(U+31C7)	
14	𠃍	SZ	TCA:斷、繼 Cook:山、互	TCA、Cook		CJKU: U+200CA (𠃍) U+200CB

						(⺁)
15		SW	TCA:區、亡、妄 Cook:四	TCA、Cook	(U+31C4)	
16		SWZ	肅、嘯、蕭、簫	TCA	KX counting convention	CJKU: U+200CE (㇄)
17		ST	TCA:食、良、艮、很、狠、鄉 Cook:民	TCA、Cook		CJKU: U+2010C (㇀)
18		SG	爭、事、求、水	TCA		CJKU: U+4E85 (亅)
19		BXG	TCA:心、必、沁、忽、慈	TCA、Cook	(U+31C3)	
20		XG	TCA:戈、弋、戰、我 Cook: 我、弋	TCA、Cook	(U+31C2)	
21		WG	TCA:狐、嶽、貓、家、逐 Cook:家	TCA、Cook	(U+31C1)	
22		PD	TCA:巡、獵、災、甾 Cook:女、巛	TCA、Cook		CJKU: U+21FE8 (㇒)
23		PZ	互、彙、牙、弘、么、公、翁	TCA	Previously identified with PH	CJKU: U+200CB (⺁)
25		TN	TCA:⺍ Cook:八、入、曳	TCA、Cook		CJKU: U+4E40 (乀) U+4E41 (乁)
26		HZZ	TCA:卍 Cook: 凹	TCA、Cook	(U+31C5)	
27		HZW	TCA:殳、投 Cook: 朵	TCA、Cook	(U+31CD)	
28		HZG	TCA:羽、習、包、匀、匍、用、青、甫 Cook:月、勺	TCA、Cook	(U+31C6)	CJKU: U+200CC (㇌)
29 A		HZWG (HXG)	TCA:飛、風、瘋、九、几、氣 China:虱	TCA、Cook China	(U+31C8)	Rad supp: U+2E84 (乁)
30		SZZ	TCA:亞、鼎、丐 Cook:亞、龍	TCA、Cook		CJKU: U+200D1 (㇇)
32		SWG	TCA:亂、己、已、巳	TCA、Cook		CJKU:

			Cook:礼、心			U+4E5A (乚)
33		HZZZ	TCA:凸 Cook: 凸	TCA、Cook	(U+31CE)	
34	乙	HPWG	TCA:乙、氿 Cook:乞	TCA、Cook		CJKU: U+4E59 (乙)
35		SZWG	TCA:號、号、亏、弓、強、丐 Cook:马、亏	TCA、Cook	(U+31C9)	
36	乃	HZZZG	TCA:乃、孕、仍 Cook:乃、仍	TCA、Cook		CJKU: U+2010E (乃)
41		PG	乂 (J source character)	Cook		
42		Q	㘝 (K source character)	Cook K		Misc Sym: U+26AA  (26AA), U+26AC  (26AC)
43		YX	 (K source character)	K		

44		HZT			(U+31CA)	
45		HZZP			(U+31CB)	
46		HPWG			(U+31CC)	

End of document

附件 9

IRG_N1178：CJK、CJK_A 及 CJK_B

字集錯誤的填表規定

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N 1178

Date: 2005-11-30

Doc. Type:	Member body contribution
Title:	Recording of CJK(+A&B) Errors
Source:	CJK_C1 editorial group
Status:	Input to IRG
Action:	To be reported to the plenary
Distribution:	IRG Members and Ideographic Experts
Attached files:	1. Recording Format of CJK(+A&B) Errors 2. IRG N1172
No. of pages:	
Medium:	Electronic

The CJK_C1 editorial group met on Nov. 30, 2005, at IRG #25. Editors are from China, Japan, R.O.Korea, Vietnam, USA, Hong Kong SAR, TCA and Unicode Consortium.

The editors agreed to keep and update an internal list of errors found in CJK, CJK_A and CJK_B is necessary. Editors also agreed:

- The IRG internal CJK(+A&B) error list should contain not only the errors confirmed by editors but also the original error reports from the editors, as well as the detailed reasons for the results of discussion among the editors.
- The IRG internal CJK(+A&B) error list should be updated dynamically according to the newly submitted error reports and discussion results made by the editors.
- The IRG chief editor is responsible for maintaining the IRG internal CJK(+A&B) error list.
- A concise error list based on the IRG internal CJK(+A&B) error list should be generated and submitted to WG2 by the IRG chief editor when it is required by WG2 resolution or IRG resolution.

The editors reviewed and discussed a draft recording format for CJK(+A&B) errors (IRG N1164), and agreed to submit error reports and maintain the IRG internal CJK(+A&B) error list with the modified format. For detail of the modified format, read the attachment 1 please.

Attachment: Recording Format of CJK(+A&B) Errors

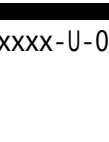
The recording format described in this document should be used for both of the IRG editors' CJK(+A&B) reports and The IRG internal CJK(+A&B) error list. It can be improved later if IRG editors require so.

For table use:

- When editors are submitting reports relating to characters duplicated or might be unified, Table 1 should be used.
- When editors are submitting reports relating to characters might be dis-unified, Table 2 should be used.
- When editors are submitting reports relating to glyphs with errors, Table 3 should be used.

Table 1: for characters duplicated or should be unified

Example:

Serial No.	UCS code point 1	Glyph 1 in ISO/IEC 10646: 2003	Source References	UCS code point 2	Glyph 2 in ISO/IEC 10646: 2003	Source References	Rationale provided by submitter	Conclusion (Y or N) drawn by all editors
xxxx-U-0001	219B4		G_HZ	219AD		T7-4D52	Because … …, see Annex S…. …, etc.	Yes. Reason: … …
xxxx-U-0002	27729		G_KX T7-3B5F	04666			Because … …, see Annex S…. …, etc.	Yes. Reason: … …
xxxx-U-0003								

Since some glyphs had been corrected in ***ISO/IEC 10646: 2003***, i.e. may be different to the original glyphs in ***SuperCJK 14.0***, glyphs in ***ISO/IEC 10646: 2003*** should be used in the table.

The UCS code point 1 should be larger than the UCS code point2. It means that the character in the right was encoded

before the left one, therefore, the left one should be unified to it.

The source references in this table are optional. They are used to make the IRG members of the sources pay more attention to their characters. They can be found in the CJKC_SR.txt and CJKU_SR.text in the CDROM of ***ISO/IEC 10646: 2003***.

Table 2: for characters should be dis-unified

Example:

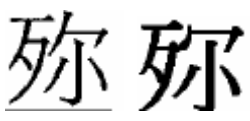
Serial no.	UCS code point	Glyph in ISO/IEC 10646: 2003	Glyphs should be dis-unified.	Glyphs are from	Rationale provided by submitter	Conclusion (Y or N) drawn by all editors
xxxx-D-0001	239E2	 239E2		Super CJK 14, 239E2	Because … …, see Annex S…. …, etc.	Yes. Reason: … …
xxxx-D-0002						

Table 3: for glyphs with errors

Example:

Serial no.	UCS code point	Current glyph in ISO/IEC 10646: 2003	Suggested glyph	Preceding Glyphs are from	Rationale provided by submitter	Conclusion (Y or N) drawn by all editors
xxxx-G-0001	25016			Super CJK 14, 25016		Yes. Reason: … …
xxxx-G-0002						

Note:

- The error report submitter should request for a document number at least 1 week before submission and make sure the document number is given by the IRG Rapporteur.
- “xxxx” in column “serial No.” means IRG Nxxxx of the original error report from the submitter.
- Every serial number in the above 3 tables should be unique, it should be kept with no change in the updating IRG internal CJK(+A&B) error list.
- The column “Conclusion (Y or N) drawn by all editors” should be kept empty when editors are submitting their error reports. Other columns in the table should be filled by editors.

- The column “Rationale provided by submitter” should be filled in detail as much as possible, such as “according to Annex S, compatibility ideographs”, etc., not “they should be unified… …” or something else only.
- The submitted error reports should be Microsoft Word documents.

Caution: The UCS code points in tables may not match glyph images by mistake. Edit and review error reports carefully.

附件 10

IRG_N1182：古漢字工作組報告

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 IRG N1182
Date: 2005-12-1

Title:	**Report from the Old Hanzi Interest Group**
Source:	**Old Hanzi Interest Group**
Status:	**Input to IRG**
Action:	
Distribution:	**IRG Members and Ideographic Experts**
Reference:	
No. of pages:	**6**
Medium:	**Electronic**

The Old Hanzi Interest Group has discussed the issues of the repertoire development of Old Hanzi. The group decided to obtain a common consensus between all the members before processing the realignment work separately, and to focus on the development of Oracle bone inscription repertoire first. The submission document will follow the N 1102 and N 1135, with one addition, the field "Unifiable Shapes" described below under "3. Format Update".

The Group discussed N1165 submitted by TCA regarding an Old Hanzi Interest Group forum on old Hanzi information format and working procedure.

The group came up with the following agreements:

1. The Old Hanzi experts group plans to arrange the Old Hanzi attributes as follows:

(1) Start with constructing an Old Hanzi script database, including these fields:

(1.1) Original image (figure)

(1.2) Glyph determination

(1.3) 9 Attributes of Old Hanzi inscription

(2) Edit a wordlist: edit an Oracle Bone inscription wordlist, for IRG consideration for submit to WG2 for information.

(3) Integrate all data into one consolidated database for easy use and make the database open to the IRG member bodies.

(4) Glyph determination: Establish a set of common rules for glyph determination in the next IRG meeting by consulting Old Hanzi experts after the IRG#25 meeting.

(5) The scope of mapping between Old Hanzi and modern ideographs：

(5.1) The mapping to modern ideographs is provided using encoded UCS characters.

(5.2) When there are mappings to both Simplified and Traditional characters, both will be listed.

2. The consolidated Old Hanzi database will include:

(1).The original images of the Old Hanzi and the explanation archives.

(2) The glyph determination and corresponding modern character.

(3) The attribute data of collected Old Hanzi inscriptions.

(Please refer to database illustrations in IRG N 1165 and N 1168.)

3. Format Update

ID	Rep. Script/ Glyph	Original Shape/ Glyph	Source	Period/ Epoch	Area/ site	Material	SW Radical	SW Radical Number	Glyph Determ.	Corresp. Modern Char	Unifiable Shapes	Notes
1												
2												

There are two changes of the format from what was defined by IRG N1135 as follows:

1. There is a new field, “Unifiable shapes”.

2. The field “Area/Terrain” is changed to “Area/site"

4. The work plan:

(1) Members are requested to submit the set of characters from Oracle Bone inscription categorized under the first 180 radicals of Shuowen Jiezi.

(2) Members should submit the images and attributes of the Oracle Bone inscriptions by 8th May, 2006.

(3) Data Format For Old Hanzi Data Exchange

For the data exchange and review work, members are going to use the data format specified as follows:

Images format:

3.1 use PNG storage format.

3.2 The original glyph should be scanned at 300 dpi (dots per inch).

3.3 The transcribed glyph images are named [ID]+[_R] (for example, if the ID is T00001, the transcribed glyph images should be named T00001_R).

3.4 The original glyphs are named [ID]+[_O] (for example, if the ID is T00001, the original glyph should be named T00001_O).

3.5 Glyph determination images are named [ID]+[_D] (for example, if the ID is T00001, the glyph determination image should be named T00001_D).

3.6 The images of unifiable shapes are named [ID]+[Unifiable shapes ID] (for example, if the Old Hanzi ID is T00001 and the unifiable shape ID is 000, the image of unifiable shape should be named T00001_000).

XML Schema:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified">
  <xs:element name="OldHanZi">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="Character" minOccurs="0"
maxOccurs="unbounded"/>
      </xs:sequence>
      <xs:attribute name="version" type="xs:string" use="required"
fixed="1.0"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="Character">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="Source"/>
        <xs:element ref="Period"/>
        <xs:element ref="Area"/>
        <xs:element ref="Material"/>
        <xs:element ref="Radical"/>
        <xs:element ref="ModernChar" minOccurs="0"/>
        <xs:element ref="Unified" minOccurs="0" maxOccurs="unbounded"/>
        <xs:element ref="Note" minOccurs="0"/>
      </xs:sequence>
      <xs:attribute name="id" use="required">
        <xs:simpleType>
          <xs:restriction base="xs:string">
            <xs:pattern value="(G|T|K|KP|J|V|S|H|M)[0-9]+"/>
          </xs:restriction>
        </xs:simpleType>
      </xs:attribute>
    </xs:complexType>
  </xs:element>
  <xs:element name="Source" type="xs:string"/>
  <xs:element name="Period" type="xs:string"/>
  <xs:element name="Area" type="xs:string"/>
  <xs:element name="Material" type="xs:string"/>
```

```
    <xs:element name="Radical">
      <xs:simpleType>
        <xs:restriction base="xs:unsignedShort">
          <xs:minInclusive value="1"/>
          <xs:maxInclusive value="540"/>
        </xs:restriction>
      </xs:simpleType>
    </xs:element>
    <xs:element name="ModernChar" type="xs:string"/>
    <xs:element name="Unified">
      <xs:complexType>
        <xs:attribute name="id" type="xs:string" use="required"/>
      </xs:complexType>
    </xs:element>
    <xs:element name="Note" type="xs:string"/>
  </xs:schema>
```

XML example:

```
  <?xml version="1.0" encoding="UTF-8" standalone="yes"?>
  <OldHanZi version="1.0">
    <Character id="T00001">
      <Source>甲骨文合集</Source>
      <Period>商</Period>
      <Area>河南安陽</Area>
      <Material>甲骨</Material>
      <Radical>001</Radical>
      <ModernChar>一</ModernChar>
      <Unified id="0000"/>
      <Unified id="0001"/>
      <Note/>
    </Character>
    <Character id="T00002">
      <Source>甲骨文合集</Source>
      <Period>商</Period>
      <Area>河南安陽</Area>
      <Material>甲骨</Material>
      <Radical> 005</Radical>
      <ModernChar>王</ModernChar>
      <Unified id="0000"/>
```

<Note/>
</Character>
</OldHanZi>

Members attended the meeting were Dai Hong,Li Guoying, Yin Jianghong, Zhang Deshao, Shi Jianqiao, Julie S. C. Chuang, Selena Wei, Hsu Hsueh Jen ,Chen Wen-Nan, ANAN Kanghong, Lu Qin, Tom Bishop, Cook Richard.

End of document