IRG Meeting #32 Hong Kong SAR, China

P&P Drafting Group Report

Reference documents: IRGN 1562(P&PV3.0 draft), IRGN1562 (HK Feedback), IRGN 1593(Feedback John Knightley), and IRGN1561 FeedbackUma,

1.General feedbacks:

The Group discussed feedbacks from HK and Uma Umamaheswaran, and agreed on the editorial notes included in IRGN1590AppendixA.

2. The group has discussed the suggestion of making a standing document series(IRGN 1562) in accordance with the agreement to maintain an up-to-date IRG lists of unification examples, dis-unifcation examples, over-unification examples, and the UCV list. It is agreed that such a series should be created. It is also important that the examples given in the current Annex S are also included in this IRG standing document series.

The group agreed to appoint a Standing Document Editor(SDE) and has nominated Mr. Taichi Kawabata to be the editor.

Detailed specification(IRGN 1600) for the file naming conventions, file formats, example naming conventions, classification of examples, and the review procedure will be proceeded according the following schedule:

July 31, 2009: Draft 1 of Specification of IRG Standing document Series ready by SDE to members
Aug. 28, 2009: Feedbacks from members to SDE
Sept. 25, 2009: Draft 2 ready by SDE for distribution to members
Oct. 23, 2009: Feedbacks to Draft 2 by members to SDE.
Nov. 13, 2009 Draft 3 ready by for its finalization in IRG # 33

3. The document IRGN 1593 on Record keeping Examples as part of P&P process was discussed. The group accepts the suggestion in principle,

and would like to evaluate its practicality by first practicing it for one or two IRG Meetings. With the establishment of the IRG standing document series, it is recommended that the suggestion be considered at a later time for inclusion as a part of IRG P&P.

4. The following is the agreed IRG P&P schedule:

Sept. 17, 2009 P&P V.3 ready by P&P editor for distribution
Oct. 23, 2009   Feedbacks from member bodies to P&P editor
Nov. 13, 2009 P&P V.4 draft ready by P&P editor for discussion at IRG #33.

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

Universal Multiple-Octet Coded Character Set
(UCS)

ISO/IEC JTC 1/SC 2/WG 2/IRG *N1562*
2009-06-03

Title:	**IRG Principles and Procedures Version 3 Draft 1**
Source:	**IRG P&P Drafting Group**
Action:	**For review by the IRG and WG2**
Distribution:	**IRG Members and Ideographic Experts**
References:	**IRGN1561(P&PV2 and all feedbacks),**
	IRN1559(P&P V2Draft and all feedbacks),
	IRGN 1465(P&P Draft1),
	IRGN 1487(Feedback from HKSARG),
	IRGN 1489(Feedback from Taichi Kawabata),
	IRGN 1516(Feedback from HKSARG)

Table of Contents

| Deleted: 5 |
| Deleted: 5 |
| Deleted: 6 |
| Deleted: 7 |
| Deleted: 7 |
| Deleted: 9 |
| Deleted: 9 |

1. Introduction

This document is a standing document of ISO/IEC JTC 1/SC 2/WG 2/IRG for 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 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 Characters 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 determine 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 proposals with justification. A rejected proposal may be revised and resubmitted to the IRG.

Deleted: projects can

Deleted: in IRG

Deleted: .

Deleted:

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.

  '戉'(U+620C) and '戊'(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 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 recorded examples, 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 for 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 standardized ideographs**. 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 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, IRG will not impose a limit on the collection size. However, to minimize the burden of the checking process and to achieve a higher quality of standard within a reasonably short period of time, review by IRG members normally cannot exceed 4,000 ideographs. Based on this principle, members 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 nature of the 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, FPDAM and FDAM)
c) Ideographs currently in IRG working sets including both M-sets and D-sets.
d) Ideographs mis-unified or over-unified with ideographs in current standard based on the list maintained by the IRG.

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:

IRGN*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. C labelled by "_C").

(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 as an IRG standing 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. ID should begin with a member body code (G,T,J,K,V,KP,H,M, MY, or U) followed by no more than 9 characters and should contain only Latin capital letters, Arabic numbers, and hyphens. 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 N 1105)..
e. ***Flag to show whether the ideograph is traditional (0) or simplified (1).***
f. ***First Stroke Code*** of the Non-radical Component (ref. IRG N 954 AR and IRG N 1105).

g. ***Ideographic Description Sequence*** (ref. IRGN 1183).
h. ***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.
i. ***References to evidence document*** including document number and page number.

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.
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 without grounds of 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 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 can be rejected by the technical editor for production of a working draft.

2.3.2. Principles on Assignment of Serial Number

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 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 impact on other characters in the collection. Thus it is generally not allowed. However, members may submit minor alteration of characters with provision of justification ONLY at the final stage as long as alteration is unifiable to the original. 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 checking and verifying that the alteration does not affect other characters in existing standards and working sets must be supplied. 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 requester 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 existing D-set.

g.   After consolidation, the IRG chief editor and technical editor may ask members 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, Radical, 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, members 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 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 ideographs with radical X and stroke 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, in 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 will 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

Deleted: G

Deleted: [Does this apply to D Set only or both to D Set and M Set?]…

Deleted: [Does this apply to D Set only or both to D Set and M Set?]…

Deleted: Unification

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 members 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 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 members 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 of 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.

Deleted: s

    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 technical editor may ask the submitter to resubmit or supply the additional information if only minor problems are encountered. Otherwise, the submission can 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. Members 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 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.   Members 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 conclusion 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 would 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 document 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.   Members 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.   Members 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.   Members 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 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. Members should review the comments on M-set and reach conclusion 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 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 the 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+xxxx and requester will request new Source ID to U+xxxx. • Unified with U+xxxx and requester will request that this character be treated as a Compatibility Ideograph. • Unified to U+xxxx 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 requester may consider registering it as a Compatibility Character or IVS. • Not Unifiable.

<div style="color: red; border: 1px solid red">

Deleted: make

Deleted:

Deleted: -set

Deleted: they are

Deleted: it to

</div>

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 ideographs.   For the ideographs that cannot be machine-checked by IDS matching, at least two non-requester members must check them manually to ensure that the ideographs are not unifiable with any standardized ideograph or working ideograph.   For the ideographs that might be unifiable with other ideographs, a submitter 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*. Entry is no longer used. ● 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* and this entry is no longer used. ● Unified with *xxxxx*. (*xxxxx* is removed.) ● Not unifiable, as shown by the evidence *IRG N xxxx*. Move to M-set
Checked against all standardized and working ideographs with radical *X* and stroke of *Y*±2.	● Moved to M-set, as two non-submitter members (*XX* and *YY*) confirmed that this ideograph is not unifiable with any existing standardized or working 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 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).

Deleted: -s

Deleted: D-set i

Formatted: Not Highlight

Deleted: the

Deleted:

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 Dr. 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.

Simplified Radicals		Traditional Radicals	
R119. 1	纟	R119. 0	糸
R146. 1	见	R146. 0	見
R148. 1	讠	R148. 0	言
R153. 1	贝	R153. 0	貝
R158. 1	车	R158. 0	車
R166. 1	钅	R166. 0	金
R167. 1	长	R167. 0	長
R168. 1	门	R168. 0	門
R177. 1	韦	R177. 0	韋
R180. 1	页	R180. 0	頁
R181. 1	风	R181. 0	風
R182. 1	飞	R182. 0	飛
R183. 1	饣	R183. 0	食
R186. 1	马	R186. 0	馬
R194. 1	鱼	R194. 0	魚
R195. 1	鸟	R195. 0	鳥
R196. 1	卤	R196. 0	鹵
R198. 1	麦	R198. 0	麥
R204. 1	黾	R204. 0	黽
R209. 1	齐	R209. 0	齊
R210. 1	齿	R210. 0	齒
R211. 1	龙	R211. 0	龍

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

c. **First stroke**.

> **Commented [h7]:** Examples to be added later.

Annex B: IDS Matching

B.1. Guidelines on Creation of IDS

Each member body should consult IRGN1183 on IDS creation finalized at IRG Meeting No. 25. 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. ⿰ネ申 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.

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.
    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

D1. Member body Code:
   G: China
   H: Hong Kong Special Administrative Region
   J: Japan
   K: Republic of Korea
   KP: Democratic People's Republic of Korea
   M: Macao Special Administrative Region
   MY: Malaysia (Added in Nov. 2008 in IRG Meeting No. 31)
   T: Taipei Computer Association
   U: Unicode Consortium/USA?
   V: Vietnam

D2. 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 （方正排版系統）
   G_4K Siku Quanshu （四庫全書）
D3. Hanzi H sources
   Hong Kong Supplementary Character Set(HKSCS)
D4. 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 forth 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
D5. 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
D6. Hanja K sources are
   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
D7. Hanja KP sources
   KP0 KPS 9566-97
   KP1 KPS 10721-2000
D8. ChuNom V sources
   V0 TCVN 5773:1993

V1 TCVN 6056:1995
V2 VHN 01:1998
V3 VHN 02: 1998
D9. MY sources

Annex E: List of Over-unified and Mis-unified Ideographs

E.1 Introduction

There are some known mis-unification and over-unification cases in the ISO/IEC 10646. However, once encoded, over-unified and mis-unified characters should not be separately encoded unless agreed upon by WG2. Therefore, requesters must be careful that their proposal does not include such over-unified character. In this section, known over-unified and mis-unified ideographs are listed. Requesters should check against this section before submitting their proposed ideographs to the IRG.

E.2 List of Over-unified Ideographs

<<Insert Contents of IRG N 1483 Appendix B2>>

E.3 List of Mis-unification Ideographs

<<Insert Contents of IRG N 1395>>

Commented [h9]: Taichi: I only have pdf files, cannot do insert at this time. To IRG members: I suggest that we make a number of standing documents in IRG which can be updated from time to time. This should be one of them.
Editorial: this will be included in the IRG standing document.

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 the 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 for 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 2008, Amendment 1, 4, and 5 have the 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 2008, PDAM 6 has the 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 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* 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* with the ideographs in the current standardized or working set mentioned above?	
Attribute Data Checklist	
11. Do all the proposed ideographs have the radical code, stroke count and first stroke attributes in attribute data?	
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 similar/variant ideographs information for proposed ideographs?	

Deleted: check

WG2 Principles and Procedures Annex I: Guideline for Handling of CJK Ideograph Unification or Disunification Errors

(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 disunified* 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 WG 2 to deal with them.

I.1 Guideline for "To Be Unified" Errors

A. The "*to be unified*" pair will be left disunified. 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 Disunified" Errors

A. The ideographs to be disunified should be disunified and should be given separate code positions as soon as possible (disunification 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 disunified 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 Disunification Request

There is a possibility of "pure true disunification" 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 DISUNIFIED" and "WILL NOT BE DISUNIFIED is as follows.

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

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

WG2 PnP Annex J: Guideline for Correction of CJK Ideograph Mapping Table Errors

(Source: <u>ISO/IEC JTC 1/SC 2/WG 2 N2577</u> – 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 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 WG 2 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 WG 2
    There are many mapping tables, which are included in national/regional standards or developed by third parties.   These are out of WG 2'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. Lu Qin).

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: [to be updated later]

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.

> **Commented [A10]:** Add Clause No. in ISO10646

Abstract shape:

D-set:

M-set:

Working set:

Compatibility characters:

Ideographic Description Sequence(IDS):

New Source: ISO reference