出國報告(出國類別：其他)

出席ISO/IEC JTC1/SC2/WG2
第56次會議

服務機關：經濟部標準檢驗局
姓名職稱：莊副局長素琴、陳技正星光
派赴國家：美國舊金山灣區 San Jose
出國期間：99 年 4 月 18 日至 25 日
報告日期：99 年 7 月 1 日

摘　　要

ISO/IEC JTC1/SC2/WG2工作組專責制定及維護國際標準ISO/IEC 10646〔廣用編碼字元集〕，用以解決各國間文字資訊交換、作業系統及應用軟體全球化等問題。

本次會議由Unicode協會主辦，共計34人出席，主要目的為審查ISO/IEC 10646第2版之最終委員會草案（FCD），並研討下一階段符號及文字編碼工作方向。ISO/IEC 10646第2版草案包含有CNS 11643﹝中文標準交換碼﹞第1至7字面之我國資訊用字。

依本次會議審查意見，WG2專案編輯預定於2010年9月前完成ISO/IEC 10646第2版之最終委員會草案（FCD）修訂，並提交ISO/IEC JTC1/SC2進行會員國最後審查及票決。ISO/IEC 10646第2版雖未正式經SC2會員國票決出版，惟WG2已決議展開第3版研擬工作，並預計於第57或第58次會議進行討論與審查工作，爰為順利推動我國古漢字及異體字等字集編入國際編碼標準，建請我國應及早準備相關資料，俾利爭取納入ISO/IEC 10646第3版。

ISO 10646系列標準是我國長期不斷參與制訂與維護的國際標準之一，後續WG2工作組將再持續進行相關表意文字及符號的擴編工作，我國應繼續參與該工作組活動，並與各國專家維持良好互動及合作模式，進而擴大我國中文編碼在該國際標準重要影響力。

目　　　　次

壹、會議概要及目的 ------------------------------------1

貳、會議過程 ---4

參、會議重要議題及決議 ------------------------------6

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

99年4月19日　23日。

二、會議地點

美國舊金山灣區San Jose，Adobe公司。

三、主辦單位

Unicode協會。

四、出席人員

本次會議共計34人出席會議，分屬台灣、中國、香港、日本、南韓、美國、加拿大、愛爾蘭及Unicode協會，與會人員名單如下：

姓名	會員體	服務單位
Mike KSAR	Convener, USA	Independent
Erich FICKLE	Invited Expert, China	Independent
Tom BISHOP	Invited Expert, Liaison SEI	Wenlin Institute, Inc.
Van ANDERSON	Invited Expert, Liaison SEI	Independent
LU Qin	IRG Rapporteur	Hong Kong Polytechnic University
Bear S. TSENG	TCA – Liaison Taiwan	Academia Sinica
Suh-Chyin CHUANG	TCA – Liaison Taiwan	Bureau of Standards, Metrology and Inspection, Taiwan
Hsin-Kuang CHEN	TCA – Liaison Taiwan	Bureau of Standards, Metrology and Inspection, Taiwan
Lin-Mei WEI	TCA – Liaison Taiwan	Chinese Foundation for Digitization Technology
Alain LABONTÉ	Canada; Editor 14651	Independent
V. S. (Uma) MAMAHESWARAN	Canada; Recording secretary	IBM Canada Limited
CHEN Zhuang	China	Chinese Electronics Standardization Institute
JING Yongshi	China	North University of Ethnics, Ning Xia
LI Guoying	China	Beijing Normal Universlity

姓　　名	會 員 體	服　務　單　位
SUN Bojun	China	Institute of Ethnology and nthropology, Chinese Academy of Scences
Michael EVERSON	Ireland; Contributing Editor	Evertype
Masahiro SEKIGUCHI	Japan	Fujitsu Limited
Yoshiki MIKAMI	Japan	Nagaoka University of Technology
KIM Kyongsok	Korea (Republic of)	Busan National University
Craig CUMMINGS	USA	Yahoo! Inc.
Eric MULLER	USA	Adobe Systems Inc.
Katsuhiko MOMOI	USA	Google
Lisa MOORE	USA	IBM Corp.
Mark DAVIS	USA	Google
Markus SCHERER	USA	Google
Peter EDBERG	USA	Apple Computer Inc.
Richard COOK	USA	Univ. of California, Berkeley
Rick McGOWAN	USA	Unicode Consortium
Roozbeh POURNADER	USA	High Tech Passport Ltd.
Yasuo KIDA	USA	Apple Inc.
Ken WHISTLER	USA; Contributing Editor	Sybase Inc.
Michel SUIGNARD	USA; Project Editor	Independent
Deborah ANDERSON	USA; SEI, UC Berkeley – Liaison	Dept. of Linguistics, Univ., of California, Berkeley
Peter CONSTABLE	USA; Unicode Consortium – Liaison	Microsoft Corporation

五、會議目的

為容納全球各種語言的字元和符號，ISO及IEC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），編號則訂為ISO/IEC 10646（簡稱ISO 10646）；其中「中日韓認同表意文字區」即為收容亞洲各國所使用之漢字。因漢字集的規模龐大，且為多個國家（地區）共同使用，WG2為此再設「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」分為80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，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「廣用多八位元編碼字元集（UCS）」計53,949個字。

本次會議之主要目的為審查ISO/IEC 10646第2版之最終委員會草案（FCD），並研討下一階段符號及文字編碼工作方向。

貳、會議過程

◆ 4月18日（星期日）：去程

◆ 4月19日（星期一）：

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，議程請詳見N3805 “Tentative Agenda – Meeting # 56”。

5、檢討上次會議應辦事項。

6、 主席報告ISO/IEC JTC1相關事務，包含最終委員會草案(FCD)轉換為國際標準草案(DIS)情形，以及JTC1對FDAM 7(Final Draft Amendment 7)票選結果。

7、主席報告ITTF (Information Technology Task Force)相關事務。

8、主席報告ISO/IEC JTC1/SC2相關事務，包含SC2工作計畫、提送ITTF事項、「FDAM 8」及「10646: 2010版」票選結果。

9、主席報告ISO/IEC JTC1/SC2/WG2相關事務，包含辦理表情符號(Emoji)、西夏文(Tangut)及女真文(Jurchen)專案會議可行性，「Snapshot of Pictorial」發展藍圖，日本公布「IVD」註冊機制，重新排序「yogh」提案，以及以英文拼寫字元名稱指引。

10、IRG報告工作情形，由主席陸勤博士說明，包含ISO/IEC JTC1/SC2/WG2/IRG第33會議成果，IRG相關活動報告，以及「WG2 P&P(Principles & Procedures)」本文修訂建議等。

◆ 4月20日（星期二）：

1、 依據ISO/IEC JTC1各會員國的表決意見，逐一討論ISO/IEC 10646: 2003修訂文件Amendment 8最後草案內容（詳見N3792）。

2、依據ISO/IEC JTC1各會員國的表決意見，逐一討論ISO/IEC 10646第2版最後草案內容（詳見N3793）。

◆ **4月21日（星期三）：**

討論各會員提案擬新增或修訂的字元或符號，包含

1、表情符號(Emoji)。（詳見N3829）

2、西夏文(Tangut)。（詳見N3833）

3、女真文(Jurchen)。（詳見N3817）

◆ **4月21日（星期三）晚上：大會晚宴。**

◆ **4月22日（星期四）：**

1、 續審ISO/IEC JTC1各會員國對ISO/IEC 10646: 2003修訂文件Amendment 8最後草案內容的表決意見，並確認修正內容。

2、續審ISO/IEC JTC1各會員國對ISO/IEC 10646第2版最後草案的表決意見，並確認修正內容。

3、討論並決定ISO/IEC 10646第2版的新出版時程。

◆ **4月23日（星期五）：**

1、起草會議決議。

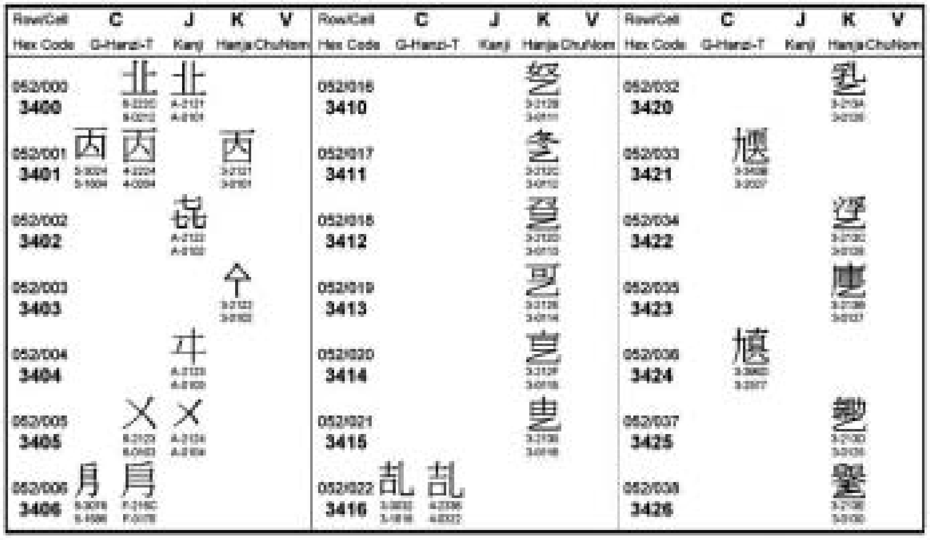
2、討論並通過會議決議。（詳見N3804）。

◆ **4月24、25日（星期六、日）：回程**

參、會議重要議題及決議

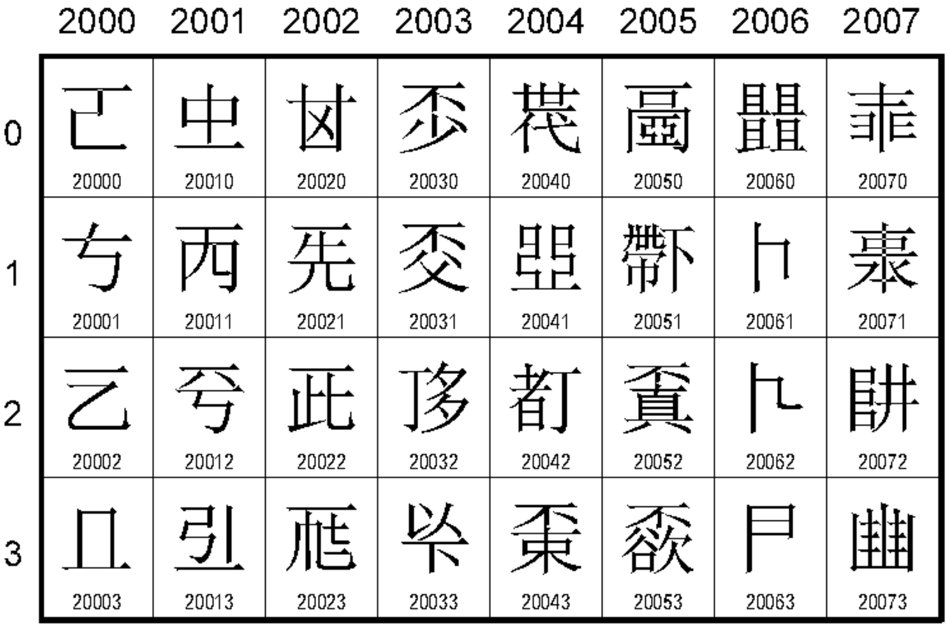
一、ISO/IEC 10646第2版之CJK-B多欄式編碼表

因IRG工作組無法在ISO/IEC 10646第2版定稿前完成CJK-B多欄式編碼表的審查工作（如圖1），WG2大會接受IRG的建議，將第2版中的CJK-B字集回復為第1版的單欄式編碼表（如圖2）。



Row/Cell Hex Code	C G-Hanzi-T	J Kanji	K Hanja	V ChuNom
052/000 3400	㐀	㐀		
052/001 3401	㐁		㐁	
052/002 3402		㐂		
052/003 3403			㐃	
052/004 3404		㐄		
052/005 3405	㐅	㐅		
052/006 3406	㐆			
052/016 3410			㐐	
052/017 3411			㐑	
052/018 3412			㐒	
052/019 3413			㐓	
052/020 3414			㐔	
052/021 3415			㐕	
052/022 3416	㐖			
052/032 3420			㐠	
052/033 3421	㐡			
052/034 3422			㐢	
052/035 3423			㐣	
052/036 3424	㐤			
052/037 3425			㐥	
052/038 3426			㐦	

圖1 多欄式編碼表



	2000	2001	2002	2003	2004	2005	2006	2007
0	𠀀 20000	𠀐 20010	𠀠 20020	𠀰 20030	𠁀 20040	𠁐 20050	𠁠 20060	𠁰 20070
1	𠀁 20001	𠀑 20011	𠀡 20021	𠀱 20031	𠁁 20041	𠁑 20051	𠁡 20061	𠁱 20071
2	𠀂 20002	𠀒 20012	𠀢 20022	𠀲 20032	𠁂 20042	𠁒 20052	𠁢 20062	𠁲 20072
3	𠀃 20003	𠀓 20013	𠀣 20023	𠀳 20033	𠁃 20043	𠁓 20053	𠁣 20063	𠁳 20073

圖2 單欄式編碼表

二、CJK字元認同規則

依據前次會議決議，IRG修訂CJK字元認同規則（詳如圖3），並獲WG2接受及決議收入下一版ISO/IEC 10646的附錄S。

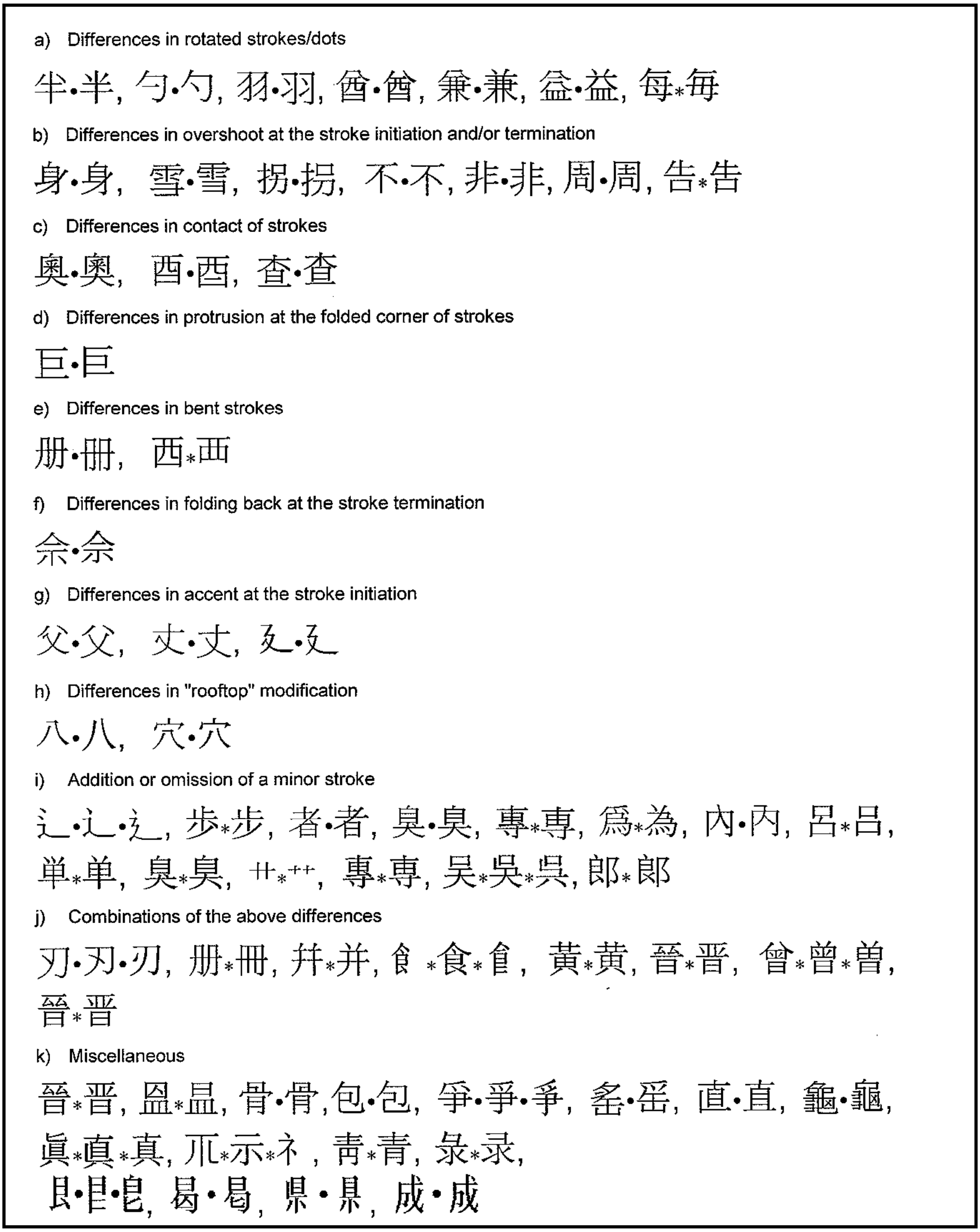

a) Differences in rotated strokes/dots
b) Differences in overshoot at the stroke initiation and/or termination
c) Differences in contact of strokes
d) Differences in protrusion at the folded corner of strokes
e) Differences in bent strokes
f) Differences in folding back at the stroke termination
g) Differences in accent at the stroke initiation
h) Differences in "rooftop" modification
i) Addition or omission of a minor stroke
j) Combinations of the above differences
k) Miscellaneous

圖3 CJK字元認同規則

三、IRG規章

WG2要求全體會員於IRG第34次會議召開前，檢視IRG規章（參見N3744），審查意見送IRG主席彙辦。

四、ISO/IEC 10646第3版之CJK-B及相容字集字型

WG2要求IRG各會員提交CJK-B及相容字集之字型檔給專案編輯，以便製作ISO/IEC 10646第3版所需之多欄式編碼表並送交IRG審查。

五、U+5FF9（忹）與 U+2F89F（忹）不予認同案

本次會議我國與中國大陸共同提案要求將U+5FF9（忹）與 U+2F89F（忹）二字分離，不應認同（詳見N3787），惟討論過程中發現U+225D6的字形與U+5FF9相同，由於本案較為複雜，WG2建請各會員體提供意見，並送IRG會議討論。

六、未來議程

1、WG2第57次會議：由南韓主辦。

地點：釜山。

日期：2010年10月4日　8日。

2、WG2第58次會議：由芬蘭主辦。

地點：Helsinki。

日期：2011年6月6日　10日。

3、WG2第59次會議：暫訂美國主辦。

地點：Mountain View, CA。

日期：2012年第1季。

肆、我國應配合辦理之工作

一、審查**IRG**規章：

WG2要求全體會員於IRG第34次會議召開前檢視IRG規章並提供審查意見。本案已請本局99年度「中文編碼及資訊處理標準之維護與推廣」委辦計畫之受託單位中文數位化技術推廣基金會配合檢視，並於規定時限內將審查意見送送IRG主席彙辦。

二、提送**CJK-B**及相容字集字型：

WG2要求IRG各會員提交CJK-B及相容字集之字型檔給專案編輯，以便製作ISO/IEC 10646第3版所需之多欄式編碼表並送交IRG審查。CJK-B及相容字集中屬我國用字部分字型，行政院主計處電子處理資料中心以委請中文數位化技術推廣基金會製作，並請專家審查，俾便提供WG2印製ISO 10646「資訊技術－廣用多八位元組編碼字元集(UCS)」碼本。

三、提供**U+5FF9**（忹）與 **U+2F89F**（忹）不予認同案資料

有關我國與中國大陸共同提案要求將U+5FF9（忹）與 U+2F89F（忹）二字分離不應認同案，因本案較為複雜，爰WG2建請各會員體提供意見，並送IRG會議討論。為便於各會員體審查，已請本局99年度「中文編碼及資訊處理標準之維護與推廣」委辦計畫之受託單位中文數位化技術推廣基金會準備更充備佐證資料，以利本案獲得WG2接受。

伍、心得與建議

一、依本次會議決議，WG2專案編輯預定於2010年9月前完成ISO/IEC 10646第2版之最終委員會草案（FCD）修訂，並提交ISO/IEC JTC1/SC2進行會員國最後審查及票決。ISO/IEC 10646第2版雖未正式經SC2會員國票決出版，惟WG2已決議展開第3版研擬工作，並預計於第57或第58次會議進行討論與審查工作，爰為順利推動我國古漢字及異體字等字集編入國際編碼標準，建請我國應及早準備相關資料，俾利爭取納入ISO/IEC 10646第3版。

二、本次會議，由於各會員體對所提案擴編文字或符號不熟悉，爰屢屢提出不同意見。針對不同意見，WG2主席會請提案會員充份說明，如有說明不清情形，WG2主席會請提案會員準備更充份資料並於下次會議再予討論。由此可見，出席WG2或IRG等工作會議並予以提案時，必須準備充份相當資料，以備當工作組討論時，可適時提出說明並輔以佐證資料。

三、本次會議於美國加州舊金山灣區Adobe公司舉行，Adobe視訊會議室有18支小型麥克風從天花板垂下，12個喇叭嵌入天花板，加上兩個可以一起或分別投影的銀幕，讓未能出席者透過該等視訊設備均能充分了解各國對議題討論情形及表達意見，而會議室中每個人的發言都可以被清晰的收音，同時清晰的收聽遠端的發言。比照國內的會議室，多半將麥克風和喇叭組固定安裝於會議桌上，此會造成非正前方的人發言不便，若能參考Adobe的設計，再運用我國豐沛的ICT技術能力及新穎遠距會議硬體設備，將可解決此問題。

四、ISO 10646標準可說是資訊產業重要的國際標準之一，並廣泛使用於Windows、Apple OS、Linux等作業系統，而我正體字又為該標準最重要內容。申請納入ISO 10646編碼字均須備齊佐證資料，並經多輪審查及充分討論，耗費許多時間及人力方可定案。後續WG2工作組將再持續進行相關表意文字及符號的擴編工作，我國應繼續參與該工作組活動，並與各國專家維持良好互動及合作模式，進而擴大我國中文編碼在該國際標準重要影響力。

陸、附件

1、會議照片集錦

2、N3805：Tentative Agenda － Meeting # 56

3、N3792：Draft disposition of comments on SC2 N 4123 (FPDAM text for Amendment 8 to ISO/IEC 10646:2003)

4、N3793：Draft disposition for fcd10646

5、N3804：Resolutions Meeting 56

6、N3744：IRG P&P (IRGN 1646)

7、N3787：Request for disunifying U+2F89F from U+5FF9

8、N3829：Emoji ad hoc report

9、N3833：Ad hoc report on Tangut

10、N3817：Comments on Jurchen

附件 1

會議照片集錦

圖 1：會議召開地點－Adobe 公司

圖 2：WG2 會議室

圖 3：會議場景之 1

圖 4：會議場景之 2

圖 5：會議場景之 3

附件 2

N3805：Tentative Agenda – Meeting # 56

DATE: 2010-04-10

ISO/IEC JTC 1/SC 2/WG 2

Universal Multiple-Octet Coded Character Set (UCS) - ISO/IEC 10646

Secretariat: ANSI

DOC TYPE:	Calling Notice Meeting 56
TITLE:	2nd Call Meeting # 56 from 2010-04-19/23 hosted by the US National Body at Adobe campus in San Jose, CA
SOURCE:	Mike Ksar, Convener
PROJECT:	JTC 1.02.18 – ISO/IEC 10646
STATUS:	
ACTION ID:	ACT – Submit all pending documents by 2010-04-12 to ensure inclusion in the draft tentative agenda. *Notify* convener the experts that plan to come to the meeting from national bodies, liaison organizations or ones that have been invited and approved by the convener. Logistics details are included in this document. ***Meeting Contact host at Adobe: Eric Muller***
DUE DATE:	2010-04-12
DISTRIBUTION:	SC2/WG2 members and Liaison organizations
MEDIUM:	Electronic
NO. OF PAGES:	6

This is the 2nd call for WG2 meeting # 56 from 19-23 April 2010 in San Jose, CA, U.S.A. It is hosted by the US National Body at the Adobe campus in San Jose, CA. A tentative draft agenda (N3805-A) is included in this document and will have a direct link in the document register.

WG2 will be meeting starting at 10:00 am Monday, 19 April 2010 and conclude by noon time Friday 23 April 2010. Note: Internet access via wireless LAN is available in the meeting room area.

The objective of this meeting is to continue the WG2 program of work, with focus on Amendment 8 as well as FCD for the next edition of ISO/IEC 10646. Contributions, including ones that have been carried forward from earlier meetings will be reviewed as well.

The detailed logistics and host contact information are given below:

Eric Muller
Adobe Systems
345 Park Avenue
San Jose CA 95110

USA
[1] 408 536 4085
emuller@adobe.com

VISAS

ANSI, as the US hosting body, will provide a letter of invitation for those delegates requiring one. ANSI's ISO Team (ISOT) will provide a letter upon receipt of the following information:

- Designation of ISO or ISO/IEC JTC 1 meeting: ISO/JTC1/SC2/WG2 meeting #
- Name (include male/female)
- Organization
- Email
- Date of birth
- Passport Number

Requests should be sent to ANSI at **isot@ansi.org** by the ISO Member of the delegate requiring a letter of invitation (i.e. SAC for China, DSM for Malaysia).

MEETING PLACE

The meeting is in the Albertus Conference Room, in the West Tower. The conference room phone number is (408) 536 3041. The street address is:

Adobe Systems
345 Park Avenue
San Jose CA 95110
USA

See map on the last page of this document for directions. The San Jose International Airport is about 15 minutes away, cab rides are about $20 each way.

HOTEL INFORMATION

All these hotels are within walking distance from the meeting venue.

*Ask for the Adobe corporate rate.   See list of hotels below.

Mike Ksar
ISO/IEC/JTC 1/SC 2/WG 2 Convener
22680 Alcalde Rd
Cupertino, CA 95014
U. S. A.

Phone: **+1 408 255-1217**

e-mail: **mikeksar@10646.com**

ISO/IEC JTC 1/SC 2/WG 2 *N3805-A*

Tentative Agenda – Meeting # 56

2010-04-10

Topic (Document No.)	Proposed Outcome
1. Opening and roll call	
2. Approval of the agenda (N3805)	*Approved agenda*
3. Approval of minutes of meeting 55 (N3703)	*Approved Minutes*
4. Review action items from previous meeting (N3703-AI)	*Updated Action Item List*
5. JTC1 and ITTF matters:	FYI
5.1. Transition from FCD to DIS (N3808)	
5.2. FDAM7 JTC1 ballot results	
6. SC2 matters:	FYI
6.1. SC2 Program of Work	FYI
6.2. Submittals to ITTF	FYI
6.3. Ballot results	FYI
6.3.1. FPDAM8 (N3790)	
6.3.2. 10646: 2010 edition	
6.4. Liaison from JTC1/SC34 on FPDAM 8 (N3809)	
6.5.	
7. WG2 matters:	
7.1. Possible ad hoc meetings: Emoji, Tangut, Jurchen	
7.2. Snapshot of Pictorial view of Roadmaps (N3807)	
7.3. Announcement of Japan's IVD Registration (N3796)	
7.4. Proposal to re-order yogh in the UCA and ISO 14651 (N3782)	
7.5. Guidelines on English spellings in character names (N3815)	
8. IRG status and reports	
8.1. IRG working document series (N3746)	
8.2. IRG Meeting 33 Resolutions (N3741)	
8.3. Summary of IRG activities (N3742)	
8.4. Suggested Text Revision of WG2 P&P (N3743)	
8.5. IRG P&P (N3744)	
8.6. Text version of Annex S examples (N3745, N3794)	
8.7. Suggested solution to M55.19 for incorrectly mapped compatibility characters (N3747)	
8.8. Request for disunifying U+2F89F from U+5FF9 (N3787)	
9. Script contributions related to ballots:	

9.1. Willcom input on Emoji (N3783)
9.2. Proposed additions to 10646:2003 (N3810)
9.3. Updated proposal to change some glyphs & names of emoticons (N3778)
9.4. Rational for proposal of N3778 (N3806)
9.5. KDDI input on Emoji (N3777)
9.6. DoCoMo input on Emoji (N3776)
9.7. Proposal to encode an emoticon “Neutral Face” (N3769)
9.8. Problems concerning U+1F471 “WESTERN PERSON” (N3785)
9.9. Proposal on use of ZERO WIDTH JOINER between two regional indicator symbols (N3779)
9.10. Comments on spelling SULPHUR vs SULFUR (N3813)
9.11. On the proposed A+A78F LATIN LETTER MIDDLE DOT (N3812)

10. Script contributions not related to ballots

10.1. Proposal to encode Arabic sign SAMVAT (N3734)
10.2. Proposal to encode six punctuation characters (N3740)
10.3. Proposal to encode the Landa script (N3768)
10.4. Proposal to encode the Sindhi script (N3767)
10.5. A roadmap for scripts of the Landa Family (N3766)
10.6. Towards an encoding for the Maithili script (N3765)
10.7. Revised proposal to encode mathematical diagonals (N3763)
10.8. Preliminary proposal for encoding the Kpelle script (N3762)
10.9. Preliminary proposal for encoding the Bassa Vah script (N3760)
10.10. Proposal for extending the annotation of the character “Y WITH DIAERESIS” (N3759)
10.11. Proposal to encode the Takri script (N3758)
10.12. Preliminary proposal for encoding the Mende script (N3757)
10.13. Preliminary proposal for encoding the Loma script (N3756)
10.14. Preliminary proposal for encoding the Palmyrene script (N3749)
10.15. Proposal to encode nine Cyrillic characters for Slavonic (N3748)
10.16. Proposal to encode an Abbreviation sign for Gujarati (N3764)
10.17. Proposal to encode the Typikon Symbols (N3772)
10.18. Proposal to encode an Armenian Dram currency symbol (N3771)
10.19. Proposal to encode two dashes required by the Chicago Manual of Style (N3770)
10.20. Proposal to encode the Old North Arabian script (N3773)
10.21. Proposal for encoding the Linear A script (N3755, N3774)
10.22. Proposal for encoding Georgian and Nuskhuri letters (N3775)
10.23. Revised code chart and namelist for the Modi script (N3780)
10.24. Pau Cin Hau script
    10.24.1. Preliminary proposal to encode the Pau Cin Hau script (N3781)

10.24.2. Properties of tone marks (N3784)

10.25. Proposal for Coptic numbers (N3786)

10.26. Final proposal for encoding the Miao script (N3761, N3789)

10.27. Revised proposal to encode the Jurchen characters (N3788)

10.28. Arabic characters for Koranic use

10.28.1. Proposal to encode four combining Arabic characters (N3791)

10.28.2. Proposal to change some combining characters

10.29. Proposal for Arabic Mathematical Alphabetic Symbols (N3799)

10.30. Proposal for Gangga Malayu script (N3798)

10.31. Final Proposal for encoding Tangut (N3797, N3797-A, N3797-B)

10.32. Preliminary proposal to encode the Tolong Siki script (N3811)

10.33. Proposed additions to 10646:2003 (N3810)

10.34.

11. Working draft for future amendment

11.1. Summary of repertoire proposal for future amendment (N3735, N3795)

11.2.

12. Contributions carried forward from earlier meetings

12.1. Hungarian (N3693)

12.2. Proposal for encoding generic punctuation used with Hungarian Runic (N3664)

12.3. Proposal for encoding generic punctuation (N3670)

12.4. Proposal for encoding the Hungarian Runic script (N3697)

12.5. Revised proposal for encoding the Manichaean script (N3644)

12.6. Proposal to encode te Pahawh Hmong script (N3667)

12.7. Proposal for encoding the Obsolete Simplified Chinese characters

12.8. Proposal for encoding the hentaigana characters (N3698)

12.9. Nushu (N3598, N3705, N3719)

12.10. Meetei Mayek extensions (N3478)

12.11.

13. Ballots disposition of comments

13.1. FPDAM 8 (N3792)

13.2. FCD for next edition (N3793)

14. Architecture issues

15. Publication issues

16. Defect reports

17. Liaison reports

17.1. Unicode Consortium

17.2. IETF

17.3. SC22

17.4. W3C

17.5. SEI (N3814)

18. Other business

18.1. Web Site Review

18.2. Future Meetings

18.2.1. Meeting 57 - Fall 2010, Korea (Republic of) (pending confirmation of location)

18.2.2. Meeting 58 - Spring 2011, Finland (pending confirmation) (along with SC2 plenary)

18.2.3. Meeting 59 - Fall 2011, Mountain View, CA, USA (pending confirmation), Germany (as backup)

18.2.4. Meeting 60 – Spring 2012 – Looking for host

19. Closing

19.1. Approval of Resolutions of Meeting 56

19.2. Adjournment

■ Logistics for meetings

SC2/WG2 #56 and SC2/OWG-SORT #10

San Jose CA, USA, April 19-23, 2010

◆ HOST CONTACT

Eric Muller
Adobe Systems
345 Park Avenue
San Jose CA 95110
USA
[1] 408 536 4085
emuller@adobe.com

◆ MEETING PLACE

The meeting is in the Albertus Conference Room, in the West Tower. The conference room phone number is (408) 536 3041. The street address is:

Adobe Systems
345 Park Avenue
San Jose CA 95110
USA

See the attached map for directions. The San Jose International Airport is about 15 minutes away, cab rides are about $20 each way.

◆ VISAS

ANSI, as the US hosting body, will provide a letter of invitation for those delegates requiring one. ANSI's ISO Team (ISOT) will provide a letter upon receipt of the following information:

- Designation of ISO or ISO/IEC JTC 1 meeting: ISO/JTC1/SC2/WG2 meeting #
- Name (include male/female)
- Organization
- Email
- Date of birth
- Passport Number

Requests should be sent to ANSI at **isot@ansi.org** by the ISO Member of the delegate requiring a letter of invitation (i.e. SAC for China, DSM for Malaysia).

◆ HOTEL INFORMATION

All these hotels are within walking distance from the meeting venue.
Ask for the Adobe corporate rate.

The Fairmont Hotel*

170 South Market St.
San Jose, CA
408-998-1900
800-527-4727
www.fairmont.com/sanjose

The DeAnza Hotel*

233 West Santa Clara St.
San Jose, CA
800-843-3700
www.hoteldeanza.com

The Hilton and Towers*

Almaden Blvd.
San Jose, CA
408-287-2100
800-HILTONS
www.hilton.com

Hotel Montgomery*

211 South First St.
San Jose, CA
408-282-8800
www.hotelmontgomerysj.com

San Jose Marriott*

301 South Market St.
San Jose, CA
408-280-1300
408-280-0212
www.sanjosemarriott.com

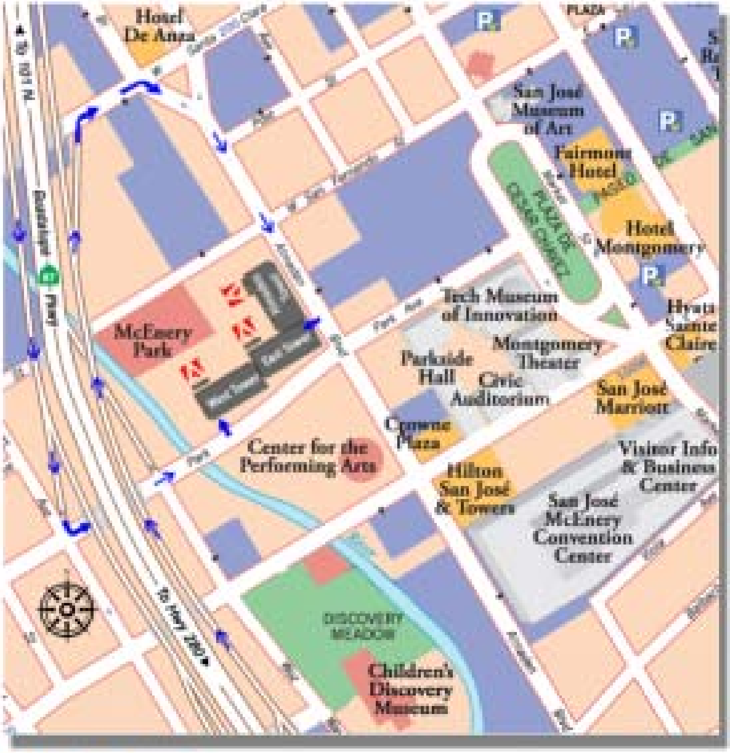
The Sainte Claire, A Larkspur Hotel

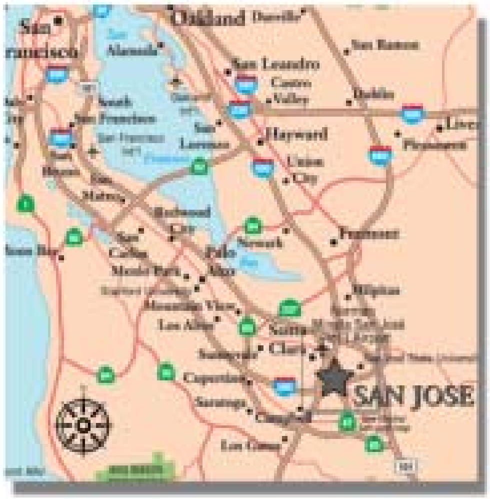
302 South Market Street
San Jose, CA 95113
Main: 408-295-2000
www.TheSainteClaire.com

Adobe

Directions to Adobe's Headquarters

San Jose, California


Hotel De Anza
Santa Clara
To 101 N.
Plaza
P
San José Museum of Art
Fairmont Hotel
Paseo
Hotel Montgomery
Plaza de Cesar Chavez
W. San Fernando St.
Guadalupe Pkwy
Almaden Blvd
Park Ave
Tech Museum of Innovation
Hyatt Sainte Claire
McEnery Park
Almaden Tower
East Tower
West Tower
Parkside Hall
Montgomery Theater
Civic Auditorium
San José Marriott
Crowne Plaza
Center for the Performing Arts
Hilton San José & Towers
Visitor Info & Business Center
San José McEnery Convention Center
To Hwy 280
Discovery Meadow
Children's Discovery Museum


San Francisco
Oakland
Alameda
San Leandro
Castro Valley
Dublin
San Ramon
Hayward
Union City
Pleasanton
Livermore
South San Francisco
San Bruno
San Mateo
Half Moon Bay
Redwood City
San Carlos
Menlo Park
Palo Alto
Newark
Fremont
Mountain View
Los Altos
Santa Clara
Milpitas
Sunnyvale
Cupertino
Saratoga
Campbell
Los Gatos
SAN JOSE

Visitors: Parking is free at Adobe for all visitors. Once you check in with security at the garage kiosk, you will be directed to park in any open visitor space or any open space on the lower levels.
The parking garages for all three buildings are connected on the ground floor.

BUILDING ADDRESS

Almaden Tower	151 Almaden Blvd.
East Tower	321 Park Ave.
West Tower	345 Park Ave.

ADOBE OPERATOR (408) 536-6000
Mailing Address 345 Park Ave., San Jose Ca 95110

---

280 Southbound from San Francisco
Take Guadalupe Pkwy/87 exit North
Continue using 87 Northbound directions

680 Southbound from Sacramento/Concord
680 South until it turns into 280 North
Take Guadalupe Pkwy/87 exit North
Continue using 87 Northbound directions

17/880 Northbound from Santa Cruz
17/880 North to 280 South (San Jose exit)
Continue using 87 Northbound directions

101 Northbound from Gilroy/Los Angeles
101 North to 280 North
Take Guadalupe Pkwy/87 exit North
Continue using 87 Northbound directions

87 Northbound
Take Santa Clara St./Downtown exit
Turn right on Santa Clara St.
Turn right on Almaden Blvd.

To access 321 Park Ave. and 151 Almaden Blvd -
Make a right turn into the Adobe garage,
which is half a block past San Fernando.

To access 345 Park Ave. - Turn right on Park Ave.
Make a right turn into the Adobe garage.

101 Southbound (from San Francisco & SFO airport)
Take 101 South towards San Jose
Take Guadalupe Pkwy/87 exit South
Continue using 87 Southbound directions

87 Southbound (from San Jose Airport)
Take Guadalupe Pkwy/87 exit South (downtown)
Take Park Ave./San Carlos St. exit, turn left

To access 345 Park Ave. - Make U-turn at Almaden Blvd.
Make a right turn into the Adobe garage.

To access 321 Park Ave. and 151 Almaden Blvd from Park Ave. -
Make left turn onto Almaden Blvd.
Make a U-turn on San Fernando
Make a right turn into the Adobe garage.

附件 3

N3792：Draft disposition of comments on SC2 N 4123 (FPDAM text for Amendment 8 to ISO/IEC 10646:2003)

ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)

Doc. Type: **Draft disposition of comments**

Title: Draft disposition of comments on SC2 N 4123 (FPDAM text for Amendment 8 to ISO/IEC 10646:2003)

Source: **Michel Suignard (project editor)**
Project: JTC1 02.10646.00.08
Status: For review by WG2
Date: 2010-04-02
Distribution: WG2
Reference: SC2 N4078, 4087, WG2 N3776, N3777, N3778, N3779
Medium: Paper, PDF file

Comments were received from Germany, India, Indonesia, Ireland, Japan, United Kingdom, and USA. The following document is the draft disposition of those comments. The disposition is organized per country.

Note – The full content of the ballot comments have been included in this document to facilitate the reading. The dispositions are inserted in between these comments and are marked in **Underlined Bold Serif text**, *with explanatory text in italicized serif.*

Germany: Negative

Germany votes "Disapproval with comments".
The vote is turned into "Approval" if the request in comment (3, [*T2 as edited*]) is accepted..)

Technical comments:

T1. Name of block 1F300-1F5FF "Miscellaneous Pictographic Symbols"

Germany suggests the block name to be changed into:
"Miscellaneous Symbols and Pictographs".
Rationale:
In fact, not all symbols proposed into this block are pictographic. This is taken into account by the proposed new name, which is more generic while retaining the original intent.
Acknowledgement: The name change originally was proposed by Asmus Freytag on the Unicode mailing list 2010-02-10, text in brackets added: "... to suggest some additional consistency:
Miscellaneous Symbols [2600-26FF; name as it is now; block already full]
Miscellaneous Symbols and Arrows [2B00-2BFF; name as it is now]
Miscellaneous Symbols and Pictographs [1F300-1F5FF; name change suggested]
This would mark all three blocks as containing some mixture of symbols, but making clear that for 2B00 there's an emphasis on arrows (and by implication on mathematical symbols) while for the new block the emphasis is on pictographs. Yet all three blocks can and do contain miscellaneous symbols."
Propose acceptance
See also comment T2 from Ireland

T2. Regarding the name part "WESTERN" in U+1F471 WESTERN PERSON

Germany requests that the term WESTERN is in no case used for denoting a special physical appearance of humans.
Germany suggests to replace "WESTERN" by "... WITH BLOND HAIR" (as in N3607).
Rationale
Any link from a specific physical appearance to membership of a cultural area can be considered racist (e.g. persons of African origin which feel themselves integrated into Western culture may legitimately think so). Especially, the link of "blonde hair" to "Western culture" may invoke associations to Nazi ideology.
WG2 discussion
Note that the comment T11 from Germany for PDAM8 included such a request and was rejected by the Emoji ad hoc group at meeting M55 (Tokyo). Rationale provided:
"In reference to German comment T11 on PDAM8, Germany requested a name change for 1F46F WESTERN PERSON along with the addition of a distinct character, creating a disunification based on gender (MAN vs. WOMAN). The Ad-hoc determined that the proposed dis-unification based on gender was inappropriate, not correctly reflecting the source Emoji character. Therefore, it is recommended that no changes be made to the character name and that the additional character not be added."
However the former request (T11) has now been split between T2 (name change) and T3 (dis-unification).

T3. Regarding the "PERSON" aspect in U+1F471

Germany requests that one of the following modifications is applied, strongly preferring the "Solution A":
Solution A: The character is replaced by two characters, as it was shown in N3607 at 1F46F/1F470:
1Fxxx MAN WITH BLOND HAIR
1Fyyy WOMAN WITH BLOND HAIR
Solution B: The character is renamed to "COMPATIBILITY SYMBOL PERSON WITH BLOND HAIR". It is to be listed under an appropriate header like "Emoji compatibility character" or "Compatibility Character for Japanese Industry Standard xxx".
An informative note like the following is to be added:

- the source standard allows a man's head as well as a woman's head being depicted

Additionally, it is suggested to move the character to U+1F5FA (near the "Cultural Symbols").
Rationale

U+1F471 "... PERSON" was included in FPDAM8 to be mapped to Emoji e-01A4, which in turn results from a unification in the Japanese Emoji set from two different symbols, KDDI #705 which shows a blond-haired woman's head, and SoftBank #290 showing a blond-haired man's head.
This unification is regarding being an error: it cannot be envisaged that a user exchanging this symbol between KDDI and SoftBank does not care whether his input is displayed as man's or woman's head.
In the same way, having such a character in Unicode is useless and fails to be a definite character includable into Unicode otherwise.

Anyway, the Emoji list is no more than an informal agreement between three Japanese companies (at least no claim is known to the German NB that it is a national or industry standard.)
Accordingly, documents like N3728 "Emoji sources" do not refer to the Emoji list. Rather, that document refers to the source standards of the single companies, and thus it can be adjusted with no problems to two different Unicode characters associated with the two different source symbols.

However, if SC2/WG2 has strong reasons to implement the Emoji list strictly 1:1 (even if the references to it are to be eventually replaced by references to their sources, as in N3728), and thus carving the error of an informal industry agreement into stone forever, it must be made clear that the "PERSON WITH BLOND HAIR" is no ordinary Unicode character which can be used without special precautions. Then, Solution B has to apply. Then, placing the character near the "Cultural Symbols" also emphasizes the fact that it is only accepted for special reasons.
See also the more detailed discussion in N3785.
WG2 discussion
See disposition of comment T2.

Editorial comment:

E1. Regarding the header "UPA letters" for U+A790/A791 in the "Latin Extended-D" block

This header should be something like "Additions for Janalif".
Rationale:
See N3581.
WG2 discussion
This could also be accommodated by annotations. Are those letters only used for Janalif?

India: Positive with comments

Technical comments

T1. Addition of characters in the ARABIC Block

1. The character 065F (ARABIC WAVY HAMZA BELOW) was proposed by Indian National Body for representation of Kashmiri language in Perso-Arabic script. We agree for its encoding.
Noted

2. The character at Code Point 0620 (ARABIC LETTER KASHMIRI YEH) was not proposed by Indian National Body. Annotation “Used in Kashmiri” for the code point 06CC (ARABIC LETTER FARSI YEH) was proposed. We have a view that there is need to encode one more character “ARABIC LETTER PALATALIZED YEH”. Hence the decision to encode 0620 (ARABIC LETTER KASHMIRI YEH) may be reviewed.”
Propose Noted
The proposed addition of 0620 ARABIC LETTER KASHMIRI removes the need to annotate 06CC. Even if there is a need to encode another character as suggested by India, this should not delay the encoding 0620. India needs to submit an encoding proposal form for ARABIC LETTER PALATALIZED YEH if such encoding is desired.

T2. Addition of characters in the DEVANAGARI Block

1. The following ten characters were proposed by Indian National Body for representation of Kashmiri language in Devanagari. These are as per our requirement. Hence we agree for their encoding.

093A (DEVANAGARI VOWEL SIGN OE)
093B (DEVANAGARI VOWEL SIGN OOE)
094F (DEVANAGARI VOWEL SIGN AW)
0956 (DEVANAGARI VOWEL SIGN UE)
0957 (DEVANAGARI VOWEL SIGN UUE)
0973 (DEVANAGARI LETTER OE)
0974 (DEVANAGARI LETTER OOE)
0975 (DEVANAGARI LETTER AW)
0976 (DEVANAGARI LETTER UE)
0977 (DEVANAGARI LETTER UUE)
Noted

T3. Addition of Characters in the MALAYALAM Block

MALAYALAM LETTER DOT REPH is proposed for encoding at 0D4E. This character has historic use. We agree for its encoding.
Noted
See also comment E1 from US.

Editorial comment

E1. Addition of Characters in the MALAYALAM Block

The following two characters, encoded in the ISO/IEC 10646:2003/Amd.8:2010(E), also have only "historic use". Hence it is proposed to add annotation "Character has historic use only" to these two characters.
a. 0D29 (MALAYALAM LETTER NNNA)
b. 0D3A (MALAYALAM LETTER TTTA)
Propose acceptance
These would be annotations on characters not part of this ballot. Assuming the annotation reflects usage, these additions are acceptable.

Indonesia: Positive with comments

Technical comments

T1 Batak

Referring to :

1. the guideline of Practical Batak Toba writing by Ama ni Par do muan
2. Article written by Uli Kozok : Batak Language Script and Literature
3. Surat Batak version 1.2, true type font for Microsoft and Macintosh for 5 type letter of Batak provided by Uli Kozok & Leander Seige

Batak letter is divided into 5 types:

A. Southern Group comprises to:

1. Batak letter Toba
2. Batak letter Simalungun
3. Batak letter Mandailing

B. Northern Group comprises to:

4. Batak letter Karo
5. Batak letter Pakpak

Proposed change only for LETTERS at page 31 (not include Sign, Dependent vowel signs, dependent consonant signs, Signs and punctuation)

WG2 discussion

It is not 100% clear what is requested here. Except for the two symbol characters in 1BFA and 1BFB, all characters are already encoded, therefore their names cannot be changed.

However a preamble could be added in front of the Letter section to explain the naming convention for these letters, something like (information extracted from N3320R):

```
@+ Unless explicitly encoded separately, Batak letters are shared among the five Batak
alphabets divided between Northern (Karo and Pakpak and Southern (Mandailing, , Simalungun,
and Toba). When letters have different usage among these alphabets, annotations are provided.

@            Letters
1BC0   BATAK LETTER A
       ∉  Letter a or ha  for Karo and Pakpak
1BC1   BATAK LETTER SIMALUNGUN A
1BC2   BATAK LETTER HA
       ∉  Toba letter ha or ka
       ∉  Also known as letter ka for Karo and Pakpak
1BC3   BATAK LETTER SIMALUNGUN HA
       ∉  Also known as Simalungun letter ka
1BC4   BATAK LETTER MANDAILING HA
1BC5   BATAK LETTER BA
1BC6   BATAK LETTER KARO BA
1BC7   BATAK LETTER PA
1BC8   BATAK LETTER SIMALUNGUN PA
1BC9   BATAK LETTER NA
1BCA   BATAK LETTER MANDAILING NA
1BCB   BATAK LETTER WA
       ∉  Toba letter wa can be represented by either 1BCB or 1BCD
1BCC   BATAK LETTER SIMALUNGUN WA
1BCD   BATAK LETTER PAKPAK WA
1BCE   BATAK LETTER GA
1BCF   BATAK LETTER SIMALUNGUN GA
1BD0   BATAK LETTER JA
1BD1   BATAK LETTER DA
1BD2   BATAK LETTER RA
1BD3   BATAK LETTER SIMALUNGUN RA
1BD4   BATAK LETTER MA
1BD5   BATAK LETTER SIMALUNGUN MA
1BD6   BATAK LETTER SOUTHERN TA
       ∉  Toba letter ta can be represented by either 1BD6 or 1DB7
1BD7   BATAK LETTER NORTHERN TA
1BD8   BATAK LETTER SA
       ∉  Pakpak letter sa or ca
1BD9   BATAK LETTER SIMALUNGUN SA
1BDA   BATAK LETTER MANDAILING SA
1BDB   BATAK LETTER YA
1BDC   BATAK LETTER SIMALUNGUN YA
1BDD   BATAK LETTER NGA
1BDE   BATAK LETTER LA
1BDF   BATAK LETTER SIMALUNGUN LA
1BE0   BATAK LETTER NYA
       ∉  Karo letter ca can be represented by either 1BE0 or 1BE1
1BE1   BATAK LETTER CA
1BE2   BATAK LETTER NDA
1BE3   BATAK LETTER MBA
1BE4   BATAK LETTER I
1BE5   BATAK LETTER U
```

Ireland: Negative

Ireland **disapproves** the draft with the technical and editorial comments given below.
Acceptance of these comments and appropriate changes to the text will change our vote to approval.

Technical comments

T1. Page Page 59, Row A72: Latin Extended-D.

Ireland reiterates its support for the character being balloted at A78F, LATIN LETTER MIDDLE DOT. Ireland opposes the removal of A78F LATIN LETTER MIDDLE DOT from FPDAM 8. However, in order to prevent confusion, we suggest that the name be changed to LATIN LETTER GLOTTAL DOT, which reflects its use as a phonetic letter in transcriptions of Tangut and Chinese..

WG2 discussion

See also comment GB1 from UK and T1 from US.
This has been a controversial addition proposal with opposing views from US versus Ireland and UK in PDAM 8 and again in this FPDAM8 ballot. Because, once a character is encoded it cannot be removed, the prudent approach seems to postpone the encoding. There was no harm in keeping the character in the amendment until the final technical phase, thus giving all parties more time to refine their arguments.
The current rationale for encoding which includes a usage for a script not yet encoded (Tangut) or rare (transcriptions of Chinese) seems a good indication to postpone the encoding proposal for a future amendment.

T2. Page 81, Row 1F30: Miscellaneous Pictographic Symbols.

Ireland requests that the block name be changed to Miscellaneous Symbols and Pictographs

Propose acceptance

See also comment T1 from Germany.

T3. Page 83, Row 1F30: Miscellaneous Pictographic Symbols.

With reference to N3777, "KDDI Input on Emoji", Ireland requests that the following characters be inserted into the block (pushing the rest of the characters in the column down by two:
1F536 MEDIUM RED CIRCLE
1F537 MEDIUM BLUE CIRCLE.

Propose non acceptance

The KDDI characters which have color indications are currently mapped to two 'colorless' existing characters: 26AA MEDIUM WHITE CIRCLE and 26AB MEDIUM BLACK CIRCLE. Unless there is a need to distinguish more than two circles, there is no need to add these two characters. 'Black' and 'White' in character names do not imply a specific color, they mean filled versus outline or filled with a darker color versus filled with a lighter color.

T4. Page 92, Row 1F60: Emoticons.

With reference to L2/10-061R "Emoji: Review of FPDAM8", and to earlier documents illustrating the glyphs and names used in the sources, Ireland requests a number of name changes, shown below in the formula OLD NAME > NEW NAME. The annotations to those characters are also given here; in a few cases where the annotation has changed or been added this is indicated. Naming conventions have been normalized for eye-shapes such as SQUINTING and CLOSED, and WINKING and TIGHTLY-CLOSED, reflecting some typical choices in the sources and in other emoticon implementations. The word HAPPY has not been used as in favour of the descriptive SMILING already in use in the UCS. A few of the faces have been given names descriptive of their "emotional comment", but in cases where the interpretation seems unclear a more direct description of the glyph has been used; in these cases, we believe, annotations should be used if necessary. The FACE WITH STUCK-OUT TONGUE for instance, may indicate "kidding" in Japanese telephony, but this gesture means something quite different in Europe and North America. (We have tried to be inclusive and neutral.) In the chart at the end of the document we have given the balloted name as a character annotation to assist WG2 and the UTC in identifying the characters, but it is not proposed that these names be retained as aliases.

WG2 discussion

Proposed dispositions have been incorporated reflecting the large consensus between the feedback from US and Japan (as reflected in N3778). However an ad hoc discussion is probably required to address differences between that consensus and the Irish comments.

a) 1F605 EXASPERATED FACE > AGITATED FACE
= e-320
Propose non acceptance
Propose instead FACE WITH COLD SWEAT per US comment T4c and Japanese comment JP10 (N3778).

b) 1F606 EXPRESSIONLESS FACE > UNAMUSED FACE
= e-326
Propose acceptance
Also supported by US comment T4b and Japanese comment JP10 (N3778).

c) 1F607 FACE WITH HEART-SHAPED EYES > SMILING FACE WITH HEART-SHAPED EYES
= e-327
Propose non acceptance or acceptance in principle
Propose instead HAPPY FACE WITH HEART SHAPED EYES per US comment T4b and Japanese comment JP10 (N3778).

d) 1F608 FACE WITH LOOK OF TRIUMPH > GRINNING FACE WITH LOOK OF TRIUMPH
= e-327
Propose non acceptance
Assuming Japanese comment JP10 (N3778) on this character is accepted, which requests a glyph change, this request is no longer applicable.

e) 1F609 WINKING FACE WITH STUCK OUT TONGUE > FACE WITH STUCK OUT TONGUE AND WINKING EYE
= e-329
* kidding, not serious (new annotation)
1F60A FACE WITH STUCK OUT TONGUE > FACE WITH STUCK OUT TONGUE AND TIGHTLY-CLOSED EYES
= e-32A
* kidding, not serious (new annotation)
Propose non acceptance or acceptance in principle
Propose instead 1F609 KIDDING AND WINKING FACE WITH STUCK OUT TONGUE and 1F60A KIDDING FACE WITH STUCK OUT TONGUE respectively per US comment T4b and Japanese comment JP10 (note the US comment says: STUCK-OUT with a dash). Also, given the proposed new name, the annotations are unnecessary.

f) 1F60C FACE THROWING A KISS > KISSING FACE WITH WINKING EYE
= e-32C
Propose non acceptance
Propose instead HAPPY FACE THROWING A KISS per US comment T4b and Japanese comment JP10 (N3778).

g) 1F60D FACE KISSING > KISSING FACE WITH CLOSED EYES
= e-32D
Propose non acceptance
Propose instead HAPPY FACE KISSING per US comment T4b and Japanese comment JP10 (N3778). Note that Irish comment E3 also asks for a glyph change.

h) 1F60E FACE WITH MASK > FACE WITH MEDICAL MASK
= e-32D
Propose acceptance
Also supported by US comment T4c and Japanese comment JP10 (N3778).

i) 1F610 HAPPY FACE WITH OPEN MOUTH > SMILING FACE WITH OPEN MOUTH
= e-330
1F611 HAPPY FACE WITH OPEN MOUTH AND COLD SWEAT > SMILING FACE WITH OPEN MOUTH AND COLD SWEAT
= e-331
Propose acceptance
Consistent with other accepted changes
j) 1F612 HAPPY FACE WITH OPEN MOUTH AND CLOSED EYES > SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES
= e-332
Propose acceptance in principle
Consistent with other accepted changes, but do we need to add 'tightly' to 'closed eyes'?

k) 1F613 HAPPY FACE WITH GRIN > GRINNING FACE
= e-333
Propose non acceptance
Grin does not imply happiness.

l) 1F614 HAPPY AND CRYING FACE > GRINNING AND CRYING FACE
= e-334
Propose non acceptance
Happiness disappears with the proposed name change. Note that Japanese comment JP10 (N3778) also requests a glyph change.

m) 1F615 HAPPY FACE WITH WIDE MOUTH AND RAISED EYEBROWS > SMILING FACE WITH SQUINTING EYES
= e-335
Propose non acceptance
Propose instead HAPPY FACE WITH WIDE MOUTH AND SMILING EYES per US comment T4b and Japanese comment JP10 (N3778).

n) 1F626 FACE WITH HALO > SMILING FACE WITH HALO
1F627 FACE WITH HORNS > SMILING FACE WITH HORNS
1F628 FACE WITH SUNGLASSES > SMILING FACE WITH SUNGLASSES
Propose acceptance
Consistent with other accepted changes

o) 1F62B CAT FACE WITH OPEN MOUTH > SMILING CAT FACE WITH OPEN MOUTH
= e-348
Propose non acceptance
Propose instead HAPPY CAT FACE WITH OPEN MOUTH per US comment T4b and Japanese comment JP10 (N3778). Note N3378 mentions the character code point as 1F62A in error.

p) 1F62C HAPPY CAT FACE WITH GRIN > GRINNING CAT FACE
= e-349
Propose non acceptance
See k.

q) 1F62D HAPPY AND CRYING CAT FACE > GRINNING AND CRYING CAT FACE
= e-34A
Propose non acceptance
See l. Note that Japanese comment JP10 (N3778) also requests a glyph change (N3778 uses 1F62C in error).

r) 1F62E CAT FACE KISSING > KISSING CAT FACE WITH CLOSED EYES
= e-34B

Propose non acceptance
Propose instead HAPPY CAT FACE KISSING per US comment T4b and Japanese comment JP10 (N3778). Note N3378 also requests a glyph change (uses 1F62D in error).

s) 1F62F CAT FACE WITH HEART-SHAPED EYES > SMILING CAT FACE WITH HEARTSHAPED EYES
= e-34C
Propose non acceptance
Propose instead HAPPY CAT FACE WITH HEART-SHAPED EYES per US comment T4b and Japanese comment JP10 (N3778). Note N3378 mentions the character code point as 1F62E in error.

t) 1F632 CAT FACE WITH TIGHTLY-CLOSED LIPS > SMIRKING CAT FACE
= e-34F
Propose instead SMART CAT FACE WITH TIGHTLY-CLOSED LIPS per US comment T4b and Japanese comment JP10 (N3778). Note N3378 mentions the character code point as 1F631 in error.

u) 1F633 ANGUISHED CAT FACE > WEARY CAT FACE
= e-350
Propose non acceptance
Because following E3 requests a glyph change as well it is unclear whether this is still the originally proposed character.

T5. Page Page 92, Row 1F60: Emoticons.

Ireland requests a number of character additions. Most of of these are “generic” forms of faces which are otherwise used in the source environment of Japanese telephony. For instance, there are three different grinning faces used (with eyes and eyebrows, with squinting eyes, and with squinting eyes and tears) but an ordinary grinning face is missing. In T6 below we propose to order the emoticons according to mouth shape, and so the generic mouth shapes are proposed here. Some other characters are proposed because of differences in cultural identity. For example, SLEEPY FACE is recognized in Japan (and perhaps elsewhere) by the “snot bubble”; this is completely unknown in the West (indeed the image has quite different connotations than the “cute sleepiness” known in Asia). On the other hand, the face with three zeds is very widely recognized, and is implemented in instant messaging emoticons, so SLEEPING FACE has been proposed for addition. (We do not care for the name SLEEPY FACE for the former, but doubt that FACE WITH SNOT BUBBLE would be more felicitous.) Another example would be WEARY FACE here, which correctly maps to a KDDI character; it differs from ANGUISHED FACE in terms of its eye-shape. The proposed additions are shown below in the chart in yellow-highlighted glyph cells.
1F640 GRINNING FACE
1F641 NEUTRAL FACE
1F642 EXPRESSIONLESS FACE
1F643 CONFUSED FACE
1F644 KISSING FACE
1F645 KISSING FACE WITH SQUINTING EYES
1F646 FACE WITH STUCK-OUT TONGUE
1F647 WORRIED FACE
1F648 FROWNING FACE WITH OPEN MOUTH
1F649 WEARY FACE
= e321
1F64A GRIMACING FACE
1F64B FACE WITH OPEN MOUTH
1F64C HUSHED FACE
1F64D SLEEPING FACE
WG2 discussion
Need Emoji ad hoc. However, the editor is not in favor such a large addition at FDAM stage without peer review. Note that 1F641 NEUTRAL FACE is also requested by the US comment T4a but at location 1F62A.

T6. Page 92, Row 1F60: Emoticons.

Ireland requests that the following characters be rearranged in the code chart. A number of the annotations have been changed. The characters have been ordered by mouth-shape, and within each mouth shape have been ordered by eye-shape. We have attempted to be accurate here and above; the chart at the end is what we intend.

@ Grinning faces
1F600 GRINNING FACE (moved up from 1F640)
1F601 GRINNING FACE WITH LOOK OF TRIUMPH (moved up from 1F608)
= e-328
1F602 GRINNING FACE WITH SQUINTING EYES (moved up from 1F613)
= e-333
1F603 GRINNING AND CRYING FACE (moved up from 1F614)
= e-334

@ Smiling faces with open mouths
1F604 SMILING FACE WITH OPEN MOUTH (moved up from 1F610)
= e-330
1F605 SMILING FACE WITH OPEN MOUTH AND SQUINTING EYES (moved up from 1F615)
= e-338
1F606 SMILING FACE WITH OPEN MOUTH AND COLD SWEAT (moved up from 1F611)
= e-331
1F607 SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES (moved up from 1F612)
= e-332

@ Smiling faces
1F608 SMILING FACE WITH HALO (moved up from 1F626)
1F609 SMILING FACE WITH HORNS (moved up from 1F627)
1F60A WINKING FACE (moved up from 1F625)
= e-347
1F60B SMILING FACE WITH SQUINTING EYES (moved up from 1F615)
= e-335
1F60C FACE SAVOURING DELICIOUS FOOD (moved down from 1F60B)
= e-32B
1F60D RELIEVED FACE (moved up from 1F61C)
= e-33E
1F60E SMILING FACE WITH HEART-SHAPED EYES (moved down from 1F607)
= e-327
1F60F SMILING FACE WITH SUNGLASSES (moved up from 1F628)
1F610 SMIRKING FACE (moved up from 1F621)
= e-343

@ Faces with flat mouths
1F611 NEUTRAL FACE (moved up from 1F641)
* used for the West Wind in some Mahjong annotation
1F612 EXPRESSIONLESS FACE (moved up from 1F642)
1F613 UNAMUSED FACE (moved down from 1F606)
= e-326
1F614 FACE WITH COLD SWEAT (moved up from 1F622)
= e-344
1F615 PENSIVE FACE (moved up from 1F61E)
= e-340
1F616 CONFUSED FACE (moved up from 1F643)
1F617 CONFOUNDED FACE (moved up from 1F61D)
= e-33F

@ Kissing faces
1F618 KISSING FACE (moved up from 1F644)

1F619 KISSING FACE WITH WINKING EYE (moved down from 1F60C)
= e-32C
1F61A KISSING FACE WITH SQUINTING EYES (moved up from 1F645)
1F61B KISSING FACE WITH CLOSED EYES (moved down from 1F60D)
= e-32D

@ Faces with stuck-out tongues
1F61C FACE WITH STUCK-OUT TONGUE (moved up from 1F646)
1F61D FACE WITH STUCK-OUT TONGUE AND WINKING EYE (moved down from 1F609)
= e-329
1F61E FACE WITH STUCK-OUT TONGUE AND TIGHTLY-CLOSED EYES (moved down from 1F60A)
= e-32A

@ Frowning faces
1F61F DISAPPOINTED FACE (moved down from 1F603)
= e-323
1F620 WORRIED FACE (moved up from 1F647)
1F621 ANGRY FACE (moved down from 1F600)
= e-320
1F622 POUTING FACE (moved down from 1F61B)
= e-33D
1F623 CRYING FACE (moved down from 1F617)
= e-339
1F624 PERSEVERING FACE (moved down from 1F61A)
= e-33C
1F625 DISAPPOINTED BUT RELIEVED FACE (moved down from 1F623)
= e-345

@ Frowning faces with open mouths
1F626 FROWNING FACE WITH OPEN MOUTH (moved up from 1F648)
1F627 ANGUISHED FACE (moved down from 1F601)
1F628 FEARFUL FACE (moved down from 1F619)
= e-33B
1F629 WEARY FACE (moved up from 1F649)
= e-321
1F62A SLEEPY FACE (moved down from 1F620)
= e-342
1F62B TIRED FACE (moved down from 1F624)
= e-346

@ Grimacing faces
1F62C GRIMACING FACE (moved up from 1F64A)
1F62D LOUDLY CRYING FACE (moved down from 1F618)
= e-33A

@ Faces with open mouths
1F62E FACE WITH OPEN MOUTH (moved up from 1F64B)
1F62F HUSHED FACE (moved up from 1F64C)
1F630 AGITATED FACE (moved down from 1F605)
= e-325
1F631 FACE SCREAMING IN FEAR (moved down from 1F61F)
= e-341
1F632 ASTONISHED FACE (moved down from 1F602)
= e-322
1F633 FLUSHED FACE (moved down from 1F60F)

= e-32F
1F634 SLEEPING FACE (moved up from 1F64D)
1F635 DIZZY FACE (moved down from 1F604)
= e-324

@ Faces without mouths
1F636 FACE WITHOUT MOUTH (moved down from 1F629)
* used for the South Wind in some Mahjong annotation
1F637 FACE WITH MEDICAL MASK (moved down from 1F60E)
= e-32E

@ Cat faces
1F638 GRINNING CAT FACE WITH SQUINTING EYES (moved down from 1F62C)
= e-349
1F639 GRINNING AND CRYING CAT FACE (moved down from 1F62D)
= e-34A
1F63A SMILING CAT FACE WITH OPEN MOUTH (moved down from 1F62B)
= e-348
1F63B SMILING CAT FACE WITH HEART-SHAPED EYES (moved down from 1F62F)
= e-34C
1F63C SMIRKING CAT FACE (moved down from 1F632)
= e-34F
1F63D KISSING CAT FACE WITH CLOSED EYES (moved down from 1F62E)
= e-34B
1F63E POUTING CAT FACE (moved down from 1F631)
= e-34E
1F63F CRYING CAT FACE (moved down from 1F630)
= e-34D
1F640 WEARY CAT FACE (moved down from 1F633)
= e-350

@ Gesture symbols
1F645 FACE WITH NO GOOD GESTURE (moved down from 1F634)
= e-351
1F646 FACE WITH OK GESTURE (moved down from 1F635)
= e-352
1F647 PERSON BOWING DEEPLY (moved down from 1F636)
= e-353
1F648 SEE NO EVIL MONKEY (moved down from 1F637)
= e-354
1F649 SPEAK NO EVIL MONKEY (moved down from 1F638)
= e-355
1F64A HEAR NO EVIL MONKEY (moved down from 1F639)
= e-356
1F64B PERSON RAISING ONE HAND (moved down from 1F63A)
= e-357
1F64C PERSON RAISING BOTH HANDS IN EXULTATION (moved down from 1F63B)
= e-358
1F64D PERSON FROWNING (moved down from 1F63C)
= e-359
1F64E PERSON WITH POUTING FACE (moved down from 1F63D)
= e-35A
1F64F PERSON WITH FOLDED HANDS (moved down from 1F63E)
= e-35B

WG2 discussion

Need Emoji ad hoc. This is related to the acceptance of the addition request in T5.

Editorial comments

E1. Page 30, Row 1BC: Batak.

Ireland suggests that the glyphs for 1BFA and 1BFB be increased in size somewhat. Note that the size of the characters is not significant except in terms of paleography. While the top three Batak examples below were taken from van der Tuuk in the 1860s, modern researchers do not require such presentation, and fonts which display the characters as in the fourth through seventh examples are equally as acceptable:
(See examples in original Irish Ballot in SC2 N4130)
WG2 discussion
Could become not applicable if comment T2 from US is accepted. See also comment JP3 from Japan.

E2. Page 88, Row 1F30: Miscellaneous Pictographic Symbols.

Ireland requests that the following annotation be added:
1F4A6 DROP OF WATER
* also represents a drop of sweat
Propose acceptance
See also comment JP11d from Japan.

E3. Page 92, Row 1F60: Emoticons.

Ireland requests a number of glyph changes. The proposed changes are shown below in the chart in blue-highlighted glyph cells.
1F607 (1F612) SMILING FACE WITH OPEN MOUTH AND TIGHTLY-CLOSED EYES eyes have changed
1F60D (1F61C) RELIEVED FACE eyes changed from squinting to closed
1F617 (1F61D) CONFOUNDED FACE mouth fixed; stray streamer deleted.
1F61B (1F60D) KISSING FACE WITH CLOSED EYES eyes changed from squinting to closed
1F628 (1F619) FEARFUL FACE mouth moved down
1F62A (1F620) SLEEPY FACE mouth moved down
1F62B (1F624) TIRED FACE mouth moved down
1F631 (1F61F) FACE SCREAMING IN FEAR hands added, mouth made bigger
1F63C (1F632) SMIRKING CAT FACE should eyebrows change like 1F621?
1F63D (1F62E) KISSING CAT FACE WITH CLOSED EYES eyes changed to be like 1F61B, mouth fixed
1F63E (1F631) POUTING CAT FACE eyebrows changed
1F640 (1F633) WEARY CAT FACE eyes changed
WG2 discussion
The editor added a parenthetical notation to show code points used in the FPDAM8 document (prior to the moves suggested in T6).
Unclear what the change is for 1F607 (1F612)
Glyph change for 1F60D (1F61C) also supported by Japan comment JP 10 (N3378)
Glyph change for 1F617 (1F61D) also supported by Japan comment JP 10 (N3378) and US comment T4d.
Controversy on 1F61B (1F60D), 1F631 (1F61F), 1F63E (1F631) between Ireland and Japan
Controversy on 1F63D (1F62E) between Ireland and Japan/US
Changes for 1F628 (1F619), 1F62A (1F620), 1F62B (1F624) probably acceptable
Open question for 1F63C (1F632)
Change for 1F640 (1F633) is linked to resolution of comment T4.u.

Japan, Negative

(Japan did not provide instruction as to which comment satisfaction would change its vote)

JP.T1 (technical), Page 1, Changes to Page 20, Clause 26, Special features of individual scripts and symbols repertoires:

Change The current text lacks the indication that the emoji source reference information contained in EmojiSrc.txt is informative. Also, some reviewers pointed out that the current text on EmojiSrc.txt is somewhat ambiguous; it can lead to a misunderstanding that the EmojiSrc.txt is intended for transcoding.

Proposed change:

Add the following sentence at the end of the first paragraph to the new "26.4 Source references for pictographic symbols":

The source reference information shall be informative.

Add the following sentence at the end of the NOTE to the clause:

EmojiSrc.txt is for reference purpose only and is not intended for transcoding.

Propose acceptance in principle

Text at end of the first paragraph reads:

The source reference information is informative.

Concerning the note addition, the editor would like to have a clarification of what is meant by transcoding. The term itself is ambiguous. The term 'mapping' could be better, if intended.

JP.T2 (technical): Page 1, Changes to Page 20, Sub-clause 27.1, Source references for CJK Unified Ideographs

The name "Hanyo-Denshi Program (汎用電子情報交換環境整備プログラム)" should not be followed by a year.

Proposed change:

Remove "2009".

WG2 discussion

Most of the source references (except for G sources) are currently dated. This is an important feature which improves the immutability of the references. Without it, the references may be invalidated by a further revision of the source document without any way for the user of the standard to be alerted.

Because this principle is not formally enforced, the Japanese request could be accommodated but is nevertheless not optimal. An alternative is to have some mentions of edition number.

JP.E3 (editorial): Page 30, Batak codechart

The glyphs for the newly added 1BFA and 1BFB are too small and hardly recognizable. Japan understands that these two characters are actually very wide and that the editor had to shrink them to fit in the cells in the code chart. However, it is important to make the code chart recognizable, and we need a special handling for these two characters.

Proposed change:

Remove these two glyphs from the code chart, put them in a separate figure (or chart) of sufficient width, and put in the code chart some words indicating "the representative glyphs for 1BFA and 1BFB are in figure X.X".

WG2 discussion

Could become not applicable if comment T2 from US is accepted. In all cases, removing characters from code charts and putting them in a different section or clause is setting a precedent that the editor is very reluctant to start. There are many other characters in the standard that are naturally large and are reduced in size to fit the cell box. The solution should be removal of these two characters from the amendment until reasonable consensus is achieved. See also comment E1 from Ireland.

JP.E4 (editorial): Page 42, Miscellaneous Symbols codechart

The glyphs for the newly added 26E5 and 26E6 should have some clear gaps to represent the interlacements of the edges, but in the current code chart the gaps are too narrow and hardly recognizable.

Proposed change:

Adjust the glyphs to make the gaps wider so that the intended interlacements are clear.

Propose acceptance in principle
Agree that the design is not optimal. The editor will work with the submitter to get better aligned characters.

JP.T5 (technical): Page 49, Name list for Dingbats (miscellaneous)

As emoji, three characters for Rock, Paper, and Scissors are used as a set. The newly added two, Rock and Paper, have appropriate annotations referring to "Rock, Paper, Scissors game" after the character name, but the existing character 270C VICTORY HAND, that is now unified with Scissors, has no such annotation. It is dangerous that a user may misunderstand that the 270C is *not* for Scissors, because RAISED FIST and RAISED HAND have appropriate annotations on the game, but VICTORY HAND doesn't.
Proposed change:
Add the following annotation to 270C VICTORY HAND:
= scissors in Rock, Paper, Scissors game
Propose acceptance
Strictly speaking, because this concerns an annotation addition, this is not a technical comment, it is however acceptable.

JP.T6 (technical): Page 77, Name list for Enclosed Alphanumeric Supplement, the explanatory text for Regional Indicator Symbols

Japan national body supports the proposed uses of the newly added characters in range 1F1E6 to 1F1FF (Regional Indicator Symbols) in the document N3779.
Proposed change:
Replace the appropriate texts with the ones proposed in N3779:
(added by editor):
In a sentence between the name list for Regional Indicator Symbols, a phrase something like: "in a pairs joined with ZERO WIDTH JOINER" should be used to explain the intended use of the characters.
In EmojiSrc.txt file, put ZERO WIDTH JOINER between two Regional Indicator Symbols to show the source reference, i.e.,
1F1E8 200D 1F1F3;;F3D2;FBB3
1F1E9 200D 1F1EA;;F3CF;FBAE
1F1EA 200D 1F1F8;;F348;FBB1
1F1EB 200D 1F1F7;;F3CE;FBAD
1F1EE 200D 1F1F9;;F3D0;FBAF
1F1EF 200D 1F1F5;;F6A5;FBAB
1F1F0 200D 1F1F7;;F3D3;FBB4
1F1F7 200D 1F1FA;;F349;FBB2
1F1FA 200D 1F1F0;;F3D1;FBB0
1F1FA 200D 1F1F8;;F790;FBAC.
Propose non acceptance
This seems to be overkill. Valid processes will not produce incorrect sequences and mitigation of invalid sequences is simple. A singleton is invalid, and sequences of more than two characters can be analyzed on the first two characters. In fact, adding a third character (ZWJ) increases possible cases of ill-formed sequences containing random numbers of these 26 characters and ZWJ. If anything, a note describing suggested processing of ill-formed sequences could be added.

JP.G7 (general): Page 84, Explanatory text at the top of the name list for Miscellaneous Pictographic Symbols

The current text is ambiguous and allusive. It may confuse readers rather than making the thing clear. Japan National Body believes that the text is intended for a clarification that the existing practice of showing emoji in animated and/or colourful graphics is fine. We need a more explicit text here. Japan is not wedded to the proposed alternate text; it accepts any better ones as long as they are clear.
Proposed change:
Replace the texts with something more explicit, e.g.:
Although the representative glyphs in this International Standard are consistently presented in black and white drawing, it does not prevent implementations from using graphic symbols with some specific colour or even with

multiple colours, fully or partly animated graphics, or both. It should also be noted that the actual graphics shown on the code chart are informative as all other characters of this International Standard.

Propose acceptance in principle

The current text says:

"In the case of characters that are associated in typical usage with a particular colour, conventions of European heraldry are used to represent those colours in monochromatic line drawings. This aspect of the representative glyphs is informative, not normative."

It would probably be better to remove the note on top of page 84 and replace it by updated text in clause 17 'Structure of the code charts and list's that would therefore apply to all pages of the code chart. It could also address concerns about 'black' and 'white' terminology used in character names. The updated text in clause 17 could replace the second paragraph as follows (new text underlined):

Graphic characters specified in ISO/IEC 10646 are to be regarded as typical visual representations of the characters. ISO/IEC does not attempt to prescribe the exact shape or color of each character. The shape is affected by the design of the font employed, which is outside the scope of ISO/IEC 10646. Although the representative glyphs in this International Standard are consistently presented in black and white, it does not prevent implementations from using graphic symbols with some specific colour or even with multiple colours, fully or partly animated graphics, or both. When characters are typically associated with a particular colour, conventions of European heraldry are used to represent those colours in monochromatic line drawings in the code charts. Furthermore, the usage of 'BLACK' and 'WHITE' in character name does not imply a specific colour. It is simply a distinction between a filled character and an outline character.

In addition, a new note could be added in clause 34.1 Code chart:

NOTE – Graphic symbols corresponding to the representation of graphic characters are informative. See 17.

JP.T8 (technical): Page 88, Character name for 1F4BE

The character name for a symbol representing a floppy disk is currently "FLOPPY DISK", but a floppy disk is called "flexible disk cartridge" in ISO/IEC. See appropriate International Standards, e.g., ISO 6596, ISO 8860, or ISO/IEC 9529.

Proposed change:

Change the character name to "FLEXIBLE DISK CARTRIDGE"

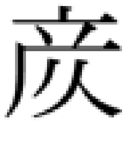
Propose non acceptance or acceptance in principle

The current term is vastly more commonly used than the proposed term. A non scientific search results in 4120000 versus 142000 hits (ratio of 1 to 29). To accommodate this suggestion in principle, the current annotation could be changed as following:

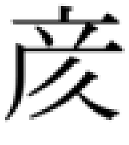
= flexible disk cartridge, floppy disc, floppy, diskette

JP.E9 (editorial): Page 99, JK Unified Ideographs Extension-D code chart

The glyphs for the newly added 2B778 (JH-IB0679) is slightly wrong. It is by a mistake of Japan National Body when it produced and submitted the TrueType font for CJK D. The correct one is shown below (right):



FPDAM 8



correct design

Proposed change:

Replace the glyph with the correct one. (Japan National Body will provide a TrueType font that contains the correct glyph.)

Propose acceptance

Assuming such font is delivered to the editor for publication.

JP.T10 (technical): All, Emoji symbols

Japan National Body supports the concerns and proposals expressed in the document N3778 regarding some emoji symbols.

Proposed change:

Change Make appropriate changes to the representative glyphs and character names as proposed in the document N3778.

(The editor corrected all shifted values between 1F62A and 1F63D to 1F62B-1F63E)

a) Name changes (from 1F601 to 1F63E)

e-321 / 1F601

Changing into the word that seems closer to the nuances of the original sources.

ANGUISHED FACE→WEARY FACE

e-325 / 1F605

Changing into the word that seems closer to the nuances of the original sources.

EXASPERATED FACE→FACE WITH COLD SWEAT

e-326 / 1F606

Changing into the word that seems closer to the nuances of the original sources.

EXPRESSIONLESS FACE→UNAMUSED FACE

e-327 / 1F607

Adding the word of emotion or condition.

FACE WITH HEART SHAPED EYES→HAPPY FACE WITH HEART SHAPED EYES

e-329 / 1F609

Adding the word of emotion or condition.

WINKING FACE WITH STUCK OUT TONGUE→KIDDING AND WINKING FACE WITH STUCK OUT TONGUE

e-32A / 1F60A

Adding the word of emotion or condition.

FACE WITH STUCK OUT TONGUE→KIDDING FACE WITH STUCK OUT TONGUE

e-32C / 1F60C

Adding the word of emotion or condition.

FACE THROWING A KISS→HAPPY FACE THROWING A KISS

e-32D / 1F60D

Adding the word of emotion or condition.

FACE KISSING→HAPPY FACE KISSING

e-32E / 1F60E

Adding the word of emotion or condition and the term MEDICAL for strictness.

FACE WITH MASK→FACE WITH MEDICAL MASK

e-335 / 1F615

Although the draft character name includes RAISED EYEBROWS SoftBank #2, one of the sources of this character, does not have eyebrows. The parallels between two sources (KDDI #454 and SoftBank #2) are WIDE MOUTH and SMILING EYES (eyes of upper half circles).

HAPPY FACE WITH WIDE MOUTH AND RAISED EYEBROWS→HAPPY FACE WITH WIDE MOUTH AND SMILING EYES

e-338 / 1F616

Although the draft character name includes RAISED EYEBROWS, SoftBank #1, the single source of this character, does not have eyebrows. The element in the source glyph except OPEN MOUTH is SMILING EYES (eyes of upper half circles).

HAPPY FACE WITH OPEN MOUTH AND RAISED EYEBROWS→HAPPY FACE WITH OPEN MOUTH AND SMILING EYES

e-348 / 1F62B

Adding the word of emotion or condition.

CAT FACE WITH OPEN MOUTH→HAPPY CAT FACE WITH OPEN MOUTH

e-34B / 1F62E

Adding the word of emotion or condition.

CAT FACE KISSING→HAPPY CAT FACE KISSING
e-34C / 1F62F
Adding the word of emotion or condition.
CAT FACE WITH HEART SHAPED EYES→HAPPY CAT FACE WITH HEART SHAPED EYES
e-34F / 1F632
Adding the word of emotion or condition.
CAT FACE WITH TIGHTLY CLOSED LIPS→SMART CAT FACE WITH TIGHTLY CLOSED LIPS
e-357 / 1F63A
Adding the word of emotion or condition.
PERSON RAISING ONE HAND→HAPPY PERSON RAISING ONE HAND
e-35B / 1F63E
Adding the word of emotion or condition.
PERSON WITH FOLDED HANDS→PLEADING PERSON WITH FOLDED HANDS
Propose acceptance
See also comment T4 from Ireland and T4b and T4c from US which among the two cover all comments from Japan. The US has similar requests for 1F601, 1F605, 1F606, 1F607, 1F609, 1F60A, 1F60C, 1F60D, 1F60E, 1F615, 1F616, 1F62B, 1F62E, 1F62F, 1F632, 1F63A, 1F63E.

b) glyph changes for 1F606, 1F608, 1F60C, 1F60D, 1F614, 1F615, 1F616, 1F61C, 1F61D, 1F61F, 1F621, 1F62D, 1F62E, 1F631
WG2 discussion
See also comments E3 from Ireland and T4e from US.
US support same changes for 1F61D and 1F62E.
Ireland supports same changes for 1F61C and 1F61D
Ireland has conflicting changes for 1F60D, 1F61F, 1F631

JP.G11 (general): all, Emoji set

Japan National Body received the documents N3776 and N3777 from interested parties. Japan wants WG 2 to consider these inputs.

a) N3776 (DoCoMo) character 'color', make new emojis or change mapping
Concerns mapping to 2600, 2665, 2666, 267B, 2702, 2764
Propose non acceptance or acceptance in principle
Black and White do not imply specific colors, see disposition of comment JP7 for further details.

b) N3776 (DoCoMo) U+27BF DOUBLE CURLY LOOP
"S-JIS code F984" shows "Free dial service" which is the NTT Communications' trademark of toll-free service in Japan, so the current mapping is not appropriate. New emoji corresponding to "toll-free service" shall be defined, or this mapping shall be deleted.
WG2 discussion
See also comment JP12.

c) N3776 (DoCoMo) U+1F35C STEAMING BOWL
The glyph of "1F35C" recalls noodle, but its explanation shows it is not limited to noodle. Therefore a new glyph which includes not only "noodle" but also other bowls e.g. beef bowl shall be replaced.
WG2 discussion
The explanation (annotation) is only informative, it does not seem necessary to disunify the character. If done, it is important to differentiate from 1F375 TEACUP WITHOUT HANDLE.

d) N3776 (DoCoMo) U+1F4A7 DROP OF WATER
The glyphs of "S-JIS code F9AC" and "1F3A7" is very similar, but their explanations are different, SWEAT" and "WATER". The name of "1F4A7" shall be changed to "DROP OF SWEATS", or the mapping shall be deleted..
Propose acceptance in principle
See also comment E2 from Ireland.

Adding the annotation 'also represents a drop of sweat' as suggested by Ireland should suffice.

e) N3777 (KDDI) character 'color', make new emojis or change mapping
Concerns 26AA, 26AB, 2705, 1F0CF
Propose acceptance in principle
Black and White do not imply specific colors, see disposition of comment JP7 for further details.

f) N3777 (KDDI) U+1F686 TRAIN
Provided TRAIN character is "front side train" while KDDI's TRAIN character is "sideways train". Please change TRAIN character from "front side train" to "sideways train". Also, please kindly note that KDDI Shift-JIS"F68E" includes both "TRAIN" and "RAILWAY CAR"(No need to differentiate TRAIN and RAILWAY CAR).
Propose acceptance in principle
Because glyphs are not normative, front or side view could be used. However, the best solution is probably to remap to 1F683 RAILWAY CAR.

g) N3777 (KDDI) U+267B BLACK UNIVERSAL RECYCLING SYMBOL
KDDI's character "F47D" does not stand for recycling symbol; thus, please remove this unicode character mapping for "F47D".
WG2 discussion
Could have changed mapping for F47D to 1F503 CLOCKWISE DOWNWARDS AND UPWARDS OPEN CIRCLE ARROWS, but it is already in use for F7E4.

h) N3777 (KDDI) Mapping changes for KDDI F65A, F6D5, F6D3, F6EA, F688
Propose non acceptance
Accepting these mapping changes would compromise interoperability.

JP.T12 (technical): EmojiSrc.txt

The source emoji symbols, DoCoMo F984 and Softbank F7B1, are intended for a particular symbol that is believed to be a registered trademark. Japanese mobile phone implementations will surely keep using the exact shape that is registered in Japan Patent Office. On the other hand, the UCS symbol DOUBLE CURLY LOOP should have some other uses today, and the best design that fits to the existing practice will differ from the shape of the registered trademark. By tying these symbols, Japanese mobile phones will be unable to support the users' expectation based on the current practice of the DOUBLE CURLY LOOP character. It breaks interoperability..
Proposed change:
Remove the following entry for DOUBLE CURLY LOOP:
27BF;F984;;F7B1"
WG2 discussion
See also comment JP11b.
Even if the symbol may have other use, the glyph is unlikely to be changed. It is difficult to imagine how having the mapping breaks interoperability. It seems to be exactly the opposite. We would expect most implementers will keep use the mapping anyway. In other words, we see little gain in removing it.

United Kingdom: Positive with comments

Technical comments:

T.1. Clause 34 Latin Extended-D A78F LATIN LETTER MIDDLE DOT

We reaffirm our support for the encoding of this character, which we can confirm is required for transliteration and phonetic transcription by scholars working with Phags-pa and Tangut.

WG2 discussion

See also comment T1 from Ireland and T1 from US.
Discussion in disposition of comment T1 from Ireland.

T.2. Clause 34 Alchemical Symbols

1F70D ALCHEMICAL SYMBOL FOR SULPHUR
= brimstone
1F70E ALCHEMICAL SYMBOL FOR PHILOSOPHERS
SULPHUR
1F70F ALCHEMICAL SYMBOL FOR BLACK SULPHUR
= sulphur nigra, dye

We reaffirm our support for the spelling "SULPHUR" in the character names for 17F0D, 17F0E and 17F0F. Consultation with UK academics studying the alchemical texts of Newton confirms that "SULPHUR" is the correct and preferred spelling. "Sulphur" is also the spelling used in the Oxford English Dictionary, which should be the primary source for the spelling of English words in the standard**.**

WG2 discussion

It is true that British spelling has been the norm for the body of the standard; however there is no such principle for the character names. Furthermore there are no implied rules in ISO/IEC 10646 Principles and Procedures that the character names should follow the Oxford English Dictionary. Because the Chemistry organizations prefer the 'sulfur' spelling, there is really no reason not to respect it.

USA: Negative

The U.S. National Body is voting No with comments on the following SC2 ballot: SC2 N4123: Information technology - Universal Multiple - Octet Coded Character Set (UCS) - AMENDMENT 8: Additional symbols, Bamum supplement, CJK Unified Ideographs Extension D, and other characters. If comments T.1, T.2, T.5, and T.6.a are accommodated, the U.S. will change its vote to Yes.

Technical comments:

T.1. Latin Extended D

The U.S. requests the removal of U+A78F LATIN LETTER MIDDLE DOT and reiterates that this character is unnecessary and is a damaging duplication for the standard. Justification for this request is contained in N3678 (L2/09‑278). A viable alternative to encoding a separate letter middle dot, for the purposes cited by the original proposal, would be to use the already encoded modifier letter, U+02D1 MODIFIER LETTER HALF TRIANGULAR COLON.

WG2 discussion

See also comment GB1 from UK and T1 from Ireland.

T.2. Batak

The U.S. requests the removal of two Batak symbols:
1BFA BATAK SYMBOL BINDU GODANG
1BFB BATAK SYMBOL BINDU PINARJOLMA

Rationale: The evidence provided in N3320 does not demonstrate these marks to be characters, but to be graphic page elements that do not behave at all as characters. (See figures 11 and 13 in N3320 for clear examples showing that these are not characters.).

WG2 discussion

See also comment E1 from Ireland and JP3 from Japan. Given the controversy and the complication concerning rendering, it seems prudent to postpone the encoding of these proposed characters.

T.3. Alchemical Symbols

The U.S. requests three name changes in the Alchemical Symbols block, all involving a spelling change from "SULPHUR" to "SULFUR":
1F70D ALCHEMICAL SYMBOL FOR SULFUR
1F70E ALCHEMICAL SYMBOL FOR PHILOSOPHERS SULFUR
1F70F ALCHEMICAL SYMBOL FOR BLACK SULFUR

Rationale: The spelling with "F" is the one approved by the International Union of Pure and Applied Chemistry (1990) and the Royal Society of Chemistry, and is the consensus spelling used internationally.

WG2 discussion

See also comment GB2 from UK.

T.4. Emoticons

a. The U.S. requests 1F62A NEUTRAL FACE be added to the Emoticons block, with the glyph, rationale, and properties as described in N3769 (L2/10-036).

WG2 discussion

See also comment T5 from Ireland. This character is also requested by Ireland.

b. The U.S. recommends the following 15 name changes suggested in N3711:
1 e-321=U+1F601 ANGUISHED FACE → WEARY FACE
Rationale: The KDDI source is an onomatopoeia which suggests the person is tired and whining.
2. e-326=U+1F606 EXPRESSIONLESS FACE → UNAMUSED FACE
Rationale: This follows a significant glyph change agreed to in Tokyo.
3. e-327=U+1F607 FACE WITH HEART‑SHAPED EYES → HAPPY FACE WITH HEARTSHAPED EYES

4. e-329=U+1F609 WINKING FACE WITH STUCK-OUT TONGUE → KIDDING AND WINKING FACE WITH STUCK-OUT TONGUE
5. e-32A=U+1F60A FACE WITH STUCK-OUT TONGUE → KIDDING FACE WITH STUCK-OUT TONGUE
6. e-32C=U+1F60C FACE THROWING A KISS → HAPPY FACE THROWING A KISS
7. e-32D=U+1F60D FACE KISSING → HAPPY FACE KISSING
8. e-335=U+1F615 HAPPY FACE WITH WIDE MOUTH AND RAISED EYEBROWS→ HAPPY FACE WITH WIDE MOUTH AND SMILING EYES
Rationale: ″raised eyebrows″ refer to older glyph designs than what was agreed to in FPDAM8
9. e-338=U+1F616 HAPPY FACE WITH OPEN MOUTH AND RAISED EYEBROWS → HAPPY FACE WITH OPEN MOUTH AND SMILING EYES
Rationale: ″raised eyebrows″ refers to older glyph designs than what was agreed for FPDAM8.
10. e-348=U+1F62B CAT FACE WITH OPEN MOUTH → HAPPY CAT FACE WITH OPEN MOUTH
11. e-34B=U+1F62E CAT FACE KISSING → HAPPY CAT FACE KISSING
12. e-34C=U+1F62F CAT FACE WITH HEART﹣SHAPED EYES → HAPPY CAT FACE WITH HEART-SHAPED EYES
13. e-34F=U+1F632 CAT FACE WITH TIGHTLY﹣CLOSED LIPS → SMART CAT FACE WITH TIGHTLY-CLOSED LIPS
14. e-357=U+1F63A PERSON RAISING ONE HAND → HAPPY PERSON RAISING ONE HAND
15. e-35B=U+1F63E PERSON WITH FOLDED HANDS → PLEADING PERSON WITH FOLDED HANDS

Propose acceptance

See also comments T4 from Ireland and JP10 (N3778) from Japan. N3711 was superseded by N3778. All the changes requested by the US are included in JP10. Note that N3378 has code points for value 1F62A and above off by one (e.g. 1F62A should read 1F62B).

c. The U.S. does not recommend the following two name changes in N3711:
1. e-32E=U+1F60E FACE WITH MASK → SAD FACE WITH MEDICAL MASK
2. e-325=U+1F605 EXASPERATED FACE → IMPATIENT FACE
The U.S. recommends instead the following names:
1. e-32E=U+1F60E FACE WITH MEDICAL MASK
Rationale: While including "medical" in the name is good, ″sad″ seems too strong.
2. e-325=U+1F605 FACE WITH COLD SWEAT
Rationale: The KDDI meaning is "be hasty" and SoftBank literally means someone is offbalance or upset. Both reflect that someone is not calm and has lost his composure.

Propose acceptance

See also comments T4 from Ireland and JP10 (N3778) from Japan. N3711 was superseded by N3778 which recommends the same name as the US for these two characters.

d. The U.S. requests a change to the glyph of e-33F=U+1F61D CONFOUNDED FACE.
The glyph has an irregularly﹣shaped squiggle over the head of this ″confounded face″. It should be removed.
e. The U.S. also requests a glyph change to e-34B=U+1F62E CAT FACE KISSING.
The glyph shows two mouth shapes: A heart representing ″kissing″, and also a simple mouth shape. The mouth shape looks like an error and should be corrected.

Propose acceptance

See also comments T6 from Ireland and JP10 (N3778) from Japan. N3711 was superseded by N3778.These two changes are also included in N3778.

T.5. Emoji mapping data

a. The U.S. noted an error in N3728R, the mapping data for emoji. The character source emoji e﹣4EA REGIONAL INDICATOR SYMBOL LETTERS GB is currently mapped to U+1F1FA and U+1F1F0, which spell "UK." However, the ISO 3166 code is "GB," so the mapping data needs to be changed from:
1F1FA 1F1F0;;F3D1;FBB0
to:
1F1EC 1F1E7;;F3D1;FBB0

Propose acceptance

b. We propose adding the following text to the header of the EmojiSrc.txt file, N3728R:
This file provides mappings between UCS code points and sequences on one hand and Shift‑JIS codes for cell phone carrier symbols on the other hand. Each mapping is symmetric ("round trip"), for equivalent UCS and carrier symbols or sequences. This file does not include best‑fit ("fallback") mappings to similar but not equivalent symbols in either mapping direction.
Note: It is possible that future versions of this file will include additional data columns providing mappings for additional vendors.).
Propose acceptance in principle
This text could also be added as informative note in clause 26.4

T.6. CJK

a. The U.S. notes that 3 CJK sources are being mixed together in G_ GF[H]ZB. G_ GF[H]ZB is itself a union of ZhongHua ZiHai, Xiandai Hanyiu Cidian, and Ci-Hai. Xiandai Hanyiu Cidian and Ci-Hai already have their own source identifications, G_XC and G_CY, respectively. Instead of citing "G_ GF[H]ZB," the other two sources, G_XC and G_CY, should be cited.
Propose acceptance
No new source should be added if they contains characters from a source already represented.

b. The U.S. further strongly requests that all sources should have a numeric value, and this information be provided as soon as possible to the Project Editor. Although many G sources still allow a non-numeric value for the source (e.g., GBK, GCH, GFZ, GHZ), this information is nearly useless. Some sources with numeric values are easily available, such as Hanyu Dazidian.
WG2 discussion
This information was already requested at the last WG2 meeting by the editor.

Editorial comments:

E.1. Malayalam

The U.S. requests an annotation be added to 0D4E MALAYALAM LETTER DOT REPH: "not used in reformed modern Malayalam orthography."
Rationale: This annotation will make it clear that the letter is not used in the reformed orthographical context.
Propose acceptance

E.2. Emoticons

The U.S. recommends adding an annotation for e-814=U+1F3B6 MULTIPLE MUSICAL NOTES "x beamed eighth notes 266B".
Propose acceptance

附件 4

N3793：Draft disposition for fcd10646

Universal Multiple-Octet Coded Character Set
(UCS)

ISO/IEC JTC 1/SC 2 N ______
ISO/IEC JTC 1/SC 2/WG 2 N 3804
Date: 2010-04-23

Source: WG 2 meeting 56, San Jose, CA, USA; 2010-04-19/23
Title: Resolutions of WG 2 meeting 56
Action: For approval by SC 2 and for information to WG 2
Status: Adopted at meeting 56 of WG 2
Distribution: ISO/IEC JTC 1/SC 2 and WG 2

Experts from Canada, China, Ireland, Japan, Korea (Republic of), SEI - UC Berkeley (Liaison), TCA (Liaison), the Unicode Consortium (Liaison) and USA were present when the following resolutions were adopted (see attached attendance list).

---

*Character count **107517** (till end of Amd. 7)*
*Addition of **1864** in FPDAM8*
*Character count **109381** in CD of 2nd edition (till end of FPDAM8)*
Total count: 109381 (before meeting M56)
*Addition of **323** for WD of future.*

RESOLUTION M56.01 (Emoji ad hoc report):

Unanimous

WG2 accepts the recommendations in the ad hoc report in document N3829 dealing with various ballot comments on FPDAM8 and supporting documents related to Emoji symbols. The recommendations include:

- Addition of 1 new character 1F610 NEUTRAL FACE
- Changes to several glyphs, changes to character names and reordering of the characters inside the Emoticons block (the results are consolidated in document N3826)
- Changes to mappings to Emoji sources (to be incorporated in the revised EmojiSrc.txt - document N3835), and
- Accepting the following 13 Emoticon symbols for encoding in a future version of the standard:

1F600	GRINNING FACE
1F611	EXPRESSIONLESS FACE
1F615	CONFUSED FACE
1F617	KISSING FACE
1F619	KISSING FACE WITH SMILING EYES
1F61B	FACE WITH STUCK-OUT TONGUE
1F61F	WORRIED FACE
1F626	FROWNING FACE WITH OPEN MOUTH
1F627	ANGUISHED FACE
1F62C	GRIMACING FACE
1F62E	FACE WITH OPEN MOUTH
1F62F	HUSHED FACE
1F634	SLEEPING FACE

with their glyphs as shown in document N3834.

count:=1864+1 = 1865 in Amd. 8
count: 323+13=336 for future

RESOLUTION M56.02 (Character deletions in Amd. 8):

Unanimous

WG2 notes moving the following three characters from Amendment 8 to the collection of characters accepted for a future amendment:

- A78F LATIN LETTER MIDDLE DOT
- 1BFA BATAK SYMBOL BINDU GODANG
- 1BFB BATAK SYMBOL BINDU PINARJOLMA

count: 1865-3 = 1862 in Amd. 8
count: 336+3=339 for future.

RESOLUTION M56.03 (Progression of Amendment 8):

Unanimous

WG2 notes and accepts various name changes, glyph changes, annotation changes and changes to code positions of characters resulting from the disposition of FPDAM8 ballot comments, which incorporates results of resolutions M56.01 and M56.02 above. WG2 instructs its project editor to forward the final text of Amendment 8, along with the final disposition of comments document N3828 to the SC2 secretariat for an FDAM ballot. The final set of charts and names list are in document N3838. The unchanged target starting date is FDAM 2010-06.

count: addition of 1862 in Amd. 8

RESOLUTION M56.04 (CJK Charts for FCD of 2nd Edition):

Unanimous

WG2 decides to revert back to the single column format for CJK Unified Ideographs of Extension B for the 2nd edition. WG2 further instructs its editor and the IRG to review, verify and finalize the multiple-column charts for CJK Unified Ideographs in blocks other than Extension B.

RESOLUTION M56.05 (Revised text for examples Annex S.1.5):

Unanimous

IRG is instructed to review and propose updates to the revision to the examples proposed by the editor in document N3794. If an agreed upon text is made available to the editor prior to the start of the FDIS ballot, it will be included in the FDIS ballot.

RESOLUTION M56.06 (Disposition of Comments on FCD of 2nd Edition):

Unanimous

WG2 accepts the final disposition of ballot comments in document N3827, reflecting resolution M56.04 above. WG2 also accepts the following corrections proposed by the editor:

- update the publication date in reference for HKSCS to 2008 (page 39 of FCD text)
- Add the missing TC-4A76 reference to 94C4, including corresponding chart update
- Keep the CC-Data Element as a normative term in clause 4.5, but replace it with ‘code unit sequence’ in the remainder of the text of the standard. The term ‘code unit sequence’ will also be inserted in clause 4.5.

RESOLUTION M56.07 (Progression of 2nd Edition):

Unanimous

WG2 instructs its editor to forward updated final text and charts (after the IRG has verified the CJK charts), incorporating the results of resolutions M56.03 to M56.06 above, and the final disposition of ballot comments in document N3827, to SC2 secretariat for an FDIS ballot. The revised target starting date for FDIS is 2010-09.

count: 109379 in 2nd edition.

RESOLUTION M56.08 (Miscellaneous character additions):

Unanimous

WG2 accepts the following additions and corrections for future encoding in the standard:

a. 0604 ARABIC SIGN SAMVAT
   with its glyph shown in section 4 of document N3734.
b. Six punctuation characters:
   2E32 TURNED COMMA
   2E35 TURNED SEMICOLON
   2E36 DAGGER WITH LEFT GUARD
   2E37 DAGGER WITH RIGHT GUARD
   2E38 TURNED DAGGER
   2E39 TOP HALF SECTION SIGN
   with their glyphs and annotations from page 1 of document N3740.
c. Two mathematical symbols:
   27CB MATHEMATICAL RISING DIAGONAL
   27CD MATHEMATICAL FALLING DIAGONAL
   with their glyphs and annotations from pages 9 and 10 of document N3763.

d. Nine Cyrillic characters for Slavonic:
   A674 COMBINING CYRILLIC LETTER UKRAINIAN IE
   A675 COMBINING CYRILLIC LETTER I
   A676 COMBINING CYRILLIC LETTER YI
   A677 COMBINING CYRILLIC LETTER U
   A678 COMBINING CYRILLIC LETTER HARD SIGN
   A679 COMBINING CYRILLIC LETTER YERU
   A67A COMBINING CYRILLIC LETTER SOFT SIGN
   A67B COMBINING CYRILLIC LETTER OMEGA
   A69F COMBINING CYRILLIC LETTER IOTIFIED E
   with their glyphs and annotations from section 2 in document N3748.
e. 0AF0 GUJARATI ABBREVIATION SIGN,
   with its glyph as shown on line 1 page 1 of document N3764.
f. Four Typikon symbols:
   1F540 CIRCLED CROSS POMMY
   1F541 CROSS POMMY WITH HALF-CIRCLE BELOW
   1F542 CROSS POMMY
   1F543 NOTCHED LEFT SEMICIRCLE WITH THREE DOTS
   with their glyphs from left column in Table 1 of document N3772.
g. 058F ARMENIAN DRAM SIGN:
   with its glyph as shown in document N3834.
h. Two dashes required for Chicago manual of Style
   2E3A TWO-EM DASH
   2E3B THREE-EM DASH
   with their glyphs and annotations based on document N3770.
i. Seven Georgian and Nuskhuri letters
   10C7 GEORGIAN CAPITAL LETTER YN
   10CD GEORGIAN CAPITAL LETTER AEN
   10FD GEORGIAN LETTER AEN
   10FE GEORGIAN LETTER HARD SIGN
   10FF GEORGIAN LETTER LABIAL SIGN
   2D27 GEORGIAN SMALL LETTER YN
   2D2D GEORGIAN SMALL LETTER AEN
   with their glyphs and annotations shown on page 1 in document N3775.
j. Four Arabic combining marks
   08F0 ARABIC OPEN FATHATAN
   08F1 ARABIC OPEN DAMMATAN
   08F2 ARABIC OPEN KASRATAN
   08F3 ARABIC SMALL HIGH WAW
   in a new block named Arabic Extended-A with the range 08A0 to 08FF
   with their glyphs as shown in document N3791.
k. Correct the names for three Sundanese Supplement characters, based on document N3836:
   1CC4 change KA to LEU in the name
   1CC5 change BA to KA in the name, and
   1CC7 change LEU to BA in the name.
l. Two new compatibility ideograph characters, as requested in Korean comment T32.3 on FCD ballot, and in Appendix 3 of document N3747:
   FA2E with its glyph shown on page 10, corresponding to K0-522B
   FA2F with its glyph shown on page 09, corresponding to K0-6766

count: 339 + 39 = 378 for future.

RESOLUTION M56.09 (Takri script):

WG2 accepts to create a new block named Takri in the range 11680 to 116CF in the SMP, and to populate it with **66** characters in code positions 11680 to 116B7 and 116C0 to 116C9, with their code positions, names and glyphs as shown on pages 10 and 11 of document N3758. This block contains some combining marks.

count: 378 + 66 = 444 for future.

RESOLUTION M56.10 (Miao script):

Unanimous

WG2 accepts to create a new block named Miao in the range 16F00 to 16F8F in the SMP, and populate it with **128** characters in code positions 16F00 to 16F41 and 16F50 to 16F8D, with their code positions, names and glyphs as shown on pages 6 to 8 of document N3789. This block contains some combining marks.

count: 444 + 128 = 572 for future.

RESOLUTION M56.11 (Arabic Mathematical Alphabetic Symbols):

Unanimous

WG2 accepts to create a new block named Arabic Mathematical Alphabetic Symbols in the range 1EE00 to 1EEFF in the SMP, and populate it with **143** characters with their code positions, names and glyphs as shown in the charts in document N3799.

count: 572 + 143 = 715 for future.

RESOLUTION M56.12 (Subdivision of work):

Unanimous

WG2 instructs its convener and project editor to create a subdivision proposal (document N3837) for creation of the 3rd edition of ISO/IEC 10646, to incorporate items accepted for future encoding at meeting 55 (resolution M55.28 in document N3704) and characters and scripts accepted in resolutions M56.08 to M56.11 above. The 3rd edition is also to include the multiple column charts for the remaining CJK ideographs. The target starting dates are: CD 2010-06, FCD 2010-12 and FDIS 2011-11.

RESOLUTION M56.13 (Fonts for CJK multiple column charts):

Unanimous

WG2 instructs the IRG to request the IRG members to send to the project editor the necessary fonts to enable creating the multiple column charts for CJK Ext. B and for CJK compatibility ideographs for consideration to include in the 3rd edition of ISO/IEC 10646.

RESOLUTION M56.14 (Multiple column CJK charts for 3rd edition):

Unanimous

WG2 instructs its project editor to prepare the CJK multiple column charts for CJK Ext. B and CJK compatibility ideographs using the fonts received from IRG members and send them to the IRG for review, verification and finalization.

RESOLUTION M56.15 (CD for 3rd edition):

Unanimous

WG2 instructs its project editor to create the text of for the 3rd edition of ISO/IEC 10646, incorporating the characters accepted for encoding per resolution M56.12 above on subdivision of work, and send it to SC2 secretariat for a CD ballot.

count: 109379 + 715 = 110094 in CD of 3rd edition.

RESOLUTION M56.16 (Liaison response to SC34):

Unanimous

WG2 instructs its convener to prepare a response to the Liaison letter from ISO/IEC JTC1/SC34 in document N3809, and forward it to SC2 to communicate to SC34. The response will state WG 2's views on the issue that is similar to the response to comment JP.G7 in the disposition of FPDAM8 ballot comment in document N3828.

RESOLUTION M56.17 (Tangut script):

Unanimous

WG2 accepts the recommendations in the Tangut ad hoc report in document N3833 and invites the authors of the Tangut proposal in document N3797, together with other Tangut experts, to submit a revised proposal taking into account the ad hoc recommendations.

RESOLUTION M56.18 (Jurchen script):

Unanimous

WG2 invites the authors of the proposal to encode Jurchen script in document N3788 to further revise the document based on feedback in document N3817, discussion at meeting 56, and in consultation with other Jurchen experts.

RESOLUTION M56.19 (Disunification request from China and TCA):

Unanimous

With reference to the disunification request in document N3787, China and TCA are invited to revise the document clarifying what exactly is requested, have it reviewed by IRG, and submit the revised document to the next WG2 meeting.

RESOLUTION M56.20 (IRG Principles and Procedures):

Unanimous

WG2 invites experts from national bodies and liaison organizations to review and provide feedback to the IRG rapporteur, on the principles and procedures document of the Ideographic Rapporteur's Group in document N3744 and the associated Ideographic Working Document Series from document N3746 before the June meeting of IRG.

RESOLUTION M56.21 (IVD registration by Japan):

Unanimous

WG2 invites national bodies and liaison organizations to review and provide feedback to the IVD registrar by 2010-06-25 on the first set of Ideographic Variation Sequences for Hanyo-Denshi ideographic variation collection as per announcement in document N3796.

RESOLUTION M56.22 (Guideline on spelling for English character names):

Unanimous

WG2 accepts the proposed text for guideline proposed in document N3832 on spelling of English character names in the standard for inclusion in the Principles and Procedures document, and instructs Dr. Umamaheswaran to prepare an updated Principles and Procedures document (N3802) and submit it to the convener to post to the WG2 web site and communicate to SC2 secretariat.

RESOLUTION M56.23 (Roadmap snapshot):

Unanimous

WG2 instructs its convener to post the updated snapshot of the roadmaps (in document N3807) to the WG2 web site and communicate the same to SC2 secretariat

RESOLUTION M56.24 (Future meetings):

Unanimous

WG2 endorses the following schedule for future meetings of WG2 and of IRG:
WG2 meetings:
  Meeting 57 - 2010-10-04/08, Busan, Korea (Republic of)
  Meeting 58 - 2011-06-06/10, Helsinki, Finland (along with SC2 plenary)
  Meeting 59 – 1st Quarter, 2012, Mountain View, CA, USA (pending confirmation), Germany (as backup)
IRG meetings:
  IRG 35 - 2010-11-08/12, Macao SAR

RESOLUTION M56.25 (Appreciation to DKUUG for web site support):

By Acclamation

WG2 thanks DKUUG and its staff for its continued support of the web site for WG2 document distribution and the e-mail server.

RESOLUTION M56.26 (Appreciation to Host):

By Acclamation

WG2 thanks the US national body, ANSI, for hosting the meeting, Adobe Systems Inc. and its staff, in particular Mr. Eric Muller for providing excellent meeting facilities. WG2 thanks the Unicode Consortium, specifically Ms. Magda Danish, for its kind hospitality.

---

*Character count **107517** (till end of Amd. 7)*
*Addition of **1862** in FDAM8*
*Character count **109379** in FDIS of 2nd edition (till end of FDAM8)*
*Addition of **715** in CD 3rd edition*
*Character count **110094** (total in 3rd edition)*

Attendance List

The following **34** attendees representing **6** national bodies, **3** liaison organizations, including **3** invited experts were present at different times during the meeting. There were **6** additional experts who participated remotely.

Name	**Representing**	**Affiliation**
Mike KSAR	.Convener, USA	Independent
Erich FICKLE	.Invited Expert, China	Independent
Tom BISHOP	.Invited Expert, liaison SEI	Wenlin Institute, Inc.
Van ANDERSON	.Invited Expert, liaison SEI	Independent
LU Qin	.IRG Rapporteur	Hong Kong Polytechnic University
Bear S. TSENG	.TCA – Liaison	Academia Sinica
Hsin-Kuang CHEN	.TCA – Liaison	Bureau of Standards, Metrology and Inspection,
Lin-Mei WEI	.TCA – Liaison	Chinese Foundation for Digitization Technology
Suh-Chyin CHUANG	.TCA – Liaison	Bureau of Standards, Metrology and Inspection,
Alain LABONTÉ	Canada; Editor 14651	Independent
V. S. (Uma) UMAMAHESWARAN	Canada; Recording Secretary	IBM Canada Limited
CHEN Zhuang	China	Chinese Electronics Standardization Institute
JING Yongshi	China	North University of Ethnics, Ning Xia
LI Guoying	China	Beijing Normal Universlity
SUN Bojun	China	Institute of Ethnology and Anthropology, Chinese Academy of Scences
Michael EVERSON	Ireland; Contributing Editor	Evertype
Masahiro SEKIGUCHI	Japan	Fujitsu Limited
Yoshiki MIKAMI	Japan	Nagaoka University of Technology
KIM Kyongsok	Korea (Republic of)	Busan National University
Craig CUMMINGS	USA	Yahoo! Inc.
Eric MULLER	USA	Adobe Systems Inc.
Katsuhiko MOMOI	USA	Google
Lisa MOORE	USA	IBM Corp.
Mark DAVIS	USA	Google
Markus SCHERER	USA	Google
Peter EDBERG	USA	Apple Computer Inc.
Richard COOK	USA	Univ. of California, Berkeley
Rick McGOWAN	USA	Unicode Consortium
Roozbeh POURNADER	USA	High Tech Passport Ltd.
Yasuo KIDA	USA	Apple Inc.
Ken WHISTLER	USA; Contributing Editor	Sybase Inc.
Michel SUIGNARD	USA; Project Editor	Independent
Deborah ANDERSON	USA; SEI, UC Berkeley – Liaison	Dept. of Linguistics, Univ., of California, Berkeley
Peter CONSTABLE	USA; Unicode Consortium – Liaison	Microsoft Corporation

附件 5

N3804：Resolutions Meeting 56

Universal Multiple-Octet Coded Character Set
(UCS)

ISO/IEC JTC 1/SC 2 N ________
ISO/IEC JTC 1/SC 2/WG 2 N 3804
Date: 2010-04-23

Source:	WG 2 meeting 56, San Jose, CA, USA; 2010-04-19/23
Title:	Resolutions of WG 2 meeting 56
Action:	For approval by SC 2 and for information to WG 2
Status:	Adopted at meeting 56 of WG 2
Distribution:	ISO/IEC JTC 1/SC 2 and WG 2

Experts from Canada, China, Ireland, Japan, Korea (Republic of), SEI - UC Berkeley (Liaison), TCA (Liaison), the Unicode Consortium (Liaison) and USA were present when the following resolutions were adopted (see attached attendance list).

Character count ***107517*** *(till end of Amd. 7)*
Addition of ***1864*** *in FPDAM8*
Character count ***109381*** *in CD of 2nd edition (till end of FPDAM8)*
Total count: 109381 (before meeting M56)
Addition of ***323*** *for WD of future.*

RESOLUTION M56.01 (Emoji ad hoc report):

Unanimous

WG2 accepts the recommendations in the ad hoc report in document N3829 dealing with various ballot comments on FPDAM8 and supporting documents related to Emoji symbols. The recommendations include:

- Addition of 1 new character 1F610 NEUTRAL FACE
- Changes to several glyphs, changes to character names and reordering of the characters inside the Emoticons block (the results are consolidated in document N3826)
- Changes to mappings to Emoji sources (to be incorporated in the revised EmojiSrc.txt - document N3835), and
- Accepting the following 13 Emoticon symbols for encoding in a future version of the standard:

1F600	GRINNING FACE
1F611	EXPRESSIONLESS FACE
1F615	CONFUSED FACE
1F617	KISSING FACE
1F619	KISSING FACE WITH SMILING EYES
1F61B	FACE WITH STUCK-OUT TONGUE
1F61F	WORRIED FACE
1F626	FROWNING FACE WITH OPEN MOUTH
1F627	ANGUISHED FACE
1F62C	GRIMACING FACE
1F62E	FACE WITH OPEN MOUTH
1F62F	HUSHED FACE
1F634	SLEEPING FACE

with their glyphs as shown in document N3834.

count:=1864+1 = 1865 in Amd. 8
count: 323+13=336 for future

RESOLUTION M56.02 (Character deletions in Amd. 8):

Unanimous

WG2 notes moving the following three characters from Amendment 8 to the collection of characters accepted for a future amendment:

- A78F LATIN LETTER MIDDLE DOT
- 1BFA BATAK SYMBOL BINDU GODANG
- 1BFB BATAK SYMBOL BINDU PINARJOLMA

count: 1865-3 = 1862 in Amd. 8
count: 336+3=339 for future.

RESOLUTION M56.03 (Progression of Amendment 8):

Unanimous

WG2 notes and accepts various name changes, glyph changes, annotation changes and changes to code positions of characters resulting from the disposition of FPDAM8 ballot comments, which incorporates results of resolutions M56.01 and M56.02 above. WG2 instructs its project editor to forward the final text of Amendment 8, along with the final disposition of comments document N3828 to the SC2 secretariat for an FDAM ballot. The final set of charts and names list are in document N3838. The unchanged target starting date is FDAM 2010-06.

count: addition of 1862 in Amd. 8

RESOLUTION M56.04 (CJK Charts for FCD of 2nd Edition):

Unanimous

WG2 decides to revert back to the single column format for CJK Unified Ideographs of Extension B for the 2nd edition. WG2 further instructs its editor and the IRG to review, verify and finalize the multiple-column charts for CJK Unified Ideographs in blocks other than Extension B.

RESOLUTION M56.05 (Revised text for examples Annex S.1.5):

Unanimous

IRG is instructed to review and propose updates to the revision to the examples proposed by the editor in document N3794. If an agreed upon text is made available to the editor prior to the start of the FDIS ballot, it will be included in the FDIS ballot.

RESOLUTION M56.06 (Disposition of Comments on FCD of 2nd Edition):

Unanimous

WG2 accepts the final disposition of ballot comments in document N3827, reflecting resolution M56.04 above. WG2 also accepts the following corrections proposed by the editor:

- update the publication date in reference for HKSCS to 2008 (page 39 of FCD text)
- Add the missing TC-4A76 reference to 94C4, including corresponding chart update
- Keep the CC-Data Element as a normative term in clause 4.5, but replace it with ‘code unit sequence’ in the remainder of the text of the standard. The term ‘code unit sequence’ will also be inserted in clause 4.5.

RESOLUTION M56.07 (Progression of 2nd Edition):

Unanimous

WG2 instructs its editor to forward updated final text and charts (after the IRG has verified the CJK charts), incorporating the results of resolutions M56.03 to M56.06 above, and the final disposition of ballot comments in document N3827, to SC2 secretariat for an FDIS ballot. The revised target starting date for FDIS is 2010-09.

count: 109379 in 2nd edition.

RESOLUTION M56.08 (Miscellaneous character additions):

Unanimous

WG2 accepts the following additions and corrections for future encoding in the standard:

a. 0604 ARABIC SIGN SAMVAT
   with its glyph shown in section 4 of document N3734.
b. Six punctuation characters:
   2E32 TURNED COMMA
   2E35 TURNED SEMICOLON
   2E36 DAGGER WITH LEFT GUARD
   2E37 DAGGER WITH RIGHT GUARD
   2E38 TURNED DAGGER
   2E39 TOP HALF SECTION SIGN
   with their glyphs and annotations from page 1 of document N3740.
c. Two mathematical symbols:
   27CB MATHEMATICAL RISING DIAGONAL
   27CD MATHEMATICAL FALLING DIAGONAL
   with their glyphs and annotations from pages 9 and 10 of document N3763.

d. Nine Cyrillic characters for Slavonic:
   A674 COMBINING CYRILLIC LETTER UKRAINIAN IE
   A675 COMBINING CYRILLIC LETTER I
   A676 COMBINING CYRILLIC LETTER YI
   A677 COMBINING CYRILLIC LETTER U
   A678 COMBINING CYRILLIC LETTER HARD SIGN
   A679 COMBINING CYRILLIC LETTER YERU
   A67A COMBINING CYRILLIC LETTER SOFT SIGN
   A67B COMBINING CYRILLIC LETTER OMEGA
   A69F COMBINING CYRILLIC LETTER IOTIFIED E
   with their glyphs and annotations from section 2 in document N3748.
e. 0AF0 GUJARATI ABBREVIATION SIGN,
   with its glyph as shown on line 1 page 1 of document N3764.
f. Four Typikon symbols:
   1F540 CIRCLED CROSS POMMY
   1F541 CROSS POMMY WITH HALF-CIRCLE BELOW
   1F542 CROSS POMMY
   1F543 NOTCHED LEFT SEMICIRCLE WITH THREE DOTS
   with their glyphs from left column in Table 1 of document N3772.
g. 058F ARMENIAN DRAM SIGN:
   with its glyph as shown in document N3834.
h. Two dashes required for Chicago manual of Style
   2E3A TWO-EM DASH
   2E3B THREE-EM DASH
   with their glyphs and annotations based on document N3770.
i. Seven Georgian and Nuskhuri letters
   10C7 GEORGIAN CAPITAL LETTER YN
   10CD GEORGIAN CAPITAL LETTER AEN
   10FD GEORGIAN LETTER AEN
   10FE GEORGIAN LETTER HARD SIGN
   10FF GEORGIAN LETTER LABIAL SIGN
   2D27 GEORGIAN SMALL LETTER YN
   2D2D GEORGIAN SMALL LETTER AEN
   with their glyphs and annotations shown on page 1 in document N3775.
j. Four Arabic combining marks
   08F0 ARABIC OPEN FATHATAN
   08F1 ARABIC OPEN DAMMATAN
   08F2 ARABIC OPEN KASRATAN
   08F3 ARABIC SMALL HIGH WAW
   in a new block named Arabic Extended-A with the range 08A0 to 08FF
   with their glyphs as shown in document N3791.
k. Correct the names for three Sundanese Supplement characters, based on document N3836:
   1CC4 change KA to LEU in the name
   1CC5 change BA to KA in the name, and
   1CC7 change LEU to BA in the name.
l. Two new compatibility ideograph characters, as requested in Korean comment T32.3 on FCD ballot, and in Appendix 3 of document N3747:
   FA2E with its glyph shown on page 10, corresponding to K0-522B
   FA2F with its glyph shown on page 09, corresponding to K0-6766

count: 339 + 39 = 378 for future.

RESOLUTION M56.09 (Takri script):

WG2 accepts to create a new block named Takri in the range 11680 to 116CF in the SMP, and to populate it with **66** characters in code positions 11680 to 116B7 and 116C0 to 116C9, with their code positions, names and glyphs as shown on pages 10 and 11 of document N3758. This block contains some combining marks.

count: 378 + 66 = 444 for future.

RESOLUTION M56.10 (Miao script):

Unanimous

WG2 accepts to create a new block named Miao in the range 16F00 to 16F8F in the SMP, and populate it with **128** characters in code positions 16F00 to 16F41 and 16F50 to 16F8D, with their code positions, names and glyphs as shown on pages 6 to 8 of document N3789. This block contains some combining marks.

count: 444 + 128 = 572 for future.

RESOLUTION M56.11 (Arabic Mathematical Alphabetic Symbols):

Unanimous

WG2 accepts to create a new block named Arabic Mathematical Alphabetic Symbols in the range 1EE00 to 1EEFF in the SMP, and populate it with **143** characters with their code positions, names and glyphs as shown in the charts in document N3799.

count: 572 + 143 = 715 for future.

RESOLUTION M56.12 (Subdivision of work):

Unanimous

WG2 instructs its convener and project editor to create a subdivision proposal (document N3837) for creation of the 3rd edition of ISO/IEC 10646, to incorporate items accepted for future encoding at meeting 55 (resolution M55.28 in document N3704) and characters and scripts accepted in resolutions M56.08 to M56.11 above. The 3rd edition is also to include the multiple column charts for the remaining CJK ideographs. The target starting dates are: CD 2010-06, FCD 2010-12 and FDIS 2011-11.

RESOLUTION M56.13 (Fonts for CJK multiple column charts):

Unanimous

WG2 instructs the IRG to request the IRG members to send to the project editor the necessary fonts to enable creating the multiple column charts for CJK Ext. B and for CJK compatibility ideographs for consideration to include in the 3rd edition of ISO/IEC 10646.

RESOLUTION M56.14 (Multiple column CJK charts for 3rd edition):

Unanimous

WG2 instructs its project editor to prepare the CJK multiple column charts for CJK Ext. B and CJK compatibility ideographs using the fonts received from IRG members and send them to the IRG for review, verification and finalization.

RESOLUTION M56.15 (CD for 3rd edition):

Unanimous

WG2 instructs its project editor to create the text of for the 3rd edition of ISO/IEC 10646, incorporating the characters accepted for encoding per resolution M56.12 above on subdivision of work, and send it to SC2 secretariat for a CD ballot.

count: 109379 + 715 = 110094 in CD of 3rd edition.

RESOLUTION M56.16 (Liaison response to SC34):

Unanimous

WG2 instructs its convener to prepare a response to the Liaison letter from ISO/IEC JTC1/SC34 in document N3809, and forward it to SC2 to communicate to SC34. The response will state WG 2's views on the issue that is similar to the response to comment JP.G7 in the disposition of FPDAM8 ballot comment in document N3828.

RESOLUTION M56.17 (Tangut script):

Unanimous

WG2 accepts the recommendations in the Tangut ad hoc report in document N3833 and invites the authors of the Tangut proposal in document N3797, together with other Tangut experts, to submit a revised proposal taking into account the ad hoc recommendations.

RESOLUTION M56.18 (Jurchen script):

Unanimous

WG2 invites the authors of the proposal to encode Jurchen script in document N3788 to further revise the document based on feedback in document N3817, discussion at meeting 56, and in consultation with other Jurchen experts.

RESOLUTION M56.19 (Disunification request from China and TCA):

Unanimous

With reference to the disunification request in document N3787, China and TCA are invited to revise the document clarifying what exactly is requested, have it reviewed by IRG, and submit the revised document to the next WG2 meeting.

RESOLUTION M56.20 (IRG Principles and Procedures):

Unanimous

WG2 invites experts from national bodies and liaison organizations to review and provide feedback to the IRG rapporteur, on the principles and procedures document of the Ideographic Rapporteur's Group in document N3744 and the associated Ideographic Working Document Series from document N3746 before the June meeting of IRG.

RESOLUTION M56.21 (IVD registration by Japan):

Unanimous

WG2 invites national bodies and liaison organizations to review and provide feedback to the IVD registrar by 2010-06-25 on the first set of Ideographic Variation Sequences for Hanyo-Denshi ideographic variation collection as per announcement in document N3796.

RESOLUTION M56.22 (Guideline on spelling for English character names):

Unanimous

WG2 accepts the proposed text for guideline proposed in document N3832 on spelling of English character names in the standard for inclusion in the Principles and Procedures document, and instructs Dr. Umamaheswaran to prepare an updated Principles and Procedures document (N3802) and submit it to the convener to post to the WG2 web site and communicate to SC2 secretariat.

RESOLUTION M56.23 (Roadmap snapshot):

Unanimous

WG2 instructs its convener to post the updated snapshot of the roadmaps (in document N3807) to the WG2 web site and communicate the same to SC2 secretariat

RESOLUTION M56.24 (Future meetings):

Unanimous

WG2 endorses the following schedule for future meetings of WG2 and of IRG:
WG2 meetings:
    Meeting 57 - 2010-10-04/08, Busan, Korea (Republic of)
    Meeting 58 - 2011-06-06/10, Helsinki, Finland (along with SC2 plenary)
    Meeting 59 – 1st Quarter, 2012, Mountain View, CA, USA (pending confirmation), Germany (as backup)
IRG meetings:
    IRG 35 - 2010-11-08/12, Macao SAR

RESOLUTION M56.25 (Appreciation to DKUUG for web site support):

By Acclamation

WG2 thanks DKUUG and its staff for its continued support of the web site for WG2 document distribution and the e-mail server.

RESOLUTION M56.26 (Appreciation to Host):

By Acclamation

WG2 thanks the US national body, ANSI, for hosting the meeting, Adobe Systems Inc. and its staff, in particular Mr. Eric Muller for providing excellent meeting facilities. WG2 thanks the Unicode Consortium, specifically Ms. Magda Danish, for its kind hospitality.

---

*Character count **107517** (till end of Amd. 7)*
*Addition of **1862** in FDAM8*
*Character count **109379** in FDIS of 2nd edition (till end of FDAM8)*
*Addition of **715** in CD 3rd edition*
*Character count **110094** (total in 3rd edition)*

Attendance List

The following **34** attendees representing **6** national bodies, **3** liaison organizations, including **3** invited experts were present at different times during the meeting. There were **6** additional experts who participated remotely.

Name	Representing	Affiliation
Mike KSAR	.Convener, USA	Independent
Erich FICKLE	.Invited Expert, China	Independent
Tom BISHOP	.Invited Expert, liaison SEI	Wenlin Institute, Inc.
Van ANDERSON	.Invited Expert, liaison SEI	Independent
LU Qin	.IRG Rapporteur	Hong Kong Polytechnic University
Bear S. TSENG	.TCA – Liaison	Academia Sinica
Hsin-Kuang CHEN	.TCA – Liaison	Bureau of Standards, Metrology and Inspection,
Lin-Mei WEI	.TCA – Liaison	Chinese Foundation for Digitization Technology
Suh-Chyin CHUANG	.TCA – Liaison	Bureau of Standards, Metrology and Inspection,
Alain LABONTÉ	Canada; Editor 14651	Independent
V. S. (Uma) UMAMAHESWARAN	Canada; Recording Secretary	IBM Canada Limited
CHEN Zhuang	China	Chinese Electronics Standardization Institute
JING Yongshi	China	North University of Ethnics, Ning Xia
LI Guoying	China	Beijing Normal Universlity
SUN Bojun	China	Institute of Ethnology and Anthropology, Chinese Academy of Scences
Michael EVERSON	Ireland; Contributing Editor	Evertype
Masahiro SEKIGUCHI	Japan	Fujitsu Limited
Yoshiki MIKAMI	Japan	Nagaoka University of Technology
KIM Kyongsok	Korea (Republic of)	Busan National University
Craig CUMMINGS	USA	Yahoo! Inc.
Eric MULLER	USA	Adobe Systems Inc.
Katsuhiko MOMOI	USA	Google
Lisa MOORE	USA	IBM Corp.
Mark DAVIS	USA	Google
Markus SCHERER	USA	Google
Peter EDBERG	USA	Apple Computer Inc.
Richard COOK	USA	Univ. of California, Berkeley
Rick McGOWAN	USA	Unicode Consortium
Roozbeh POURNADER	USA	High Tech Passport Ltd.
Yasuo KIDA	USA	Apple Inc.
Ken WHISTLER	USA; Contributing Editor	Sybase Inc.
Michel SUIGNARD	USA; Project Editor	Independent
Deborah ANDERSON	USA; SEI, UC Berkeley – Liaison	Dept. of Linguistics, Univ., of California, Berkeley
Peter CONSTABLE	USA; Unicode Consortium – Liaison	Microsoft Corporation

附件 6

N3744：IRG P&P (IRGN 1646)

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 2/WG 2/IRG

Universal Coded Character Set
(UCS)

ISO/IEC JTC 1/SC 2/WG 2 N3744

/IRG *N1646*

2010-03-16

Title:	**IRG Principles and Procedures Version 5 Draft**
Source:	**IRG P&P Drafting Group**
Action:	**For review by the IRG and WG2**
Distribution:	**IRG Member Bodies and Ideographic Experts**
References:	**IRG N1602 (P&P Draft 4) and IRG N1633 (P&P Editorial Report)** **IRG N1601 (P&P Draft 3 Feedback from HKSAR)** **IRG N1590 and IRGN 1601(P&P V2 and V3 draft and all feedback)** **IRG N1562 (P&P V3 Draft 1 and Feedback from HKSAR)** **IRG N1561 (P&P V2 and all feedback)** **IRG N1559 (P&P V2 Draft and all feedback)** **IRG N1516 (P&P V1 Feedback from HKSAR)** **IRG N1489 (P&P V1 Feedback from Taichi Kawabata)** **IRG N1487 (P&P V1 Feedback from HKSAR)** **IRG N1465, IRG N1498 and IRG N1503 (P&P V1 drafts)**

Table of Contents

1. Introduction 3
1.1. Scope of IRG Work 3
1.2. Scope of This Document 3
2. Development of CJK Unified Ideographs 3
2.1. Principles on Identification of CJK Unified Ideographs 4
2.1.1. Principles on Encoding 4
2.1.2. Unification Procedures of CJK Ideographs...4
2.1.3. Non-cognate Rule...4
2.1.4. Maintaining Up-to-Date Unification/Non-unification Examples....................4
2.2. Principles on Submission of Ideographs to the IRG 4
2.2.1. Basic Rules on Submission 4
2.2.2. Required Font to be Submitted...5
2.2.3. Required Data to be Submitted 6
2.2.4. Required Evidence to be Submitted 6
2.2.5. Required Summary Form to be Submitted..6
2.2.6. Quality Assurance: The 5% Rule..6
2.3. Principles on Production of IRG Working Drafts 7
2.3.1. Principles on Submitted Ideographs...7
2.3.2. Principles on Assignment of Serial Numbers 7
2.3.3. Principles on Machine-checking of IDS of Submitted Ideographs...............7
2.3.4. Production of IRG Working Drafts 7
2.4. Principles on Reviewing IRG Working Drafts 8
2.4.1. General Principles on Reviews...8
2.4.2. Principles on Manual Checking 8
2.4.3. Submission of Possibly Unifiable Ideographs..8
2.5. Principles on Discussions at IRG Meetings 9
2.5.1. Document-based Discussion..9
2.5.2. Discussion Procedures...9
2.5.3. Recording of Discussions 9

2.5.4. Time and Quality Management..9
2.6. Principles on Submission of Ideographs to WG2 10
2.6.1. Checking of Stabilized M-Set..10
2.6.2. Preparation for WG2 Submission 10
3. Procedures 10
3.1. Call for Submission 10
3.2. Consolidation and Grouping of Submitted Ideographs 10
3.3. First Checking Stage 11
3.4. First Discussion and Conclusion Stage 11
3.5. Second Checking Stage 11
3.6. Second Consolidation and Conclusion Stage 11
3.7. Final Checking Stage 12
3.8. Approval and Submission to WG2 12
4. Guidelines for Comments and Resolutions on Working Sets...12
4.1. Guidelines for M Set 12
4.2 Guidelines for D Set 13
5. IRG Website..14
6. IRG Document Registration 14
6.1. Registration Procedures 14
6.2. Contact for IRG Document Registration 14
Annex A: Sorting Algorithm of Ideographs...15
Annex B: IDS Matching 17
B.1. Guidelines on Creation of IDS 17
B.2. Requirements on IDS Matching 17
B.3. Limitation of IDS Matching 17
Annex C: Urgently Needed Ideographs 18
C.1. Introduction 18
C.2. Requirements 18
C.3. Dealing with Urgent Requests 18
Annex D: Up-to-Date CJK Unified Ideograph Sources and Source References.................19
Annex E: Maintenance Procedure of the IRG Working Documents Series.........................21
E.1 Introduction 21
E.2. IRG Working Documents Series 21
E.3. Maintenance Procedure 21
Annex F: IRG Repertoire Submission Summary Form...23
Annex G: Examples of Unified CJK Submissions 25
G.1. Sample Data Files 25
G.2. Sample Evidence 25
G.3. Handling of Data with Privacy Concerns 26
[Annex H] Not Used at the Moment..28
Annex I: Guideline for Handling of CJK Ideograph Unification or Dis-unification Errors .29
I.1 Guideline for “To Be Unified” Errors 29
I.2 Guideline for “To Be Dis-unified” Errors 29
I.3 Discouragement of New Dis-unification Request 29
Annex J: Guideline for Correction of CJK Ideograph Mapping Table Errors......................30
J.1 Priority of Error Correction Procedure…………………………………………………………………..29
J.2 Announcement of Addition or Correction of Mapping Table……………………………………………29
J.3 Collection and Maintenance of Mapping Table that are not Owned by WG2………………………...29
References...31
Glossary...32

1. Introduction

This document is a standing document of the ISO/IEC JTC 1/SC 2/WG 2/IRG for the standardization of Chinese-Japanese-Korean (CJK) Unified Ideographs. It consists of a set of principles and procedures on a number of items relevant to the preparation, submission and development of repertoires of CJK Unified Ideographs extensions for additions to the standard (ISO/IEC 10646). Submitters should check the standard documents (including all the amendments and corrigenda) before preparing new submissions.

For anything not explicitly covered in this document, the IRG will follow the Principles and Procedures of WG2 and other higher level directives.

1.1. Scope of IRG Work

The IRG works on CJK ideograph-related tasks under the supervision of WG2 (SC2 Resolution M13-05). The following is a list of current and completed IRG projects:

a. CJK Unified Ideograph Repertoire and its extensions
b. Kangxi Radicals and CJK Radical Supplements
c. Ideographic Description Characters
d. International Ideographs Core (IICore)
e. CJK Strokes
f. Old Hanzi

Work on new IRG projects requires the approval of WG2 and preparation of documents for such approval is required before the IRG can officially launch these projects.

1.2. Scope of This Document

The following sections are dedicated for the standardization of CJK Unified Ideographs, describing the set of principles and procedures to be applied in the development of a new repertoire of CJK Unified Ideographs as specified under work item a. in Section 1.1.

This document does not cover other IRG work items listed in Section 1.1. Standardizing CJK Compatibility ideographs maintained in UCS for the purpose of round-trip integrity with other standards is out of IRG scope. However, CJK compatibility characters submitted to WG2 must be reviewed by the IRG to avoid potential problems. For handling mis-unification and duplicate ideographs, WG2 Principles and Procedures Annex I and J attached to this document should be referenced.

2. Development of CJK Unified Ideographs

Any new extension work must be approved by WG2 before the actual consolidation and review can be formally carried out. There are no fixed rules to initiate a new extension. Normally, some member bodies would first initiate it by submitting a proposal which states the need of a required repertoire. Submission of proposals must follow the principles and procedures stated in this document. The IRG would first review the proposal and confirm that it is within the IRG scope.

Taking into consideration of the urgency, the justification and the repertoire size in the proposal, and the current workload of the IRG, the IRG may take one of the following actions.

a. Endorse the proposal and submit it to WG2 for approval.
b. Request other member bodies to submit characters of similar nature so as to estimate the real workload before submitting to WG2 for endorsement.
c. Accept the proposal as a contribution to an ongoing IRG work item.
d. Reject the proposal with justifications. A rejected proposal may be revised and re-submitted to the IRG.

2.1. Principles on Identification of CJK Unified Ideographs

2.1.1. Principles on Encoding

Ideographs that have the same abstract shapes are unified under the unification rule (Annex S of ISO/IEC 10646) and assigned to a single character code. A CJK ideographic character can be represented in many actual forms depending on the writing style adopted. Examples of common writing styles include Song style and Ming style as typical print forms, Kai style as a hand written form, and Cao style as a cursive form. Stylistically different forms of the same character can be represented by a different number or different type of strokes or components, which may affect identification of the same abstract shape. In order to reach a common ground to identify those abstract shapes to be encoded as distinct CJK Unified Ideographs, the IRG only accepts submissions using a print form of glyphs (usually Song style or Ming style).

2.1.2. Unification Procedures of CJK Ideographs

Standard print forms of CJK ideographs are constructed with a combination of known components or stroke types. Many are determined by two components - a radical chosen to classify the character in dictionaries and possibly reflect the meaning of the character, and a phonetic component which represents the pronunciation of the character. Basically, two submitted print forms of glyphs with different radicals are distinct characters even if they have the same phonetic component such as shown in the example of '理'(U+7406) and '鯉'(U+9BC9). For non trivial cases, further shape analysis must be conducted. Two similar glyphs are decomposed into radicals, components or stroke types and evaluated by following the unification procedures described in Annex S of ISO/IEC 10646.

2.1.3. Non-cognate Rule

No matter how similar two ideographs are in actual shape, non-cognate or semantically different glyphs are considered to have different abstract shapes. The following gives examples of semantically different characters with very similar glyphs, they are considered to have different abstract shapes because they are non-cognate.

'戌'(U+620C) and '戍'(U+620D) differ only in rotated strokes or dots (S.1.5 a).

'于'(U+4E8E) and '干'(U+5E72) differ only in folding back at the stroke termination (S.1.5 f).

Because shape analysis alone may not tell non-cognateness or semantic differences, it is the submitter’s responsibility to provide information and supporting evidence in order to invoke the non-cognate rule.

2.1.4. Maintaining Up-to-Date Unification/Non-unification Examples

In Annex S, unification/non-unification examples are summarized from past practice (currently, only reflecting the first unification work in the CJK main block) and they are not exhaustive. If there is ambiguity in applying these rules, the IRG must first have a formal discussion for agreement. In case of finding worthy examples for recording, the IRG will maintain an up-to-date list of unification/non-unification examples by adding such examples. Furthermore, the list will be reported to WG2 from time to time as the input for Annex S revision.

2.2. Principles on Submission of Ideographs to the IRG

2.2.1. Basic Rules on Submission

A member body may submit the following to the IRG along with its repertoire. Different information may be handled differently as specified below.

a. **New Sources to StandardizedIdeographs**. If the submission specifies new sources (such as an existing or a new national standard) to some existing standards, it needs to be reviewed and approved by the IRG before submission to WG2. Sources and source references in the current ISO/IEC 10646 standard can be found in clause 27 of ISO/IEC 10646 First Edition (2003-12-15)(See Annex D for up-to-date IRG list of sources).

b. **New Sources to Working Sets**. In case there are some remaining characters in previous standardization stages, new sources reviewed and approved by the IRG will be incorporated into the up-to-date IRG list of sources for the current IRG working sets by the IRG technical editor.

c. **New Compatibility Ideographs.** In case a member body needs to add compatibility ideographs, these characters must be reviewed by the IRG before submission to WG2 to avoid potential problems of unification or dis-unification with other CJK characters.

d. **New Unified Ideographs.** All ideograph submissions are subject to the following rules:

(1). **Collection Size**: As the collection is defined by submitters according to their own criteria, the IRG will not impose a limit on the collection size. However, to rationalize the burden of the checking process and to achieve a higher quality of standard within a reasonably short period of time, the size of the collection to be reviewed by IRG member bodies normally cannot exceed 4,000 ideographs. Based on this principle, member bodies may be asked to divide its submitted collections into subsets to be processed in different IRG collections.

(2). **Pre-submission Unification Checking**: A member body should be EXTREMELY CAREFUL ***not to submit unified ideographs that are already standardized or previously discussed*** and recorded at IRG meetings. By the nature of ideographs, it is very difficult for reviewers to find out all unifiable ideographs. Thus, it is important to achieve high quality at the time of submission. Submitters must make sure that submitted ideographs do not fall into any of the following categories:

a) Ideographs already standardized in the ISO/IEC 10646 standard (including amendments).
b) Ideographs currently in WG2 working drafts (including PDAM (Proposed Draft Amendment), FPDAM (Final Proposed Draft Amendment) and FDAM (Final Draft Amendment)).
c) Ideographs currently in IRG working sets including both M-sets and D-sets.
d) Ideographs mis-unified or over-unified with ideographs in the current standard based on the list maintained by the IRG in its working document series, IRGWD_MUI and IRGWD_NUC, respectively.

Low quality submissions may be rejected by applying the “5% rule” described in Section 2.2.6 below.

(3). **Document Registration**: All submission documents should be registered as IRG N documents, whose file name should be in the form of:

IRG N*nnnn*_*mmmm*[_*sss*[_*ppp*]]_submission

where *nnnn* indicates a document number assigned by the IRG Rapporteur, *mmmm* indicates member body’s source ID (as listed in 2.2.3a), *sss* can be any member body designated indicator, and *ppp* indicates the working set or repertoire name (such as Ext. E labelled by “_E”).

(4). **Submission of Over-Unified or Mis-Unified Ideographs**: Submission of ideographs that are already mis-unified or over-unified within the current standard should follow the principles in Annex I of WG2 Principles and Procedures. The list of over-unified or mis-unified ideographs should be maintained by the IRG technical editor and made available for update in the IRG standing document series IRGWD_NUC and IRGWD_MUI according the maintenance procedures defined in Annex E of this document..

2.2.2. Required Font to be Submitted

a. ***Glyph Image*:** Each proposed ideograph must be accompanied by a corresponding 128 x 128 bitmap file in Song or Ming style. The file name should be the same as the Source ID (defined below in Section 2.2.3.) with .bmp as its file extension.

b. **TrueType Font** (optional): TrueType font availability is highly recommended although not necessary. Font specification can be found under point 5 of A.1. – Submitter’s Responsibilities in Annex A (url:http://std.dkuug.dk/jtc1/sc2/wg2/docs/principles.html). The IRG at certain stage of project development will set a deadline for TrueType font submission.

2.2.3. Required Data to be Submitted

The following data for each proposed ideograph must be submitted with CSV (Comma Separated Value) text format (in UTF-8) or Microsoft Excel format file:

a. ***Source ID*** to indicate the source and the name of the glyph image for tracking. The source ID should begin with a member body code (G, T, H, M, J, K, KP, MY, U or V)[this is the order in Section 4]followed by no more than 9 characters and should contain only Latin capital letters, Arabic numbers, and hyphens. The purpose of source references and accepted source references by ISO 10646 are exhaustively listed in Section 27 of ISO 10646. See Annex D for details on information about member body code.

b. ***Glyph Image file name.*** The glyph image file name of each glyph image must be the same as the source ID with file extension of .bmp in bitmap format.

c. ***KangXi Radical Code*** from 1(U+2F00) to 214(U+2FD5) with an additional 0 or 1 to indicate a traditional character or simplified character, respectively.

d. ***Stroke Count*** of the non-radical component (ref. IRG N 954 AR and IRG N1105).

e. ***Flag to show whether the ideograph is traditional (0) or simplified (1).***

f. ***Ideographic Description Sequence*** (ref. IRG N1183).

g. ***Similar Ideographs and Variant Ideographs*** if available (identified by their code points in the standard in the form of U+xxxxx) or enter “No” if no known variants, leave it empty if not checked.

h. ***References to evidence documents*** including document number and page number.

Some sample submissions are listed in Annex G for reference.

2.2.4. Required Evidence to be Submitted

a. **Supporting Evidence**: Evidence should be supplied to support the proposed glyph shape and the usage and context with pronunciations, meanings, etc., to convince the IRG that it is actually used or non-cognate with other similar ideographs. Evidence for each character must be supplied as scanned images. The provision of evidence on character usage including those for personal names should not be exempted. A declaration for character use without accompanying evidence is not acceptable. Considering privacy issues, the IRG has suggested some compromised provision. Details are given in Annex G Part 3.

b. **Questionable Characters** (optional): For candidate ideographs with possible unification questions, submitters are encouraged to provide detailed evidence of use from authoritative sources, and relationships to other standardized ideographs or variants having similar shape or meaning encoded in UCS for review.

c. **Avoidance of Derived Simplified Ideographs:** To avoid encoding derived simplified characters that are not in actual use, submission of simplified ideographs requires the actual usage evidence. Providing only their corresponding traditional ideographs will not be considered evidence.

2.2.5. Required Summary Form to be Submitted

Each submission for an ideograph collection should be accompanied by a duly completed “Proposal Summary Form for Additions of CJK Unified Ideographs to the Repertoire of ISO/IEC 10646” (see **Annex F**).

2.2.6. Quality Assurance: The 5% Rule

For any character encoding standard, a common general principle is to encode the same character once and only once. Before any submission, it is the submitter's responsibility to filter out the ideographs that are already in the ISO/IEC 10646 international coding standard:

- the published standard,
- any of its published amendments,
- any of its amendments under ballot in JTC1/SC2, or
- one of the working sets of the IRG.

In assessing the suitability of a proposed ideograph for encoding, the IRG will evaluate the credibility and quality of the submitter's proposal. If the IRG finds more than 5% of duplicated characters in the above mentioned collections from the submitter's source set during the IRG review process, the whole submission will be removed from the subsequent IRG working drafts for that particular IRG project.

2.3. Principles on Production of IRG Working Drafts

After the IRG accepts submissions based on principles specified in Section 2.2, the IRG technical editor will produce a set of IRG working drafts.

2.3.1. Principles on Submitted Ideographs

a. All the original ideograph submissions, including glyphs, IDS, radicals, stroke counts and evidence, must have registered IRG document numbers.
b. If any required information is missing, the IRG chief editor or technical editor can ask for additional information from the submitter. Without timely supply of such information, the submission may be rejected by the technical editor in producing a working draft.

2.3.2. Principles on Assignment of Serial Numbers

a. The IRG technical editor will consolidate and sort the submitted ideographs in accordance with Annex A of this document.
b. A unique *serial number* will be assigned to each submitted ideograph after consolidation.
   The serial numbers must be unique throughout the entire standardization process. They must not be changed, re-set or re-assigned unless a split happens. This principle allows easy reference to past discussions. In case of a split, one ideograph will keep the original serial number and a new serial number will be assigned to the split ideograph.
c. If ideographs submitted by different member bodies are obviously unifiable, such ideographs may be unified and assigned the same serial number by the IRG technical editor.

2.3.3. Principles on Machine-checking of IDS of Submitted Ideographs

a. The IRG technical editor will check the submitted IDS with existing IDS data to detect possible unifiable or duplicated ideographs.
b. Machine checking sometimes detects obviously non-unifiable pairs. Such cases, when detected, will be annotated before proceeding to the next stage.
c. IDS checking algorithm will satisfy the requirements described in Annex B.

2.3.4. Production of IRG Working Drafts

a. **Division of Character Subsets**: By the result of IDS checking, submitted ideographs will be grouped into the following two working sets:
   i. **M-set (main set)**: for ideographs with proper IDS, and found not to be unifiable with current standardized ideographs nor previously discussed ideographs with proper IDS.
   ii. **D-set (discussion set**): for ideographs with missing, incomplete, or inconclusive IDS, or ideographs of which the attribute data have been questioned by any member body during a review process, or ideographs that might be unifiable with standardized or previously discussed ideographs. Ideographs with missing or incomplete IDS will be commented as such, and checked intensively through manual checking. Ideographs that might be unifiable with standardized or previously discussed ideographs will also be commented as such, and their suitability for unification must be manually checked and supported by evidence for dis-unification.
b. **Naming of Working Drafts**: The file name should follow the format of “IRGN*nnnn*V*X*[*XXX*]” where *nnnn* is the IRG assigned document number and *X* is the version number. No space is allowed but use of underscore “_” for separation is allowed. Examples of version numbers are “ExtEV1.0”, V1.0Draft”, etc.
c. **Glyph Images**: Archive of consolidated glyph images whose image size should be 128x128 with file name using the Source ID with the extension .bmp.
d. **Addition of Characters**: No ideographs should be added to the working set once the development process begins.
e. **Alteration of Characters:** Generally speaking, alteration of characters indicates instability and any change may also have impact on other characters in the collection. Thus it is generally not allowed. However, member bodies may submit minor alteration of characters with provision of justification ONLY at the final stage as long as the alteration is unifiable with the original

character. Change of glyph beyond the Annex S unification criteria is considered to be an addition of new character and is NOT acceptable at this stage. The submitter must provide the results of thorough checks and verification that the alteration does not affect other characters in existing standards and working sets. The IRG, based on its evaluation, may decide to accept the alteration, reject the alteration or request the removal of such a character by the submitter. If the submitter finds that the glyph of a character is wrong at any working stages, the character will be rejected by the IRG and should be withdrawn by the submitter.

f. **Previous D-Set:** If a previously discussed D-set exists, new D-set ideographs should be merged with the previous D-set.
g. After consolidation, the IRG chief editor and technical editor may ask member bodies to review M-set and D-set based on IRG review schedule and task division.

2.4. Principles on Reviewing IRG Working Drafts

If the IRG instructs member bodies to review a working draft, member bodies' editors should review it (different portions may be assigned to different member bodies) according to the agreed schedule and they should follow the principles set out below during the review process.

2.4.1. General Principles on Reviews

a. Each member body should check the ideographs of the working sets assigned by the IRG chief editor and technical editor for the following issues:
    i. Correctness of KangXi radical and KangXi Index, Stroke Count, First Stroke and IDS.
    ii. Correctness and quality of glyphs and source information if necessary.
    iii. Any duplicate or unifiable ideographs based on Annex S guidelines.
    iv. Consistency of submitted characters with the submitted evidence and documentary proof.
b. When any data, including IDS, KangXi radical, or stroke count is found to be incorrect, such M-set ideograph should be moved to D-set as its standing data is no longer valid. Until such ideograph is assured to be unique by manual checking (procedures described in Section 2.4.2. below), it should not be moved back to M-set.

2.4.2. Principles on Manual Checking

a. **Duplication and Unification:** For D-set ideographs, member bodies should ensure that they are not duplicates of or unified with any ideographs in the standard or in another working set (including the current one).
b. **Radical Checking:** Assurance is done by enumerating all possible radicals of a target ideograph and looking for any duplicate or unifiable ideographs in the range of ±2 stroke counts of standardized and working set ideographs. For example, “聞” may have the radical of “門” with 6 strokes, or the radical of “耳” with 8 strokes. In such a case, checking standardized and working set ideographs with radical of “門” and 4-8 strokes, or ideographs with radical of “耳” and strokes of 6-10 manually can have much better assurance that such an ideograph does not have duplicate or unifiable ideographs.
c. **Recording of Review Results**: After reviewing, the reviewer should record the comment of “Checked against all standardized and working set ideographs with radical *X* and stroke count of *Y*±2.”

2.4.3. Submission of Possibly Unifiable Ideographs

a. **Preparation of Comments**: Member bodies should prepare comments and feedback with reference to the assigned serial number of the ideograph in question. The guidelines on comments are described in Section 4 of this document. Comment files should be in CSV form as a text file or a Microsoft Excel format file. All comment files must have pre-assigned IRG document numbers.
b. **Additional Evidence and Arguments**: For each proposed ideograph in the D-set that has been questioned for possible unification, the submitter should prepare arguments with further evidence of its use and further evidence (for example, from dictionaries, legal documents or other publications) showing that it is not unifiable with another standardized ideograph or an ideograph proposed in the same or another working draft.

c. **Submission Deadline:** Each member body should send feedback comments at least two months before the next IRG meeting. The IRG chief editor and technical editor will consolidate them and register the result as IRG N documents a month before the next IRG meeting so that each member body can examine the comments and prepare any additional documents for discussion at the meeting.

d. **Rejection**: Questioned ideographs with no counter arguments in support of dis-unification supplied to the meeting will be automatically marked as unified.

2.5. Principles on Discussions at IRG Meetings

2.5.1. Document-based Discussion

For efficient and smooth work, all discussion items and evidence must be prepared with registered IRG documents before the commencement of an IRG meeting. Items or evidence that are not contained in an IRG registered document are not treated as evidence and will not be discussed during IRG meetings. Any discussions on evidence or items raised after the commencement of an IRG meeting may be postponed to the next IRG meeting if any member body requests longer time to examine such items or evidence.

2.5.2. Discussion Procedures

Discussion will be based on the review comments on working sets. For non-unification issues, a submitter should present evidence document(s) showing that suspected unifiable ideographs are distinctively used as non-cognate character in the same region, or that these two characters cannot be unified in accordance with Annex S. When IRG member bodies have consensus that the ideographs are unifiable, the submitter should take one of the following actions, and the decision must be recorded.

a. Withdraw the duplicate ideograph and map the character in question to the existing standardized or working set ideograph.

b. Submit it as a compatibility ideograph character.

c. Add a new source reference to the existing standardized or working set ideograph.

When characters are reviewed by different people, different choices of radical, stroke count or first stroke code are possible for the same ideograph. IRG member bodies should resolve to agree on the most appropriate one based on the commonest abstract shape of the specific glyph. When KangXi radical or stroke count is found to be incorrect, the ideographs will be moved to D-set and wait for another manual review to prevent any unification error caused by not having conducted the review with ideographs having the correct KangXi radical or stroke count.

Guidelines on typical comments and resolutions are given in Section 4 of this document.

2.5.3. Recording of Discussions

Comments, rationales, and decisions must be recorded for each ideograph reviewed in a tabular format for reference and checking.

2.5.4. Time and Quality Management

Before discussion begins, the number of ideographs under review will be counted and the estimated schedule will be determined based on it. During the discussion, the number of comments reviewed per hour will be noted and the schedule will be adjusted by this rate (Note: It is recognized that some comments may take longer than others to discuss and resolve). If the comments cannot be handled in one IRG meeting, they may be partitioned and resolved in subsequent IRG meetings. Due to the limited time CJK Editorial Group has to deal with individual characters during an IRG meeting, member bodies can use emails to discuss and reach agreement on simple, straightforward cases before and after an IRG meeting.

2.6. Principles on Submission of Ideographs to WG2

2.6.1. Checking of Stabilized M-Set

a. Once M-set is consolidated and stabilized, the ideographs in M-set will be checked intensively as a complete set at least once to ensure data and glyph integrity.
b. Approval by member bodies by majority is needed before the set can be prepared for WG2 submission.

2.6.2. Preparation for WG2 Submission

After the approval by majority of IRG member bodies, the IRG technical editor will prepare the proposal to be forwarded to WG2. The preparation includes the following:

a. Sort the final stable M-set ideographs by the sorting algorithm described in Annex A.
b. Assign provisional UCS code positions to the sorted M-set ideographs (with agreement from ISO 10646 project editor on block assignment).
c. Make available the TrueType fonts for each member body with assigned provisional UCS code positions (fonts have to be available in accordance with the requirement stated in point 5 of A.1. – Submitter's Responsibilities in Annex A, WG2 Principles and Procedures).
    i. Each submitter is encouraged to prepare and submit its own font for best font quality.
    ii. If a submitter has difficulty creating the font, other member bodies or the IRG technical editor may help creating the font. In this case, the glyph style of the submitter must be respected.
    iii. If the submitter cannot provide the TrueType font by this time, the collection by the submitter will be withdrawn from this working set.
d. Prepare a list of source references.
e. Produce a packed Multi-column Ideograph Chart using the TrueType fonts.

The IRG will conduct at least one round of review of the proposal and the chart generated using TrueType font before submission to WG2.

3. Procedures

This section describes the basic development procedures of CJK Unified Ideograph extensions. The ultimate purpose of the procedures outlined in this section is to realize the production of high quality CJK Unified Ideograph sets in an efficient manner.

Development procedures described in this section consists of 8 stages, and it may take two to three years to create a high quality ideograph set for standardization.

3.1. Call for Submission

a. When a member body requests a new project for CJK Unified Ideograph extension and when the project is agreed upon at an IRG meeting, the IRG may call for submission of new ideographs. The IRG will also determine the deadline for the submission.
b. Each member body with proposed ideographs must submit the ideographs before the specified deadline with required data described in Section 2 of this document.
c. Member bodies must check whether the submitted ideographs are accompanied with all required information. If some required information is missing or misplaced, the IRG chief editor or technical editor may ask the submitter to re-submit or supply the additional information if only minor problems are encountered. Otherwise, the submission may be rejected because consolidation with other member bodies' submissions cannot be carried out.

3.2. Consolidation and Grouping of Submitted Ideographs

Consolidation of submissions is normally done between IRG meetings. The consolidation includes the following tasks:

a. The IRG technical editor will sort and assign *serial numbers* to submitted ideographs as described in Section 2.3.2.

b. After serial numbers are assigned, submitted ideographs must undergo IDS checking to detect any duplication and unification. By the result of IDS checking as described in Section 2.3.3, submitted ideographs will be grouped into M-set and D-set as described in Section 2.3.4.
c. After consolidation, a working draft will be assigned an IRG N document number with a version number, and will be distributed to member bodies' editors and made available on the official website of the IRG so that any other experts can have access to it. The IRG chief editor and technical editor may ask and assign member editors to check M-set and D-set ideographs either for the entire collection or certain portions of it depending on reasonable estimation of workload by the IRG chief editor and technical editor.

3.3. First Checking Stage

This stage, which is between IRG meetings, involves the following tasks:

a. Each member body's editor must check the assigned M-set and D-set for data integrity, correctness, missing data and duplication. Checking for unification is not mandatory, but desirable. Typical review comment examples for each set are provided in Section 4.
b. Member bodies must submit their comments to the IRG chief editor and technical editor at least two months before the next IRG meeting.
c. The IRG chief editor and technical editor must consolidate the comments and produce an IRG registered document for circulation and discussion at least one month before the next IRG meeting.
d. Submitters and outside experts are encouraged to prepare and submit supplementary documents (with IRG document numbers) so that they can be discussed at the next IRG meeting.

3.4. First Discussion and Conclusion Stage

This stage, which is during an IRG meeting, includes the following tasks:

a. Member bodies should review the comments which are officially submitted before the meeting with assigned IRG document numbers and the editorial group must reach conclusions for each commented ideograph in writing. Guidelines for typical conclusions are provided in Section 4.
b. All the conclusions must be agreed to and endorsed by the IRG plenary in its resolutions. As a result of resolution, some ideographs may be removed or moved between M-set and D-Set.
c. The IRG technical editor will create a new M-set and D-set a month after the IRG meeting, and register them as IRG registered documents with version information.
d. If more than 5% of ideographs submitted by a specific submitter are removed as a result of duplication or unification with existing standardized set, the entire submission of this submitter will be removed to ensure high quality of the project. This is known as the 5% Rule described in Section 2.2.6 above.

3.5. Second Checking Stage

This stage, which is between IRG meetings, involves the following tasks:

a. Each member body's editor must check the newly created M-set and D-set for correctness and any duplication.
b. Member bodies should submit their comments with registered IRG document number to the IRG chief editor and technical editor at least two months before the next IRG meeting.
c. The IRG chief editor and technical editor will consolidate the comments and produce a registered IRG document for circulation and discussion at least a month before the next IRG meeting.
d. Member bodies and outside experts are encouraged to prepare and submit supplementary documents to facilitate discussion during the next IRG meeting.

3.6. Second Consolidation and Conclusion Stage

This stage, which is during an IRG meeting, includes the following tasks:

a. Member bodies must review the comments and draw conclusion for each ideograph. Typical comment and conclusion examples for each set are provided in Section 4.

b. All the conclusions must be agreed to and endorsed by the IRG plenary in its resolutions. As a result of the resolutions, some ideographs may be removed or moved between M-set and D-set.
c. The IRG technical editor will create a new M-set and D-set a month after the IRG meeting, and produce an IRG registered document.
d. If more than 5% of the ideographs submitted by a specific submitter are removed as a result of duplication or unification with existing standardized set, the entire submission of this submitter will be removed to ensure high quality of the project.

3.7. Final Checking Stage

This stage, which is between IRG meetings, involves the following tasks:

a. All member bodies' editors are requested to check M-set intensively based on comments and conclusions made in all previous stages. In the final checking stage, no ideographs are allowed to be moved from D Set to M Set.
b. Member bodies' editors must submit their comments to the IRG chief editor and technical editor at least two months before the next IRG meeting.
c. The IRG chief editor and technical editor will consolidate the comments and produce an IRG registered document for circulation and discussion at least a month before the next IRG meeting so that member bodies' editors can have time to review them before the next IRG meeting.

3.8. Approval and Submission to WG2

This stage, which is during an IRG meeting, involves the following tasks:

a. Member bodies should review the comments on M-set and reach conclusions for each ideograph.
b. If there is no positive decision on an M-set ideograph, it will be moved to D-set. No character will be moved from D-set to M-set at this stage. Ideographs may only be moved from M-set to D-set.
c. With the approval from the majority of IRG member bodies, M-set will be frozen as the new ideograph extension set to be submitted to WG2. The IRG technical editor will prepare the document in accordance with Section 2.6 of this document.
d. The remaining D-set ideographs will not be removed. They will be kept and used in the next standardization work. To avoid repetition of discussion of previously checked ideographs, the discussion record will be maintained for future reference.

4. Guidelines for Comments and Resolutions on Working Sets

The following tables list guidelines for typical comments and conclusions during the development process. All comments must be accompanied with date (in YY-MM-DD format) and member body identifier (G, T, H, M, J, K, KP, MY, U or V). All conclusions must also be dated.

4.1. Guidelines for M Set

M-set is the ultimate target of a standardized ideograph set. As such, it must be carefully examined. If any suspicious characters are found, they will be moved to D-sets or removed from the working sets altogether.

Possible Comment by a Reviewer	Possible Resolution
Wrong or Missing Glyph	● Glyph is corrected, or the missing glyph is supplied. The ideograph is moved to D-set for manual checking.
Wrong KangXi radical / strokes count / first stroke	● Data will be corrected and this Ideograph will be moved to D-set for further manual checking.
Wrong IDS	● IDS will be corrected and the character will be moved to D-set until it is checked again by the

	IDS checker. ● Moved to D-set (in case IDS cannot be corrected).
May be unifiable with U+xxxxx (standardized ideograph)	● Unified with U+xxxxx and submitter will request new Source ID to U+xxxxx. ● Unified with U+xxxxx and submitter will request that this character be treated as a Compatibility Ideograph. ● Unified to U+xxxxx and this entry will be removed. (May consider to register it to IVS.) ● Not unifiable.
May be unifiable with xxxxx (M-set ideograph)	● Unified with xxxxx and this source ID will be attached to xxxxx. ● Unified with xxxxx and the submitter may consider registering it as a Compatibility ideograph Character or IVS. ● Not Unifiable.

4.2 Guidelines for D Set

Ideographs in D Set are the ones that either cannot be checked automatically by IDS checking algorithm or the ones of which the attribute data have been questioned by a member body or that are suspected to be unifiable with other standardized or working set ideographs. For the ideographs that cannot be machine-checked by IDS matching, at least two non-submitter member bodies must check them manually to ensure that the ideographs are not unifiable with any standardized ideograph or working set ideograph. For the ideographs that might be unifiable with other ideographs, the submitter of these ideographs is requested to prepare arguments and evidence to show that such ideographs should be separately encoded.

Possible Comment by IDS Checker	**Possible Conclusion**
● Incomplete IDS IDS with extra character ● Component is not an ideograph	● IDS will be corrected and it will be moved to M-set when next IDS-check is done. ● Proper IDS cannot be generated and manual checking is needed.
Possible Comment by a Reviewer	**Possible Conclusion**
● Wrong KangXi radical ● Wrong stroke count ● Wrong first stroke	● Data will be corrected. ● Proposal to correct data is not accepted, as it is an ambiguous case and the IRG agrees that the previous choice of *XX* is more appropriate.
● Wrong IDS	● IDS will be corrected and will be checked by the IDS checker again. ● Correct IDS cannot be generated and manual checking is needed.
May be unifiable with U+xxxxx (standardized ideograph)	● Unified with U+xxxxx and new source is added to U+xxxxx. The new candidate entry should be deleted. ● Not unifiable, as shown by the evidence *IRG N xxxx*. Moved to M-set.
May be unifiable with *xxxxx* (M-set or D-set ideograph)	● Unified with *xxxxx* in M-set and new source of this candidate ideograph is added to xxxx. The new candidate entry is deleted from D-Set.. ● Unified with *xxxxx* in D-Set and new source of this candidate ideograph is added to xxxx in D-Set. The new candidate entry is removed from D-Set. ● Not unifiable, as shown by the evidence *IRG N xxxx*. Move to M-set

Checked against all standardized and working set ideographs with radical *X* and stroke count of *Y*±2.	● Moved to M-set, as two non-submitter member bodies (*XX* and *YY*) confirmed that this ideograph is not unifiable with any existing standardized or working set ideographs. ● Checking against ideographs with radical *X* may not be enough. This ideograph will also be checked against ideographs with radical *Z*.

5. IRG Website

The IRG maintains its own web site at http://www.cse.cuhk.edu.hk/~irg/, hosted by the Department of Computer Science and Engineering at The Chinese University of Hong Kong. IRG meeting notices, minutes, resolutions, document register, documents and standing documents are made available at this site. Hyperlinks to WG2 websites will be provided for member bodies' easy access. For faster retrieval of documents and searching, documents should not be compressed as far as possible and the site search engine window should be made available. Documents larger than 4MB must be split into multiple files for easy uploading, downloading and searching. The compressed files must be in WinZip format with .zip extension.

6. IRG Document Registration

All documents to be formally discussed by the IRG must be registered with assigned IRG document numbers (assigned by the IRG Rapporteur) and containing submission date, title, submitting member body, or the author, purpose (or summary), and the 'IRG Ideographic Repertoire Submission Summary Form' (when applicable).

6.1. Registration Procedures

The following gives the registration procedures:

a. **Request for Document Number:** All documents submitted to the IRG must be given a registered document number. The assignment is done by the IRG Rapporteur. A member body will first contact the IRG Rapporteur for a document number with a document title. Once the document number is assigned, the information will be posted on the IRG website. Some document numbers can be pre-assigned during IRG meetings for activities between IRG meetings.
b. **Submission of Documents:** All registered documents must be submitted to the IRG Rapporteur. The submitted documents must also contain an assigned IRG document number in text form so that searching can be supported.
c. **Posting of Documents**: Properly submitted documents are then posted by the IRG Rapporteur on the IRG website as official documents.
d. **Disqualified Documents**: Documents with certain basic information missing such as submitter's name, title and purpose may be rejected by the IRG Rapporteur for posting. All other documents which fail to comply with the above registration process and the preliminary review by the IRG Rapporteur for basic information will not be treated as IRG documents. As such, issues to be addressed contained in such documents will not be discussed by the IRG formally.

6.2. Contact for IRG Document Registration

The current IRG Rapporteur is Prof. Qin LU and her contact information is as follows:

Professor Qin Lu
Department of Computing
The Hong Kong Polytechnic University
Hung Hom, Hong Kong
Tel. (852) 2766 7247
Fax. (852) 2774 0842
Email: csluqin@comp.polyu.edu.hk

Annex A: Sorting Algorithm of Ideographs

Ideographs must be sorted by the following order.

a. **KangXi Radical Order**.
Note: When radicals are in simplified forms given below, ideographs with simplified radicals must be placed after the ideographs with corresponding traditional radicals.

Traditional Radicals		Simplified Radicals	
R119.0	糸	R119.1	纟
R146.0	見	R146.1	见
R148.0	言	R148.1	讠
R153.0	貝	R153.1	贝
R158.0	車	R158.1	车
R166.0	金	R166.1	钅
R167.0	長	R167.1	长
R168.0	門	R168.1	门
R177.0	韋	R177.1	韦
R180.0	頁	R180.1	页
R181.0	風	R181.1	风
R182.0	飛	R182.1	飞
R183.0	食	R183.1	饣
R186.0	馬	R186.1	马
R194.0	魚	R194.1	鱼
R195.0	鳥	R195.1	鸟
R196.0	鹵	R196.1	卤
R198.0	麥	R198.1	麦
R204.0	黽	R204.1	黾
R209.0	齊	R209.1	齐
R210.0	齒	R210.1	齿

R211.0	龍	R211.1	龙

b. **Stroke Count**.
Note: Simplified characters must be placed after traditional characters within the same stroke-number group.

Annex B: IDS Matching

B.1. Guidelines on Creation of IDS

Each member body should consult IRG N1183 on IDS creation finalized at IRG Meeting No. 25. It should be noted that in addition to the CDC (Character Description Components) defined in IRG N1183, all unified CJK ideographs accepted by ISO 10646 in its amendments are also qualified as CDC in constructing IDS[1].

The use of “overlapping” IDC or more than four IDCs is considered to be ‘inappropriate’ and may not be a subject of IDS comparison.

B.2. Requirements on IDS Matching

The IDS matching algorithm used by the IRG should support the following features:

1. IDS matching should be able to handle different split points.
   (e.g. ⿰亻頃 and ⿰化頁 should be matched.)
2. IDS matching should be able to handle different split levels.
   (e.g. ⿰亻悉 and ⿰亻⿱釆心 should be matched.)
3. IDS matching should match different glyphs of the same abstract shape.
   (e.g. ⿰礻申 and ⿰示申 should be matched.)
4. IDS matching should match similar glyphs.
   (e.g. ⿰忄生 and ⿰小生 should be matched.)
5. IDS matching should match IDS with different orderings of overlapping IDC.
   (e.g. ⿻三丨 and ⿻丨三 should be matched.)
6. IDS matching should match unifiable IDC patterns.
   (e.g. ⿰麥离 and ⿺麥离 should be matched.)
7. IDS matching should be able to handle any combination of the above.
8. IDS matching should be able to detect any inappropriate IDS, such as IDS being too long, IDS with non-ideographic DC, or missing or extra DC or IDC.

B.3. Limitation of IDS Matching

It should be noted that IDS matching cannot detect unification or duplication if a component cannot be encoded by an IDS, or if the glyph itself is very complex. IDS matching is done algorithmically. It is not versatile on detection of the unifiable ideographs unless rules are explicitly given to the algorithm. Thus, it is not meant to be the replacement of manual checking. Rather, it is an assistive tool for quality assurance to identify duplication and known cases of unification. Therefore, it is very important for submitters to make sure that their submitted ideographs are not going to be unified with any standardized or previously discussed ideographs or working set ideographs.

---

[1] Currently, Amendment 1, 4, 5, 6 and 8 have unified CJK ideographs

Annex C: Urgently Needed Ideographs

C.1. Introduction

When a member body urgently needs a few ideographs to be standardized for some good reasons (such as they are ideographs in Regional or National Standard), the member body may, with the approval of the IRG, submit the ideographs independent of any of the current IRG working set to WG2.

C.2. Requirements

The submitter of urgently needed ideographs must prepare the following documents:

a. All the documents required as in normal ideograph submissions.
b. In addition to the above, a document to show any unifiable ideographs in the current IRG working sets against the submitted ideographs. When the urgently needed ideographs are accepted by WG2, the unifiable ideographs in the current working sets will be removed as explained in C.3 below.
c. For ideographs not mentioned above, the document must prove that their submitted ideographs are not unifiable with any ideographs in the current working set. The proof may be provided by listing the documents the submitter has checked, and for each proposed ideograph, a list of ideographs whose radicals and strokes were checked against. It is an important responsibility of the submitter to check with not only current standardized CJK ideographs, but also the IRG working set for any unifiable characters against its submission. If a submitter fails to do the above, the submission will not be approved by the IRG as an IRG-endorsed independent submission to WG2.

C.3. Dealing with Urgent Requests

The IRG may at its discretion accept the document from the submitter of urgently needed ideographs for discussion if the amount of work is considered to be reasonably small for IRG review without unreasonable disruption to its on-going projects. Accepted submissions must be checked by the IRG for correctness, duplication and unification. All accepted ideographs as an independent submission must be checked with the current IRG working sets. When an ideograph is found to be identical or unifiable with the ones in the current IRG working sets, such ideograph must be noted and removed from the current IRG working sets if approval by WG2 is given.

Annex D: Up-to-Date CJK Unified Ideograph Sources and Source References

D.1. Member body code:
  G: China
  H: Hong Kong Special Administrative Region, China
  J: Japan
  K: Republic of Korea
  KP: Democratic People’s Republic of Korea
  M: Macao Special Administrative Region, China
  MY: Malaysia (Added in Nov. 2008 at IRG Meeting No. 31)
  T: Taipei Computer Association
  U: Unicode Consortium
  V: Vietnam

D.2. The Hanzi G sources
  G0 GB2312-80
  G1 GB12345-90 with 58 Hong Kong and 92 Korean “Idu” characters
  G3 GB7589-87 traditional forms
  G5 GB7590-87 traditional forms
  G7 General Purpose Hanzi List for Modern Chinese Language, and General List of Simplified Hanzi
  GS Singapore Characters
  G8 GB8565-88
  GE GB16500-95
  G_KX Kangxi Dictionary ideographs（康熙字典）including the addendum（康熙字典補遺）

  G_HZ Hanyu Dazidian ideographs（漢語大字典）

  G_CY Ci Yuan （辭源）

  G_CH Ci Hai （辞海）

  G_HC Hanyu Dacidian （漢語大詞典）

  G_BK Chinese Encyclopedia （中國大百科全書）

  G_FZ Founder Press System （方正排版系统）

  G_4K Siku Quanshu （四庫全書）

D.3. Hanzi H sources
  Hong Kong Supplementary Character Set (HKSCS)
D.4. Hanzi T sources
  T1 TCA-CNS 11643-1992 first plane
  T2 TCA-CNS 11643-1992 second plane
  T3 TCA-CNS 11643-1992 third plane with some additional characters
  T4 TCA-CNS 11643-1992 fourth plane
  T5 TCA-CNS 11643-1992 fifth plane
  T6 TCA-CNS 11643-1992 sixth plane
  T7 TCA-CNS 11643-1992 seventh plane
  TF TCA-CNS 11643-1992 fifteenth plane
D.5. Kanji J sources
  J0 JIS X 0208-1990
  J1 JIS X 0212-1990
  J3 JIS X 0213:2000 level-3
  J4 JIS X 0213:2000 level-4
  JA Unified Japanese IT Vendors Contemporary Ideographs, 1993
D.6. Hanja K sources
  K0 KS C 5601-1987

K1 KS C 5657-1991
K2 PKS C 5700-1 1994
K3 PKS C 5700-2 1994
K4 PKS 5700-3:1998

D.7. Hanja KP sources
KP0 KPS 9566-97
KP1 KPS 10721-2000

D.8. ChuNom V sources
V0 TCVN 5773:1993
V1 TCVN 6056:1995
V2 VHN 01:1998
V3 VHN 02: 1998

D.9. MY sources
MY1 “ Dictionary Of Chinese Rustic Language In South-East Asia”, written by Xu Yunqiao, published by Singapore Shjie Publisher, 1961. 《南洋华语俚俗辞典》，新加坡世界书局有限公司，1961 年 8 月

D.10. Macao sources
M1 Macao Supplementary Character Set

Annex E: Maintenance Procedure of the IRG Working Documents Series

E.1 Introduction

The IRG Working Documents Series is a set of IRG maintained documents which keeps the up-to-date examples of CJK unification related example cases to supplement the published Annex S of ISO/IEC 10646 for IRG unification work.

E.2. IRG Working Documents Series

The document formats and the specific lists are maintained as a separate set of documents as the IRG Working Documents (IWD).

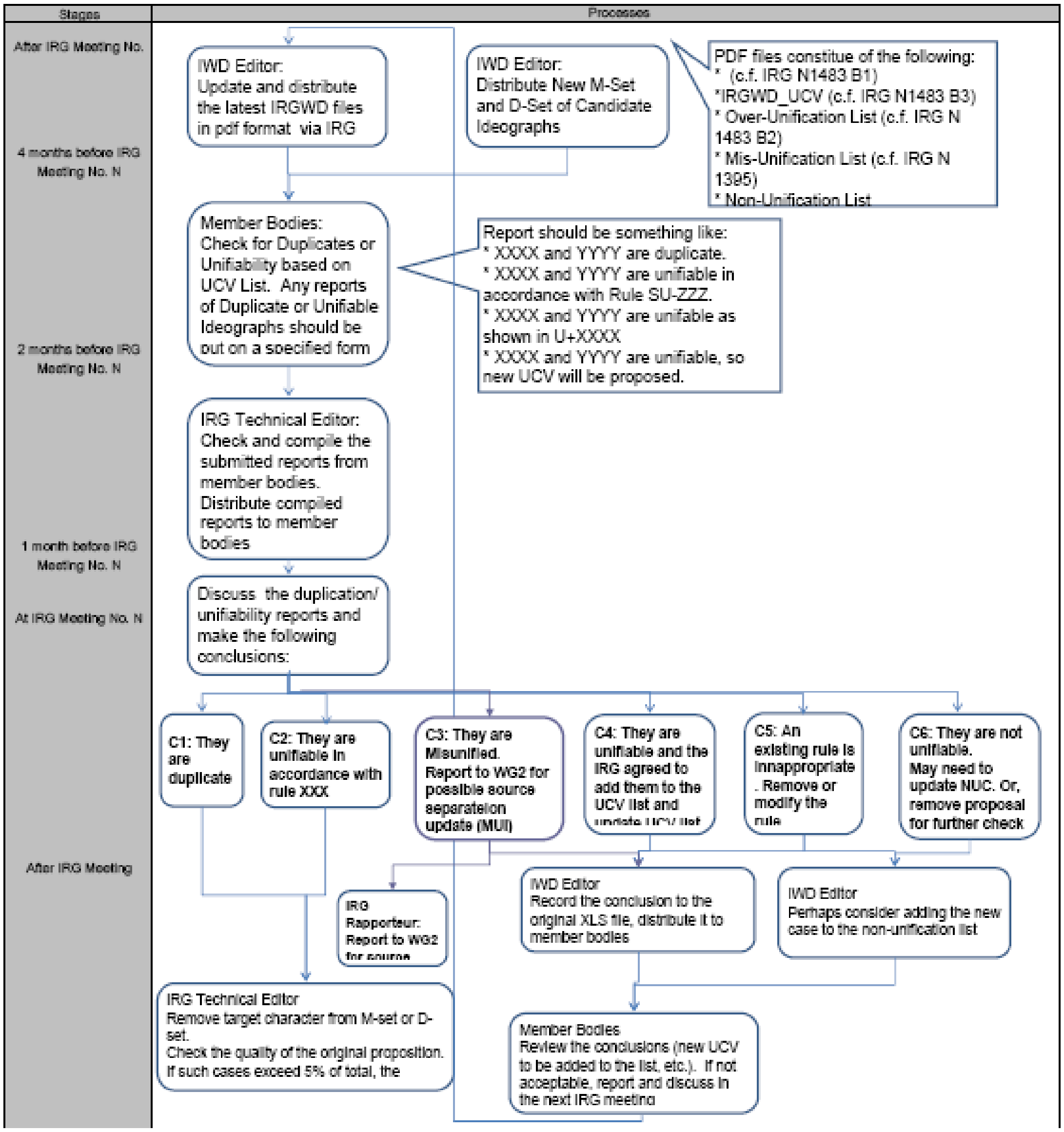
Series 1:	Summary of unification rules and sample examples (File name: IRGWD_SUM.pdf)
Series 2:	List of UCV (Unifiable Component Variations) of Ideograhs (File name: IRGWD_UCV.pdf)
Series 3:	List of Non-Unifiable Components of Ideographs and Overly-Unified Ideographs (File name: IRGWD_NUC.pdf)
Series 4:	List of Possibly Mis-Unified Ideographs (File name: IRGWD_MUI.pdf).

E.3. Maintenance Procedure

The maintenance procedure describes how entries in the IWD are added, removed, or changed. The IRG has an appointed IWD Editor (currently, Mr. Taichi Kawabata) who is in charge of maintaining the IWDS.

In principle, all update requests are results of IRG unification review work. A review cycle between two IRG meetings is needed. Every update must be discussed in at least one IRG meeting and confirmed in writing. The update is normally started from the assigned unification review work given to member bodies in the past IRG meeting (Meeting No. N-1). Then, during the review work before the next IRG meeting (Meeting No. N), if member bodies found duplicates, unifiable cases or mistakes, which may warrant a change in the IWDS, they need to report these cases in a specified form attached to IWD Series 1.These reported cases will then be consolidated by the IRG technical editor before IRG Meeting No. N. During IRG meeting No. N, time must be allocated to discuss these reported cases and conclusions must be recorded during this IRG meeting. Based on the confirmed conclusion on IWDS updates, the IWDS editor will update the IWDS documents. Any unclear conclusions will be further discussed in future meetings.

Below is the description of the maintenance procedure as a flow chart.


Stages
Processes
After IRG Meeting No.
4 months before IRG Meeting No. N
2 months before IRG Meeting No. N
1 month before IRG Meeting No. N
At IRG Meeting No. N
After IRG Meeting
IWD Editor: Update and distribute the latest IRGWD files in pdf format via IRG
IWD Editor: Distribute New M-Set and D-Set of Candidate Ideographs
PDF files constitue of the following:
* (c.f. IRG N1483 B1)
*IRGWD_UCV (c.f. IRG N1483 B3)
* Over-Unification List (c.f. IRG N 1483 B2)
* Mis-Unification List (c.f. IRG N 1395)
* Non-Unification List
Member Bodies: Check for Duplicates or Unifiability based on UCV List. Any reports of Duplicate or Unifiable Ideographs should be put on a specified form
Report should be something like:
* XXXX and YYYY are duplicate.
* XXXX and YYYY are unifiable in accordance with Rule SU-ZZZ.
* XXXX and YYYY are unifable as shown in U+XXXX
* XXXX and YYYY are unifiable, so new UCV will be proposed.
IRG Technical Editor: Check and compile the submitted reports from member bodies. Distribute compiled reports to member bodies
Discuss the duplication/ unifiability reports and make the following conclusions:
C1: They are duplicate
C2: They are unifiable in accordance with rule XXX
C3: They are Misunified. Report to WG2 for possible source separateion update (MUI)
C4: They are unifiable and the IRG agreed to add them to the UCV list and
C5: An existing rule is innappropriate . Remove or modify the rule
C6: They are not unifiable. May need to update NUC. Or, remove proposal for further check
IRG Rapporteur: Report to WG2
IWD Editor Record the conclusion to the original XLS file, distribute it to member bodies
IWD Editor Perhaps consider adding the new case to the non-unification list
IRG Technical Editor Remove target character from M-set or D-set. Check the quality of the original proposition. If such cases exceed 5% of total, the
Member Bodies Review the conclusions (new UCV to be added to the list, etc.). If not acceptable, report and discuss in the next IRG meeting

Please refer to the separate Excel file for a more clear diagram.

Annex F: IRG Repertoire Submission Summary Form

ISO/IEC JTC 1/SC 2/WG 2/IRG
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS OF CJK UNIFIED IDEOGRAPHS TO THE REPERTOIRE OF ISO/IEC 10646
Please fill in all the sections below.
Please read Principles and Procedures Document (P & P) from http://appsrv.cse.cuhk.edu.hk/~irg/irg31/IRGN1562.pdf
for guidelines and details before filling in this form.
Please ensure you are using the latest Form from http://appsrv.cse.cuhk.edu.hk/~irg/SubmissionForm.pdf **.**
See also http://appsrv.cse.cuhk.edu.hk/~irg/UCV.html **for latest *Unifiable Calligraphic Variations*.**

A. Administrative

1. **IRG Project Code:** e.g. Extension-E
2. **Title:**
3. Requester's region/country name:
4. Requester type (National Body/Individual contribution):
5. Submission date:
6. Requested Ideograph Type (Unified or Compatibility Ideographs)
   If Compatibility, does requester have the intention to register them as IVS (See UTS #37) with the IRG's approval? (Registration fee will not be charged if authorized by the IRG.)
7. Request Type (Normal Request or Urgently Needed)
8. Choose one of the following:
   This is a complete proposal:
   (or) More information will be provided later:

B. Technical – General

1. Number of ideographs in the proposal:
2. Glyph format of the proposed ideographs: (128x128 "bmp" files or TrueType font file)
   If 'bmp' files, their file names are the same as their Source IDs?
   If TrueType font, all proposed glyphs are put into BMP PUA area?
   If TrueType font, data for Source IDs vs. character codes are provided?
3. Source IDs:
   Do all the proposed ideographs have a unique, proper Source ID (country/region code and less than 9 alphanumeric characters)?
4. Evidence:
   a. Do all the proposed ideographs have the separate evidence document which contains at least one scanned image of printed materials (preferably dictionaries)?
   b. Do all the printed materials used for evidence provide enough information to track them by a third party (ISBN numbers, etc.)?
5. Attribute Data Format: (Excel file or CSV)

C. Technical - Checklist

Understandings of the Unification Checklist

1. Has the requester read ISO/IEC 10646 Annex S and did the requester understand the unification policy?
2. Has the requester read the “Unifiable Calligraphic Variations” (contact IRG technical editor through the Rapporteurfor the latest one) and did the requester understand the unifiable variation examples?
3. Has the requester read this P&P document and did the requester understand the 5% rule?

Character-Glyph Duplication Checklist(http://www.itscj.ipsj.or.jp/sc2/open/pow.htm contains all the published ones and those under ballot**)**

4. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the unified or compatibility ideographs of ISO/IEC 10646?
   If yes, which version of ISO/IEC 10646 did requester check? (e.g. 10646:2003)
5. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the ideographs in Amendments of current ISO/IEC 10646? (As of 2009, Amendment 1, 4, 5, 6and 8 have CJK ideographs.)
   If yes, which amendments did requester check?
6. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the ideographs in the current IRG working sets or proposed amendments of ISO/IEC 10646? (As of 2009, PDAM 6 and PDAM 8 have CJK ideographs.)
   If yes, which draft amendments did requester check?
7. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the ideographs in the current working M-set and D-set of the IRG? (Contact IRG chief editor and technical editor through the IRG Rapporteur for the newest list)
   If yes, which document did requester check?
8. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the over-unified or mis-unified ideographs in ISO/IEC 10646? (Check Annex E of this document).
9. Has the requester checked that any of the proposed ideographs ***has similar ideograph(s)*** with the ideographs in the current standardized or working set mentioned above?
10. Has the requester checked that any of the proposed ideographs ***has variant ideograph(s)*** with the ideographs in the current standardized or working set mentioned above?

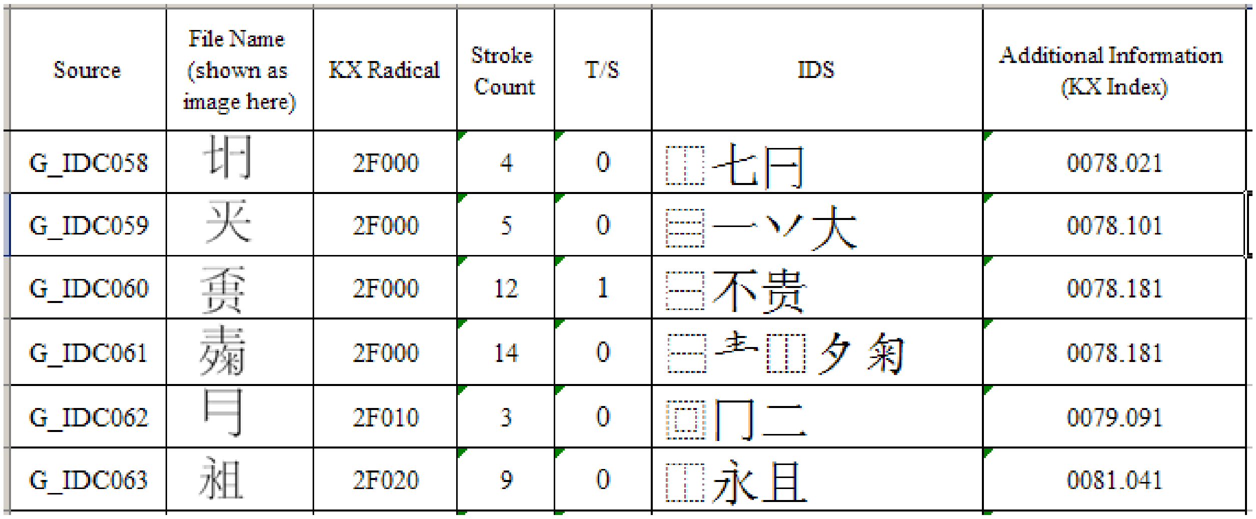
Attribute Data Checklist

11. Do all the proposed ideographs have attribute data such as the KangXi radical code, stroke count and first stroke?
12. Are there any simplified ideographs (ideographs that are based on the policy described in 簡化字總表) in the proposed ideographs?
    If YES, does your proposal include proper simplified/traditional indication flag for each proposed ideograph in attribute data?
13. Do all the proposed ideographs have the document page number of evidence documents in attribute data?
14. Do all the proposed ideographs have the proper Ideographic Description Sequence (IDS) in attribute data?
    If NO, how many proposed ideographs do not have the IDS?
15. If the answer to question 9 or 10 is yes, do the attribute data include any information on similar/variant ideographs for the proposed ideographs?

Annex G: Examples of Unified CJK Submissions

G.1. Sample Data Files

All submitted characters must follow the submission format given in Section 2.2.3.
The following gives list of samples submitted by China from its Ministry of Public Security for consideration in CJK Ext. D work (IRGN 1366 Appendix 3 and Appendix 4).

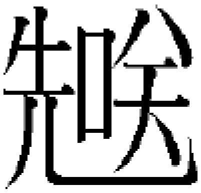


Source	File Name (shown as image here)	KX Radical	Stroke Count	T/S	IDS	Additional Information (KX Index)
G_IDC058		2F000	4	0	⿰七冃	0078.021
G_IDC059		2F000	5	0	⿳一丷大	0078.101
G_IDC060		2F000	12	1	⿱不贵	0078.181
G_IDC061		2F000	14	0	⿱龶⿰夕匊	0078.181
G_IDC062		2F010	3	0	⿴冂二	0079.091
G_IDC063		2F020	9	0	⿰永且	0081.041

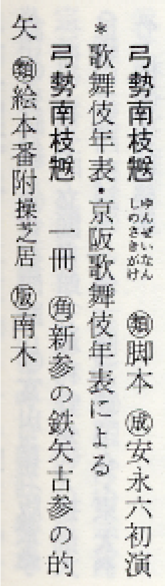
G.2. Sample Evidence

All character submissions must include evidence of use as specified in Section 2.2.4. The following shows an example of a Japanese submission with reference to the use of the character in ancient books (IRG N1225 Part2).

JMK66114



『補訂版国書総目録』(1969年4月30日第1刷発行, 2002年7月5日補訂版第4刷発行,
岩波書店)
第7巻870ページ4段

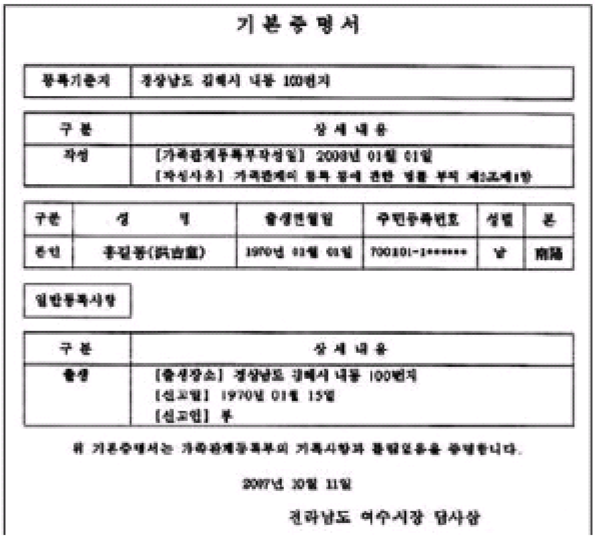


弓勢南枝魁（ゆんぜいなんしのさきがけ） (類)脚本 (成)安永六初演
＊歌舞伎年表・京阪歌舞伎年表による
弓勢南枝魁 一冊 (角)新参の鉄矢古参の的
矢 (類)絵本番附操芝居 (版)南木

G.3. Handling of Data with Privacy Concerns

The IRG understands that the current privacy laws and practices in different Countries and Regions can make the submission of complete records as evidence related to personal information difficult. As a compromise, the IRG suggests member bodies to provide evidence in such a way that it would not reveal complete personal/internal information. However, the character information itself must be shown in the supplied evidence. In other words, partial document images should be supplied with certain sensitive information blocked.

As different departments/organizations may have different types of documents, the IRG suggests that, for each type of document, a submitter provides a sample document with private information deleted. A good example is the original Basic Certificate of Family Relation Register in Korea as seen in Fig. G1. The evidence can be submitted as partial data in the form shown in Fig. G2.



기 본 증 명 서

등록기준지	경상남도 김해시 내동 100번지

구 분	상 세 내 용
작성	[가족관계등록부작성일] 2008년 01월 01일 [작성사유] 가족관계의 등록 등에 관한 법률 부칙 제3조제1항

구분	성 명	출생연월일	주민등록번호	성별	본
본인	홍길동(洪吉童)	1970년 01월 01일	700101-1******	남	南陽

일반등록사항

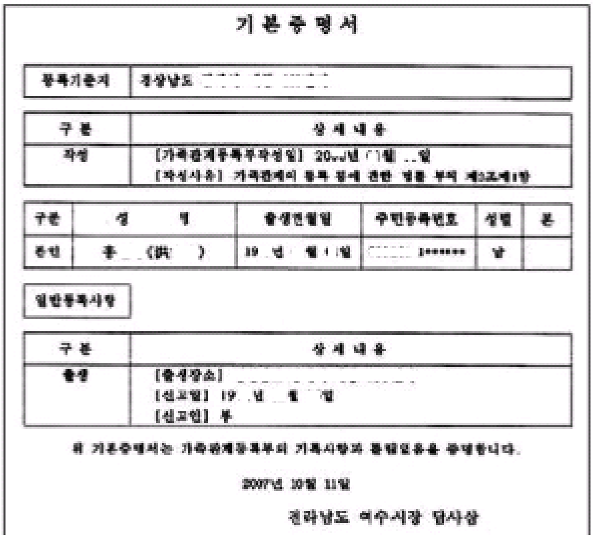
구 분	상 세 내 용
출생	[출생장소] 경상남도 김해시 내동 100번지 [신고일] 1970년 01월 15일 [신고인] 부

위 기본증명서는 가족관계등록부의 기록사항과 틀림없음을 증명합니다.

2007년 10월 11일

전라남도 여수시장 담사상

Figure G1. The original Basic Certificate of
Family Relation Register



기 본 증 명 서

등록기준지	경상남도

구 분	상 세 내 용
작성	[가족관계등록부작성일] 20 년 월 일 [작성사유] 가족관계의 등록 등에 관한 법률 부칙 제3조제1항

구분	성 명	출생연월일	주민등록번호	성별	본
본인	홍 (洪)	19 년 월 일	******	남	

일반등록사항

구 분	상 세 내 용
출생	[출생장소] [신고일] 19 년 월 일 [신고인] 부

위 기본증명서는 가족관계등록부의 기록사항과 틀림없음을 증명합니다.

2007년 10월 11일

전라남도 여수시장 담사상

Figure G2. A modified Basic Certificate of
Family Relation Register (private information
such as full name and date of birth has been deleted).

[Annex H] Not Used at the Moment

Annex H is purposely left out for the time being so that IRG Annex numbers can be the same as WG2 Annex numbers where the subjects are the same.

Annex I: Guideline for Handling of CJK Ideograph Unification or Dis-unification Errors

(Same as WG2 Principles and Procedures Annex I)
Source: www.dkuug.dk/jtc1/sc2/wg2/docs/principles.html

There are two kinds of errors that may be encountered related to coded CJK unified ideographs.

Case 1: *to be unified* error - Ideographs that should have been unified are assigned separate code points.

Case 2:*to be dis-unified* error - Ideographs that should not have been unified are unified and assigned a single code point. An example of this is the request from TCA in document N2271.

When such errors are found, the following guidelines will be used by WG2 to deal with them.

I.1 Guideline for "To Be Unified" Errors

A. The "*to be unified"* pair will be left dis-unified. Once a character is assigned a code position in the standard, it will not be removed from the standard.
B. If necessary, an additional note may be added to an appropriate section in the standard.

I.2 Guideline for "To Be Dis-unified" Errors

A. The ideographs to be dis-unified should be dis-unified and should be given separate code positions as soon as possible (dis-unification in some sense, and character name change in some sense also). These ideographs will have two separate glyphs and two separate code positions. One of these ideographs will stay at its current encoded position. The other one will have a new glyph and a new code position.
B. For the ideographs that are encoded in the BMP, the code charts in ISO/IEC 10646 are presented in multiple columns, with possibly differing glyph shapes in each column. The question of which glyph will be used for the currently encoded ideograph will be resolved as follows. In the interest of synchronization between ISO/IEC 10646 and the Unicode standard, the ideograph with the glyph shape that is similar to the glyph that is published in the "Unicode Charts" will continue to be associated with its current code position. For the ideographs outside the BMP, the glyph shape in ISO/IEC 10646 and the Unicode Charts are identical and will be used with its current code position.
C. The dis-unified ideograph will have a glyph that is different from the one that retains the current code position.
D. The net result will be an addition of new ideograph character and a correction and an additional entry to the source reference table.

I.3 Discouragement of New Dis-unification Request

There is a possibility of "pure true dis-unification" request. This is almost like the new source code separation request. This kind of request will not be accepted disregarding the reasoning behind. Key difference between "TO BE DIS-UNIFIED" and "WILL NOT BE DIS-UNIFIED is as follows.

a. If character pair is non-cognate (meanings are different), that pair of characters is TO BE DIS-UNIFIED.
b. If a character pair is cognate (means the same but different shape), that pair of characters WILL NOT BE DIS-UNIFIED.

Dis-unification request with reason of mis-application (over-application usually) of unification rule should NOT be accepted due to the principle in resolution M41.11.

Annex J: Guideline for Correction of CJK Ideograph Mapping Table Errors

(Same as WG2 P&P Annex J)
Source: ISO/IEC JTC 1/SC 2/WG 2 N2577 – 2003-09-02

In principle, mapping table or reference to code point of existing national/regional standard (in the source reference tables) must not be changed. But once a fatal error is found it should be corrected as early as possible, under the following guidelines:

J.1 Priority of Error Correction Procedure

A. Consider adding new code position and source-reference mapping for the character in question rather than changing the mapping table.
B. If change of mapping table is unavoidable, correction should be done as soon as possible.

J.2 Announcement of Addition or Correction of Mapping Table

Once any addition or correction of mapping table is made, an announcement of the change should be made immediately. Usually this will be in the form of a resolution of a WG2 meeting, followed by subsequent process resulting in an appropriate amendment to the standard.

J.3 Collection and Maintenance of Mapping Tables that are not Owned by WG2

There are many mapping tables, which are included in national/regional standards or developed by third parties. These are out of WG2's scope. Any organization (such as Unicode Consortium) that collects mapping information, maintains it consistently and makes this information widely available is invited and encouraged to do so.

References

Document numbers in the first column in the following table refer to IRG working documents (ISO/IEC JTC 1/SC 2/WG 2/IRGNxxxx), except where noted otherwise. For documents with no link, you may try http://www.cse.cuhk.edu.hk/~irg/ ; some older documents may only be available in paper form (contact the IRG Rapporteur Prof. Qin LU).

Doc. No.	Title	Source	Date
WG2 N3201	Principles and Procedures for Allocation of New Characters and Scripts and Handling of Defect Reports on Character Names	WG2	2007-03-14
N681	Annex S http://standards.iso.org/ittf/PubliclyAvailableStandards/c039921_ISO_IEC_10646_2003(E).zip	Bruce Peterson and IRG Rapporteur	1999-11-18
N881	CJK Extension C Submission Format	IRG	2001-12-04
N953	Minutes of the Adhoc meeting on submitted documents: N941, N942, N944, N945, N948, N949	CJK ad hoc group	2002-11-22
N954	Report on first stroke/stroke count by ad hoc group	CJK ad hoc group	2002-11-22
N954AR	N954 Appendix: First Stroke / Stroke Count Chart	CJK ad hoc group	2002-11-21
N955	IRG Radical Classification	Ideograph Radical Ad Hoc	2002-11-21
N956	Ideograph Unification	Ideograph Radical Ad Hoc	2002-11-21
N1105	Amendments to IRG N954AR	Macao	2005-01-03
N1183	IDS decomposition principles(Revised by the IRG)	KAWABATA, Taichi	2005-12-28
N1197	Sample evidence for CJK C1 candidates	Japan	2006-05-22
N1372	On Better use of IDS on IRG development process	KAWABATA, Taichi	2007-11-09
SC2 N3933		SC2	

Glossary:

CJK Unified Ideographs. It refers to the collection of unified Han characters in ISO 10646 standard. CJK stands for Chinese, Japanese and Korean. The term CJK Unified Ideographs was adopted at the earlier years of the IRG to reflect the development work of the Han character unification from the three languages at that time. It is obvious today that Han unification covers far beyond the scripts used in China, Japan and Korea. However, the name is consistently being used in the standardization process and is not changed.

Source: A reputable published document such as a dictionary, a standardization document, or a well published and widely read or referenced book which the IRG would consider as authoritative such that the characters in this source are considered reliable and stable for consideration of inclusion. A set of ISO 10646 accepted sources is listed in Section 27 of the ISO 10646 document.

New Source: Any CJK source that is newly submitted by IRG member bodies which is not yet accepted by ISO 10646, thus is not present in Section 27 of ISO 10646. Member bodies may first submit their new source to the IRG for acceptance. Once accepted, the characters in that source can be accepted by the IRG for consideration for inclusion in future extensions. The IRG will also submit the source to WG2 for approval and inclusion in Section 27 of ISO 10646.

Ideographic Description Characters (IDC): The 12 characters defined in ISO 10646 starting from the code point U+2FF0: ⿰⿱⿲⿳⿴⿵⿶⿷⿸⿹⿺⿻.

Ideographic Description Sequence (IDS): IDS describes a character using its components and indicating the relative positions of the components. IDCs are considered operators to the components. IDSs can be expressed by a context free grammar through the Backus Naur Form (BNF). The grammar G has four components:

Let G = {Σ, N, P, S}, where

- Σ: the set of terminal symbols including all coded radicals, coded ideographs, and the 12 IDCs.
- N:the set of 5 non-terminal symbols
  N={IDS, IDS1, Binary_Symbol, Ternary_Symbol, CDC[2]}
- S = {IDS}, which is the start symbol of the grammar
- P: a set of rewrite rules

The following is the set of rewriting rules P:

- IDS::=<Binary_Symbol><IDS1><IDS1>|<Ternary_Symbol>
  <IDS1><IDS1><IDS1>
- <IDS1> ::= <IDS> | <CDC>
- <CDC>::= coded_ideograph | coded_radical | coded_component
- <Binary_Symbol> ::=⿰|⿱ | ⿴ | ⿵ | ⿶ | ⿷ | ⿸ | ⿹ | ⿺ | ⿻
- <Ternary_Symbol> ::= ⿲ | ⿳

Note that even though the IDCs are terminal symbols, they are not part of the Character Description Components.

Abstract shape: Ideographic characters are used as symbols to represent different entities and used for different purposes. The same character conceptually can sometimes be written in different actual shapes with minor stroke differences, due to preference, which do not affect the recognition of the character as a unique symbol. These characters having the same abstract shapes are not coded separately because ISO 10646 is a character (symbol) standard, not a glyph standard. In other words, character glyphs (actual shapes) that are considered to have the same abstract shapes are to be unified under the CJK unification rules (defined in Annex S of ISO 10646).

---

[2] Stands for Character Description Components

As ideographs are formed by both the components and the relative positioning of the components, the examination of glyph difference is observed by taking into consideration the meaning, components, and their relative positions. Characters having different meanings and different actual shapes are not considered to have the same abstract shapes. Characters having the same components yet different in relative positions are generally considered to have different abstract shapes. However, component difference is subjected to examination by experts to see if they have influenced the recognition of the character as a whole with consideration of the character's origin and use. Annex S of ISO 10646 has defined the examination procedure which is given below:

> "*The examination of character glyphs are through*
> *a) the number of components,*
> *b) the relative position of the components in each complete ideograph,*
> *c) the structure of corresponding components.*
>
> *If one or more of the features a) to c) above are different between the ideographs in the comparison, the ideographs are considered to have different abstract shapes and are therefore not unified. If all of the features a) to c) above are the same between the ideographs, the ideographs are considered to have the same abstract shape and are therefore unified."*

Please also refer to Annex S in ISO 10646 for examples of characters and components that are considered to have the same abstract shape. The IRG maintains an up-to-date Unification Examples List.

Working set: A working set is the set of characters accepted by the IRG as a collection to work on for extension to ISO 10646. Characters accepted in a working set are subjected to review by IRG member bodies for inclusion in a particular extension.

M-set (main set): M-set is the set of characters that have been reviewed and accepted by IRG member bodies without pending questions in the current working set.

D-set (discussion set): D-set is the set of characters that have been reviewed by IRG member bodies with pending issues which need further discussion/evidence for inclusion in the M-set of a working set.

Compatibility Ideographs: Compatibility ideographs are a kind of compatibility characters defined in Section 22 of ISO 10646. Below is a direct quote from ISO 10646-2003:

> *"The CJK compatibility ideographs (characters that are part of the CJK COMPATIBILITY IDEOGRAPHS-2001 collection) are ideographs that should have been unified with one of the CJK unified ideographs (characters that are part of the CJK UNIFIED IDEOGRAPHS-2001 collection), per the unification rule described in annex S. However, they are included in this International Standard as separate characters, because, based on various national, cultural, or historical reasons for some specific country and region, some national and regional standards assign separate code positions for them."*

附件 7

N3787：Request for disunifying U+2F89F from U+5FF9

Universal Multiple-Octet Coded Character Set
UCS

ISO/IEC JTC1/SC2/WG2 **N3787**

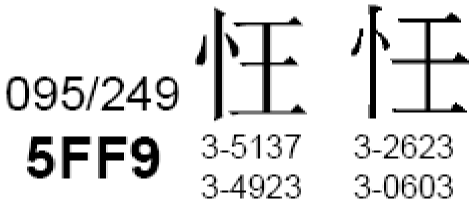
ISO/IEC JTC1/SC2/WG2 IRG N 1666

Date: 2010-3-25

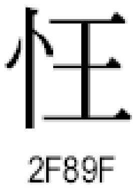
Doc. Type:	Member body contribution
Title:	Request for disunifying U+2F89F from U+5FF9
Source:	TCA and China
Status:	Input to WG2 and IRG
Action:	For discussion on the meeting
Distribution:	WG2 members and IRG members
No. of pages:	6
Appendixes	None
Medium:	Electronic

1 Background

The character U+2F89F was encoded as a compatibility ideograph because it had been unified to the character U+5FF9, please refer to the figure shown below and the entry “U+2F89F” in the electronic insertion file “CJKC_SR.txt” of ISO/IEC 10646:2003. TCA and China found that both of the glyphs shown in the G-column and T-column of the code point U+5FF9 are wrong and the characters U+5FF9 and U+2F89F are non-cognate, as mentioned below.



(see p.477 in ISO/IEC 10646:2003)

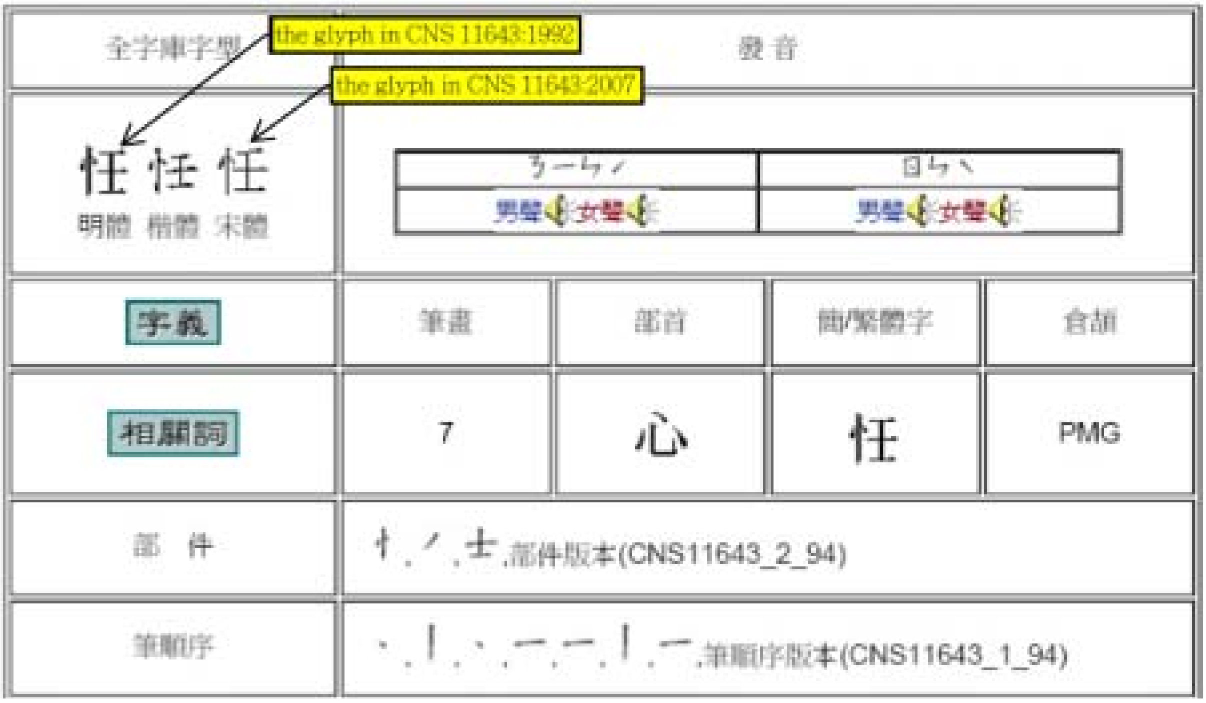


(see p.1310 in ISO/IEC 10646:2003)

1.1 The glyph, pronunciations and meanings of the character T3-2623

The CNS Character Database (http://www.cns11643.gov.tw) provides attributes of characters encoded in CNS 11643, and the Dictionary of Chinese Character Variants (http://140.111.1.40/main.htm) provides glyphs, pronunciations and meanings of 106,230 ideographs and the variantal relationships among them. Both of the websites are build and maintained by Taiwan’s government. It means that all information provided by those websites are offically.

By means of the CNS Character Database, the pronunciation of the character T3-2623 can be found as either “ren (ㄖㄣˋ)” or “nin (ㄋㄧㄣˊ)”, as shown following:



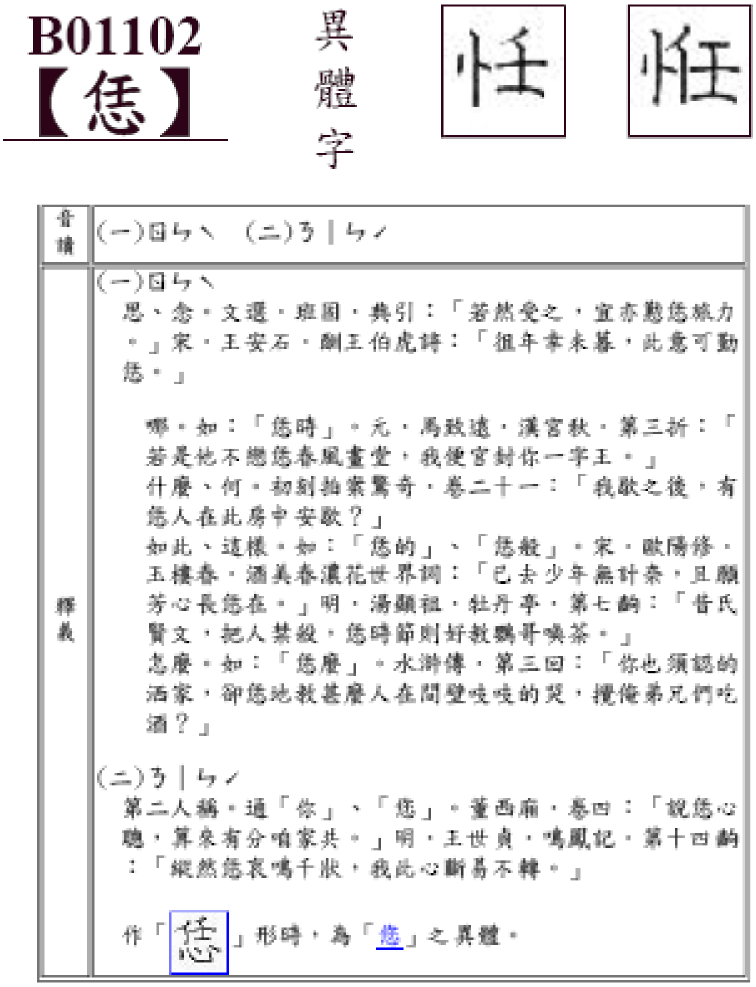
全字庫字型	發音			
忹 忹 忹 明體 楷體 宋體	ㄋㄧㄣˊ 男聲 女聲	ㄖㄣˋ 男聲 女聲		
字義	筆畫	部首	簡/繁體字	倉頡
相關詞	7	心	忹	PMG
部 件	忄, ㇒, 士,部件版本(CNS11643_2_94)			
筆順序	丶, 丨, 丶, 一, 一, 丨, 一,筆順序版本(CNS11643_1_94)			

(see http://www.cns11643.gov.tw/AIDB/query_general_view.do?page=3&code=2623)

Most of Chinese characters are consisted of a radical part and a phonetic part. For a Chinese charcter having radical-phonetical structure, the radical part (namely radical) determines its class of meaning and the phonetic part (namely phontical) determines its pronunciation and sometime more specific meaning. It’s a common rule for a radical-phonetical structured Chinese character that the pronunciation of the character always simlar to the pronunciation (or accient pronunciation) of its phonetical.

In the CNS 11643:1992, the character T3-2623 (忹) is consisted of the radical “心” and the phonetical “王” with pronunciation “wang”. Because an inconsistency exists between pronunciations of the character T3-2623 (i.e., “ren” in offical record from CNS Character Database) and its phonetical (i.e., “wang”), the phonetical shall be wrong. That’s the reason why Taiwan’s MOE modified the glyph of T3-2623 to be the compositon of the radical “心” and the phonetical “壬” with pronunciation “ren” in the CNS 11643:2007.

By means of the Dictionary of Chinese Character Variants, the meaning of the character T3-2623 with pronunciation “ren” is relative to “think about”, and the meaning with pronunciation “nin” is “you (in good manners)”, as shown below.



B01102
【恁】

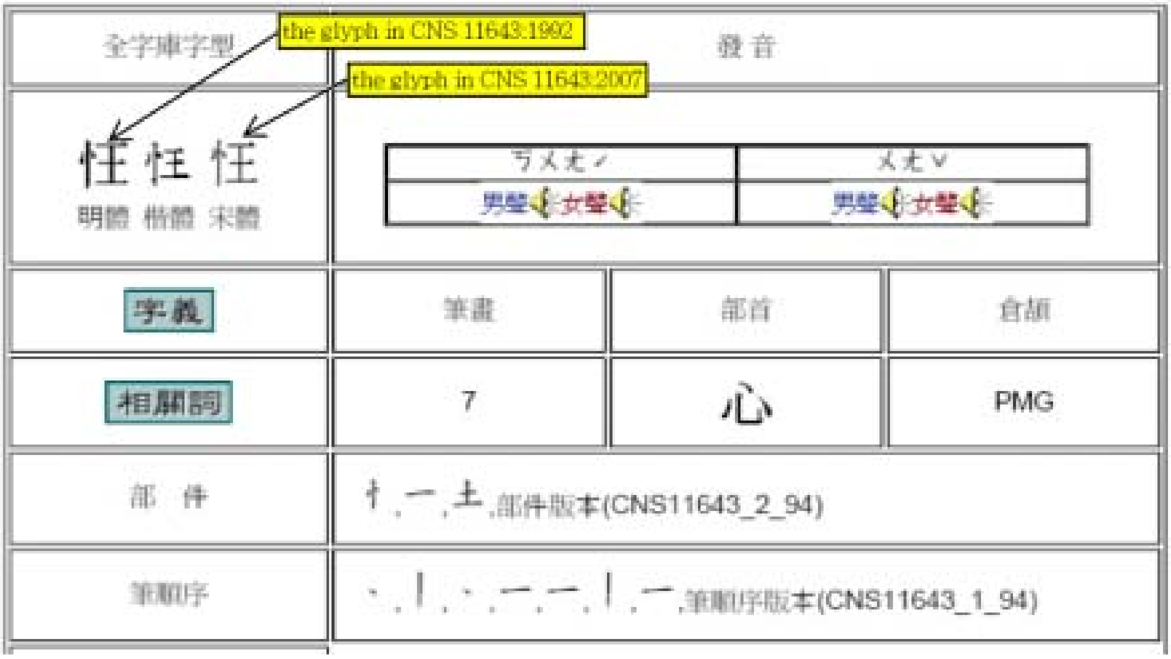
異體字

音讀	(一)ㄖㄣˋ　(二)ㄋㄧㄣˊ
釋義	(一)ㄖㄣˋ 思、念。文選．班固．典引：「若然受之，宜亦勤恁旅力。」宋．王安石．酬王伯虎詩：「徂年幸未暮，此意可勤恁。」 哪。如：「恁時」。元．馬致遠．漢宮秋．第三折：「若是他不戀恁春風畫堂，我便官封你一字王。」 什麼、何。初刻拍案驚奇．卷二十一：「我歇之後，有恁人在此房中安歇？」 如此、這樣。如：「恁的」、「恁般」。宋．歐陽修．玉樓春．酒美春濃花世界詞：「已去少年無計奈，且願芳心長恁在。」明．湯顯祖．牡丹亭．第七齣：「昔氏賢文，把人禁殺，恁時節則好教鸚哥喚茶。」 怎麼。如：「恁麼」。水滸傳．第三回：「你也須認的洒家，卻恁地教甚麼人在間壁吱吱的哭，攪俺弟兄們吃酒？」 (二)ㄋㄧㄣˊ 第二人稱。通「你」、「您」。董西廂．卷四：「說恁心聰，算來有分咱家共。」明．王世貞．鳴鳳記．第十四齣：「縱然恁哀鳴千狀，我此心斷易不轉。」 作「」形時，為「恁」之異體。

(see http://140.111.1.40/yitib/frb/frb01102.htm)

1.2 The glyph, pronunciations and meanings of the character T5-2438

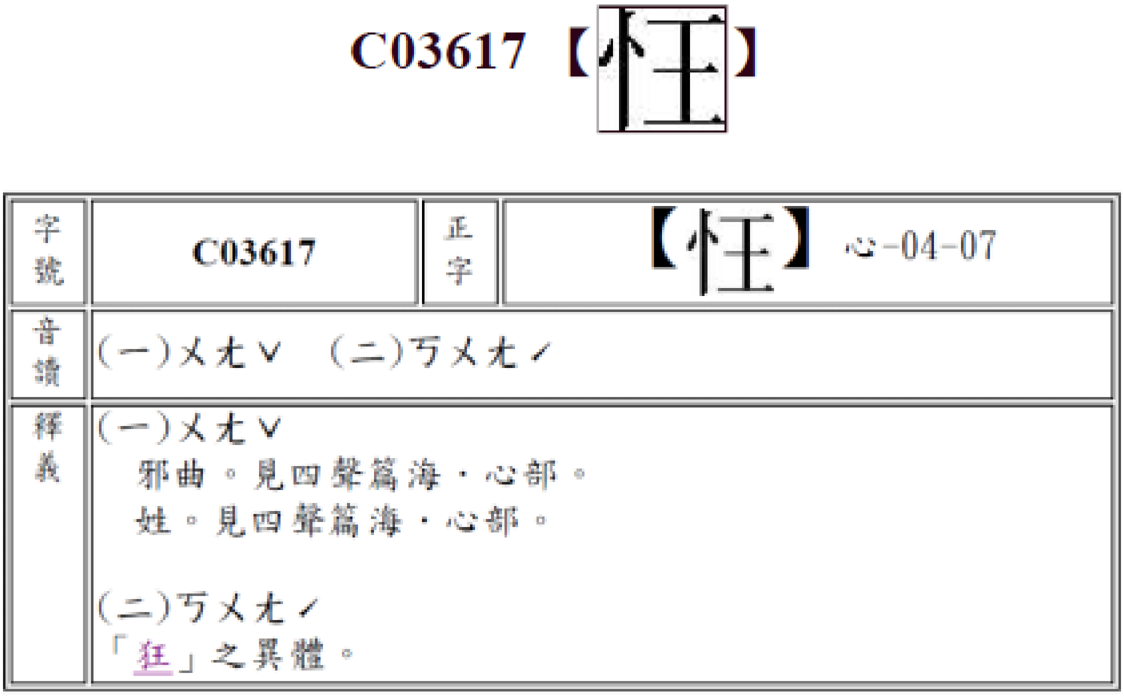
By means of the CNS Character Database, the pronunciation of the character T5-2438 can be found as either “wang(ㄨㄤˇ)” or “kuang (ㄎㄨㄤˊ)”.　In both of the CNS 11643:1992 and CNS 11643:2007, the character T5-2438 is consisted of the radical “心” and the phonetical “王” with pronunciation “wang”, as shown following:



全字庫字型	發音		
忹 忹 忹 明體 楷體 宋體	ㄎㄨㄤˊ 男聲 女聲	ㄨㄤˇ 男聲 女聲	
字義	筆畫	部首	倉頡
相關詞	7	心	PMG
部　件	忄,一,土,部件版本(CNS11643_2_94)		
筆順序	丶,丨,丶,一,一,丨,一,筆順序版本(CNS11643_1_94)		

(see http://www.cns11643.gov.tw/AIDB/query_general_view.do?page=5&code=2438)

By means of the Dictionary of Chinese Character Variants, the meaning of the character T5-2438 with pronunciation “wang” is relative to “wicked”, and the meaning with pronunciation “kuang” is relative to “undisciplined” or “crazy”, as shown below.



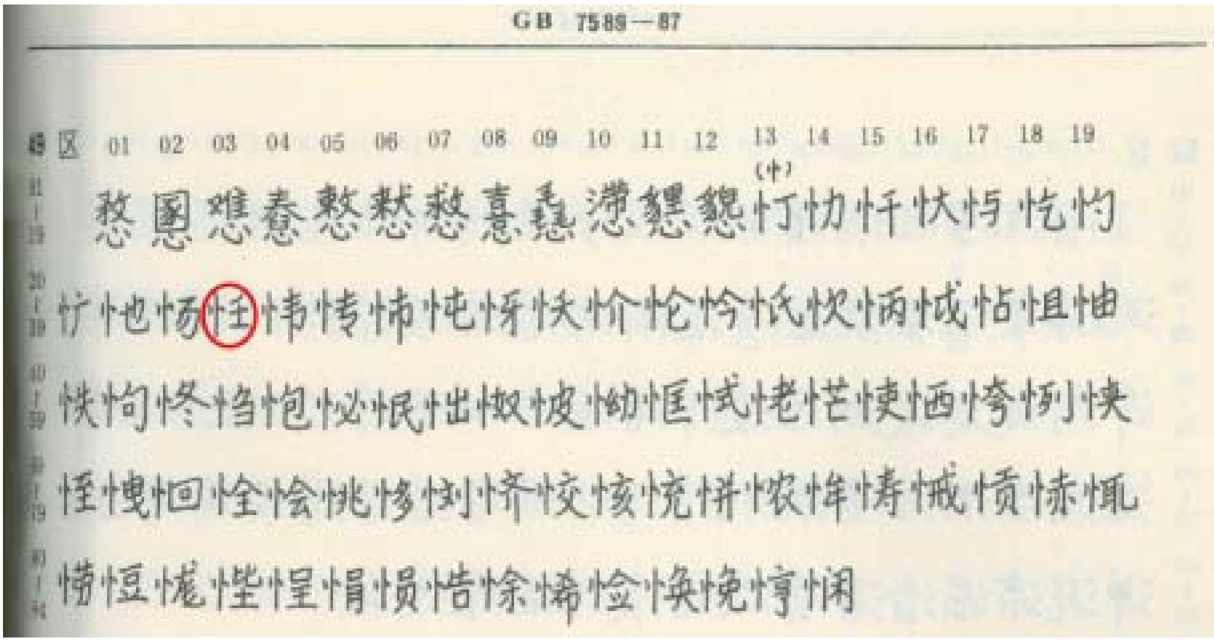
C03617 【忹】

字號	C03617	正字	【忹】 心-04-07
音讀	(一)ㄨㄤˇ　(二)ㄎㄨㄤˊ		
釋義	(一)ㄨㄤˇ 邪曲。見四聲篇海．心部。 姓。見四聲篇海．心部。 (二)ㄎㄨㄤˊ 「狂」之異體。		

(see http://140.111.1.40/yitic/frc/frc03617.htm)

1.3 The glyph, pronunciations and meanings of G3-5137

The image below shows the glyph of U+5FF9 in G3 (GB 7589-87)

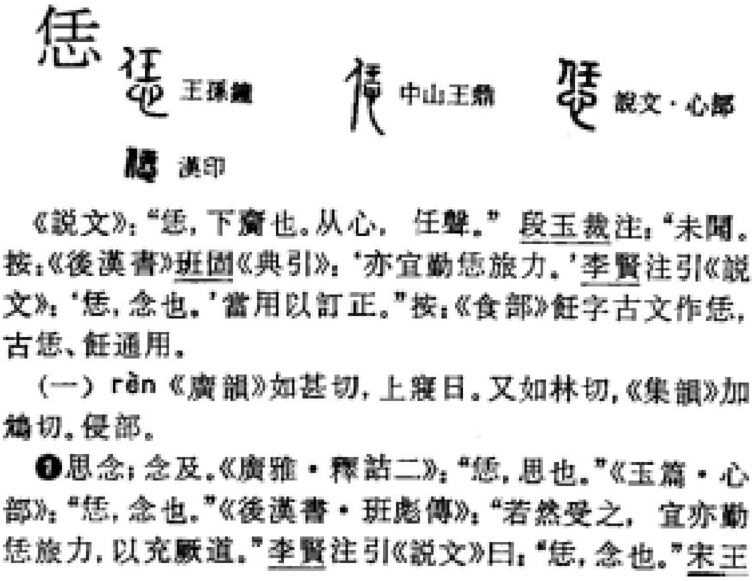


GB 7589—87

The character was selected from Page 2277 of ***Hanyu Da Zidian*** (漢語大字典

同"恁"。《正字通・心部》："[illegible]，同恁。"

It is considered as a variant of 恁 (U+6041) which is pronounced "ren" (meaning "thinking about"). See page 2291 of ***Hanyu Da Zidian*** (漢語大字典):



恁

王孫鐘

中山王鼎

說文・心部

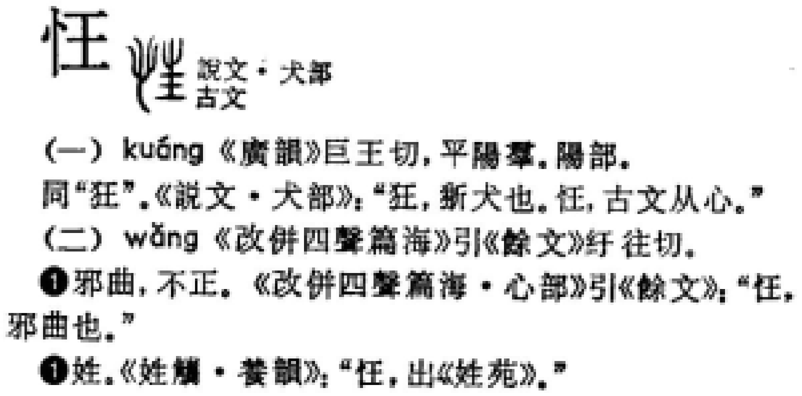
漢印

《說文》："恁，下齎也。从心，任聲。"段玉裁注："未聞。按：《後漢書》班固《典引》：'亦宜勤恁旅力。'李賢注引《說文》：'恁，念也。'當用以訂正。"按：《食部》飪字古文作恁，古恁、飪通用。

（一）rèn《廣韻》如甚切，上寢日。又如林切，《集韻》如鴆切。侵部。

❶思念；念及。《廣雅・釋詁二》："恁，思也。"《玉篇・心部》："恁，念也。"《後漢書・班彪傳》："若然受之，宜亦勤恁旅力，以充厥道。"李賢注引《說文》曰："恁，念也。"宋王

While the character 忹 (currently coded in U+5FF9) is pronounced "kuang" (meaning "crazy") or "wang" (meaning "evil, irregular"), see page 2275 of ***Hanyu Da Zidian*** (漢語大字典):



忹 說文・犬部 古文

（一） kuǎng 《廣韻》巨王切，平陽羣。陽部。

同“狂”。《説文・犬部》：“狂，狾犬也。忹，古文从心。”

（二） wǎng 《改併四聲篇海》引《餘文》纡往切。

❶邪曲，不正。《改併四聲篇海・心部》引《餘文》：“忹，邪曲也。”

❶姓。《姓觿・養韻》：“忹，出《姓苑》。”

2 Proposed Solution

2.1 Change the glyphs in G-column and T-column of U+5FF9:

In order to replce the wrong glyphs (both 忹) in the code point U+5FF9 with the correct glyphs, TCA and China will submit the correct glyphs of G3-5137 and T3-2623 (both 忹)to IRG and WG2.

2.2 Remove the glyph from U+2F89F and remove the entry “U+2F89F” from the electronic insertion file “CJKC_SR.txt” of ISO/IEC 10646:2003.

2.3 A new code point U+XXXX is assigned for encoding the character T5-2438 (忹), and insert a corresponding entry into the electronic insertion file “CJKU_SR.txt” of ISO/IEC 10646:2003.

附件 8

N3829：Emoji ad hoc report

Title: Emoji Ad-Hoc Meeting Report

Date: 2010-4-21
Source: Emoji Ad-hoc committee

An ad-hoc committee on Emoji encoding met in San Jose on April 20–21, 2010. The following were in attendance:

> Deborah Anderson, Sun Bojun, Peter Constable, Craig Cummings, Mark Davis, Peter Edberg, Michael Everson, Taichi Kawabata, Yasuo Kida, Mike Ksar, Yoshi Mikami, Katsuhiko Momoi, Lisa Moore, Eric Muller, : Katsuhiro Ogata, Roozbeh Pournader, Markus Scherer, Masahiro Sekiguchi, Michel Suignard, V. S. (Uma) Umamaheswaran, Ken Whistler, Chen Zhuang

The ad-hoc meeting was chaired by Peter Constable.

The ad-hoc discussed several open issues from national body comments on FPDAM8 related to Emoji, plus additional expert or national-body contributions on Emoji received by WG2 (N3769, N3776, N3777, N3778, N3785, N3806). Through these deliberations, the ad-hoc was able to arrive at consensus on many issues that should open the way for WG2 to resolve ballot comments on FPDAM8.

Note: in discussion that follows, references to FPDAM8 ballot comments use numbering as used in the "Draft Disposition of Comments…" (N3792).

Also, some of the recommendations given here make reference to document N3826, which shows a revision of the Emoticons block. Note that the revisions reflected in N3826 include a re-positioning of some characters in that block, but that the names list that is given in that document includes character annotations showing the character code positions in FDPAM8. In this document, all code positions and character names are those found in FPDAM8 unless explicitly stated otherwise.

The following describes the consensus decisions and recommendations of the Emoji ad-hoc committee:

1. In reference to Japan's comment T11c and item 8 in N3776, it is recommended that the glyph for U+1F35C STEAMING BOWL be changed to a glyph that shows a wide bowl with steam, but no noodles or chopsticks.
2. In reference to Germany's comment T2, it is recommended that the name for U+1F471 be changed from WESTERN PERSON to PERSON WITH BLOND HAIR.
3. In reference to Ireland's comment E3, Japan's comment T10b and corresponding comments in section 2 of N3778, US comment T4d, and additional expert discussion during the ad hoc committee meeting: it is recommended that the glyphs for the following 20 characters in the Emoticons block be changed:

1F601	ANGUISHED FACE
1F606	EXPRESSIONLESS FACE
1F608	FACE WITH LOOK OF TRIUMPH
1F60C	FACE THROWING A KISS
1F60D	FACE KISSING
1F612	HAPPY FACE WITH OPEN MOUTH AND CLOSED EYES
1F614	HAPPY AND CRYING FACE
1F615	HAPPY FACE WITH WIDE MOUTH AND RAISED EYEBROWS
1F616	HAPPY FACE WITH OPEN MOUTH AND RAISED EYEBROWS
1F619	FEARFUL FACE

1F61C	RELIEVED FACE
1F61D	CONFOUNDED FACE
1F61F	FACE SCREAMING IN FEAR
1F620	SLEEPY FACE
1F621	SMIRKING FACE
1F624	TIRED FACE
1F62D	HAPPY AND CRYING CAT FACE
1F62E	CAT FACE KISSING
1F631	POUTING CAT FACE
1F633	ANGUISHED CAT FACE

The recommended glyphs for these characters are shown in N3826.

4. In reference to Ireland’s comment T4, Japan’s comment T10a and corresponding comments in section 3 of N3778, US comment T4b, and additional expert discussion during the ad hoc committee meeting: it is recommended that the names of the following 27 characters in the Emoticons block be changed:

1F601	ANGUISHED FACE
1F605	EXASPERATED FACE
1F606	EXPRESSIONLESS FACE
1F607	FACE WITH HEART-SHAPED EYES
1F609	WINKING FACE WITH STUCK-OUT TONGUE
1F60A	FACE WITH STUCK-OUT TONGUE
1F60C	FACE THROWING A KISS
1F60D	FACE KISSING
1F60E	FACE WITH MASK
1F610	HAPPY FACE WITH OPEN MOUTH
1F611	HAPPY FACE WITH OPEN MOUTH AND COLD SWEAT
1F612	HAPPY FACE WITH OPEN MOUTH AND CLOSED EYES
1F613	HAPPY FACE WITH GRIN
1F614	HAPPY AND CRYING FACE
1F615	HAPPY FACE WITH WIDE MOUTH AND RAISED EYEBROWS
1F616	HAPPY FACE WITH OPEN MOUTH AND RAISED EYEBROWS
1F626	FACE WITH HALO
1F627	FACE WITH HORNS
1F628	FACE WITH SUNGLASSES
1F62B	CAT FACE WITH OPEN MOUTH
1F62C	HAPPY CAT FACE WITH GRIN
1F62D	HAPPY AND CRYING CAT FACE
1F62E	CAT FACE KISSING
1F62F	CAT FACE WITH HEART-SHAPED EYES
1F632	CAT FACE WITH TIGHTLY-CLOSED LIPS
1F633	ANGUISHED CAT FACE
1F63A	PERSON RAISING ONE HAND

The recommended replacement names for these characters are provided in N3826.

5. In reference to Ireland’s comment T6: it is recommended that characters in the Emoticons block be re-ordered as reflected in N3826.
6. In reference to Ireland’s comment T5, US comment T4a, and document N3769: it is recommended that one additional character be added in the Emoticons block in FDAM8:

   1F610 NEUTRAL FACE

   The representative glyph for this character is to be as shown in N3826.
7. In Ireland’s comment T5, Ireland requests addition of 13 other characters. It is recommended that these characters *not* be added in FDAM8, but that these be added to a bucket for consideration in a future amendment and that spaces be left open in the Emoticons block in anticipation of possible addition of these characters in the future, as reflected in N3826. The characters for addition to a future amendment are:

1F600	GRINNING FACE
1F611	EXPRESSIONLESS FACE
1F615	CONFUSED FACE
1F617	KISSING FACE
1F619	KISSING FACE WITH SMILING EYES[1]
1F61B	FACE WITH STUCK-OUT TONGUE
1F61F	WORRIED FACE
1F626	FROWNING FACE WITH OPEN MOUTH
1F627	ANGUISHED FACE[2]
1F62C	GRIMACING FACE
1F62E	FACE WITH OPEN MOUTH
1F62F	HUSHED FACE
1F634	SLEEPING FACE

Glyphs for these characters can be found in Ireland’s ballot comments—though code positions and some names from that document are to be revised in accordance with decisions made in the Emoji ad hoc committee. Glyphs for these characters can also be seen in N3834, which uses names and code positions listed here.

Rationale: experts in the ad hoc committee felt that these would probably make good additions in the Emoticons block but that that these should be processed in the context of a new amendment to allow opportunity for national body review.

8. It is recommended that the names list for the Emoticons block divide characters under three headings:

---

[1] In Ireland’s ballot comments, the name KISSING FACE WITH SQUINTING EYES was used, but this was revised in discussion in the ad hoc committee to KISSING FACE WITH SMILING EYES.

[2] In Ireland’s ballot comments, the proposed additional character given was WEARY FACE; in discussion in the ad hoc committee, names were revised such that the new character proposed by Ireland has the name ANGUISHED FACE.

Faces
Cat Faces
Gesture Symbols

9. Reflecting the combined recommendations in 3–7 above, it is recommended that the inventory, code positions, names and glyphs for characters in the Emoticons block in FDAM8 be as given in N3826. Also headings in the names list should be as in N3826; annotations in N3826 that reference “e-“ identifiers and FPDAM8 code positions should not be included.
10. In reference to Japan’s comments T11b, T12 and corresponding comments in item 7 of N3776: it is recommended that the row in EmojiSrc.txt for U+27BF DOUBLE CURLY LOOP (giving DoCoMo and Softbank source mappings) be deleted.
11. In reference to Japan’s comment T11g and corresponding comments in section 3 of N3777: it is recommended that *no change* be made in EmojiSrc.txt for U+267B BLACK UNIVERSAL RECYCLING SYMBOL.

    Rationale: the KDDI source mapping for this character should be kept to maintain consistency with vendor cross-mappings used in existing implementations.
12. In reference to Japan’s comment T11h, as it pertains to KDDI F65A, and to corresponding comments in section 3 of N3777: it is recommended that the vendor mappings to U+2762 HEAVY EXCLAMATION MARK ORNAMENT be changed to map instead to U+2757 HEAVY EXCLAMATION MARK SYMBOL. That is, that the row in EmojiSrc.txt

    ```
    2762;F9A7;F65A;F961
    ```

    be deleted and replaced by the row

    ```
    2757;F9A7;F65A;F961
    ```
13. In reference to Japan’s comment T11h, as it pertains to KDDI F6D5 and F6D3, and to corresponding comments in section 3 of N3777: it is recommended that KDDI mappings to U+1F466 BOY and U+1F467 GIRL be changed to map instead to U+1F468 MAN and U+1F469 WOMAN. That is, that the four rows in EmojiSrc.txt

    ```
    1F466;;F6D5;F941
    1F467;;F6D3;F942
    1F468;;;F944
    1F469;;;F945
    ```

    be revised as follows:

    ```
    1F466;;;F941
    1F467;;;F942
    1F468;; F6D5;F944
    1F469;; F6D3;F945
    ```
14. In reference to Japan’s comment T11h, as it pertains to KDDI F6EA, and to corresponding comments in section 3 of N3777: it is recommended that *no change* be made in EmojiSrc.txt for U+1F50A SPEAKER WITH THREE SOUND WAVES.

    Rationale: the KDDI source mapping for this character should be kept to maintain consistency with vendor cross-mappings used in existing implementations.

15. In reference to Japan's comment T11h, as it pertains to KDDI F688, and to corresponding comments in section 3 of N3777: it is recommended that the vendor mappings to U+1F68D ONCOMING BUS be changed to map instead to U+1F68C BUS. That is, that the row in EmojiSrc.txt

    1F68D;F8C1;F688;F79A

    be deleted and replaced by the row

    1F68C;F8C1;F688;F79A

Note: the Emoji ad hoc committee met after some discussion of national body comments had taken place in the WG2 plenary, including discussion of disposition for various other comments relating to the Emoji repertoire in FPDAM8 beyond those discussed here. The Emoji ad hoc committee expresses no recommendations regarding disposition of national body comments related to Emoji beyond the recommendations listed above.

The chair for the Emoji ad hoc committee wishes to express thanks to Michael Everson for timely preparation of document N3826, and to all the participants of the Emoji ad hoc committees at this and prior WG2 meetings for their dedication and collaborative work to encode the Emoji repertoire in Amendment 8.

附件 9

N3833：Ad hoc report on Tangut

JTC1/SC2/WG2 N3833

Tangut Ad hoc report
SOURCE: Deborah Anderson, SEI, UC Berkeley
DATE: 21 April 2010

Participants of the WG2 Tangut ad hoc meeting included: Deborah Anderson, Tom Bishop, Sun Bojun, Hsing Kuang CHEN, Suh-Chyin CHUANG, Peter Constable, Richard Cook, Michael Everson, Erich Fickle, KIM Kyongsok, Mike Ksar, Alain Labonté, Rick McGowan, Yoshi Mikami, Katsuhiko Momoi, Lisa Moore, Eric Muller, Roozbeh Pournader, Markus Scherer, Masahiro Sekiguchi, Michel Suignard, Bear TSENG, V. S. (Uma) Umamaheswaran, Ken Whistler, Jing Yongshi, and Chen Zhuang

Participating by phone: Nathan Hill and Andrew West

The members of the ad hoc met and discussed two issues that were raised in document N3821 (“Comments on Tangut proposal N3797”).

1. Naming

Ken Whistler discussed the reasoning for using names derived algorithmically from code points, and the burden that having catalog-based names could present for implementers and standardizers, particularly for such a large set of characters. China and Ireland agreed with the US on the use of such names. (UK did not comment.)

Recommendation: The names should be algorithmically derived from code points.

2. Repertoire

Andrew West (UK) described the current repertoire, which now has consists of one large block of 6055 characters, with a second, supplementary block of approximately 300 characters to be proposed later. Ireland recommended combining the two blocks into one, which will facilitate better behavior in binary sorting. However, by having two blocks, the supplemental characters will not sort correctly without explicit tailoring.

Recommendation: A single block of characters is advisable, incorporating as many of the 300 supplementary characters as is feasible and in an agreed-upon order.

The various NBs and interested parties will continue to work together and it is hoped a proposal agreeable to all can be presented at the next WG2 meeting. While the recommendations above were not unanimous, the ad hoc meeting felt this represented a productive direction to follow.

附件 10

N3817：Comments on Jurchen

JTC1/SC2/WG2 N3817
2010-04-15

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type:	**Working Group Document**
Title:	**Review of Revised Proposal to Encode Jurchen (N3788)**
Source:	**Andrew West**
Status:	**Individual Contribution**
Action:	**For consideration by JTC1/SC2/WG2 and UTC**
Date:	**2010-04-15**

---

1. Character Names

A. Errors in the Source Dictionary

The following corrections of errors in the phonetic transcriptions in the source dictionary should be reflected in the character names:

- 012-02 X-0051 (湯古) *tanŋgu* should be *taŋgu*
- 021-05 X-0108 (湯古) *tanŋgu* should be *taŋgu*
- 023-01 X-0114 (希兒) *çïr* should be *çir*
- 041-05 X-0194 (只里) *ʤiri* should be *ʤïri*
- 045-03 X-0216 (哈称) *xaʧi* should be *xaʧï*
- 084-07 X-0405 (間) *giyen* should be *gijen*
- 118-01 X-0557 (先) *ʃiɛn* should be *çiɛn*
- 138-02 X-0639 (謙) *ʧiɛn* should be *ʨiɛn*
- 168-02 X-0766 (申) *ʃin* should be *ʃïn*
- 169-03 X-0771 *futici*? should be *fuʧiçi*
- 178-01 X-0809 *futici*? should be *fuʧiçi*
- 178-02a X-0810a *futici*? should be *fuʧiçi*
- 207-01 X-0952 (只速) *ʤisu* should be *ʤïsu*
- 223-01 X-1040 (厥) *kiye* should be *kije*

B. Simplification of the Phonetic Transcription

The phonetic reconstructions used in the source dictionary are strictly phonetic, and so use different phonetic symbols to represent the same phoneme in different contexts.

The following phonetic symbols only occur before back and mid vowels:

- *ʃ* before *a*, *o*, *u*, *ə*, *ï*
- *ʧ* before *a*, *o*, *u*, *ə*, *ï*
- *ʤ* before *a*, *o*, *u*, *ə*, *ï*
- *ʒ* before *u*

Whereas the following corresponding phonetic symbols only occur before front vowels:

- *ç* before *i*, *y*
- *ʨ* before *i*
- *ʥ* before *i*, *y*

The vowels *e* and *ə* are also in complementary distribution:

- *e* only occurs as *ei*, *ie* and *je*
- *ə* never occurs before/after *i* or after *j*

We believe that there is no need to preserve such non-phonemic distinctions in the character names, and suggest representing pairs of phonetic symbols that represent the same phoneme and that are in complementary distribution (i.e. *ʃ*/*ç*, *ʧ*/*ʨ*, *ʤ*/*ʥ* and *e*/*ə*) using the same ASCII letter or letters, for example using 'E' for both *e* and *ə*.

C. Use of Accented Letters in Character Names

The character names proposed in N3788 include extended characters (e.g. Ś, Š, Ź, Ž, È), which are disallowed according to the character naming rules. We propose the following scheme for translating the phonetic transcription in the source dictionary to legal character names (all unlisted characters are unchanged):

- *e*/*ə* = E
- *ɛ* = AE
- *ŋ* = NG
- *ʥ*/*ʤ* = J
- *ʨ*/*ʧ* = C
- *ç*/*ʃ* = SH
- *ʒ* = ZH
- *j* = Y (in complementary distribution with 'Y' representing the vowel [y])

We prefer "J" and "C" to "DZH" and "TSH" as Manchu romanization uses "j" and "c", so the resultant Jurchen names appear much closer to their Manchu cognates.

2. Wrongly Ordered Characters

The following characters appear to be misordered:

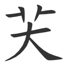
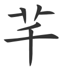
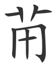
- X-0663b (Radical 7 / 4 strokes) should be moved to after X-0218
- X-0894a (Radical 32 / 6 strokes) should be moved to after X-0931d
- X-0454 (Radical 22 / 8 strokes) and X-0455 (Radical 21 / 8 strokes) should be swapped
- X-0993 (214-08) and X-0994 (214-09) should be swapped

3. Duplicate Characters

N3788 proposes encoding the character in entry 048-03 (X-0231a and X-0231b) twice as it is a variant form of two different characters (048-02 and 051-01). We believe that this character should not be encoded twice.

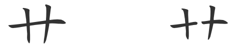
4. Radical 21

The glyph form of Radical 21 does not exactly match the shape of the radical in the characters under radical 21. The radical glyph has crossing horizontal strokes, whereas the characters under this radical do not:

- 092-05 X-0452 
- 092-06 X-0453 
- 093-02 X-0455 

We believe that although the difference is minor, it would be best to change the glyph for Radical 21 to match its actual appearance so that it is more distinct from Radical 20:

Radical 20 Radical 21

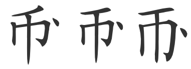
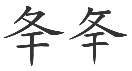
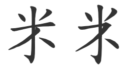
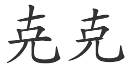


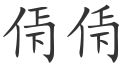
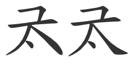
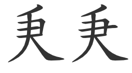
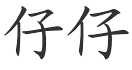
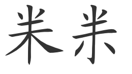
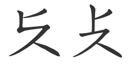
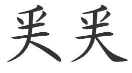
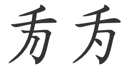
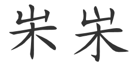
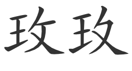
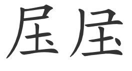
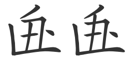
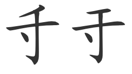
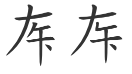
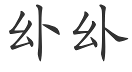
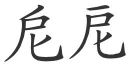
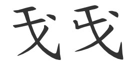
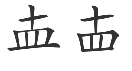
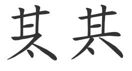
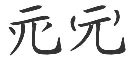
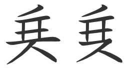
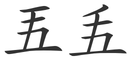
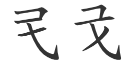
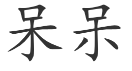
5. Glyph Variants

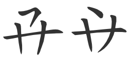
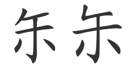
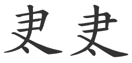
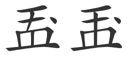
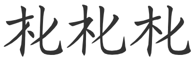
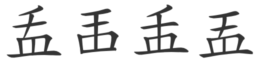
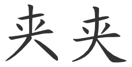
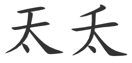
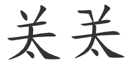
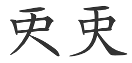
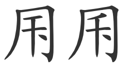
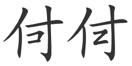
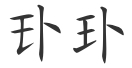
The revised proposal (N3788) proposes 1,430 characters for encoding, of which:

- 523 characters have no variant forms (including nearly a hundred characters with unknown readings that may in fact be variants of other characters proposed for encoding)
- 368 characters are the primary form of a character with multiple glyph variants
- 539 characters are glyph variants

Thus, almost 40% of the proposed Jurchen characters are glyph variants, which is a considerably higher proportion than for any other script encoded or proposed for encoding in the UCS. Moreover, very many of the glyph variants proposed for encoding show very insignificant differences, often just reflecting a slightly different way of writing the same stroke in different manuscript sources, as for example:

AI	
AMBA	
AN	
AXU	

BIRA	
BUXA	
CII	
CUEN	
DAI	
ESE	
FUN	
GE	
FI	
FO	
GE	
GU	
I	
INDA	
IU	
JAL	
JO	
JU	
KI	
MA	
MI	
MINGGAN	
MO	
MUA	

NADAN	
NAN	
O	
SHI	
SHIA	
SHII	
SHIIR	
TAIYI	
TU	
U	
UJE	
XA	
YA	

It may be appropriate to separately encode some significant character variants, but it may be more appropriate to represent insignificant glyph variants by means of variation sequences.

In several respects, the nature of the Jurchen character variants is significantly different to that of Tangut character variants, which means that although variation sequences are not appropriate in the case of Tangut, they may be the most appropriate solution for Jurchen:

Tangut	**Jurchen**
Relatively few variant characters (about 120 out of the 6,054 characters proposed for the main Tangut block)	Relatively many variant characters (539 out of 1,430 proposed characters)
Most variants are attested in more than one modern Tangut dictionary	Most variants are only used in Jin Qicong's dictionary
Most Tangut variants are distinguished by a different structural composition (one or more different component elements, or the same component elements arranged differently)	Most Jurchen variants are distinguished only by a different way of writing the same stroke (e.g. the angle or length of a stroke) or a different placement of a stroke or a difference in stroke count