ISO/IEC JTC1/SC2/WG2/IRG N1915
Doc. #: Korea JTC1/SC2 K2181D

Korea JTC1/SC2, Committee on Character Codes

Author: KIM, Kyongsok, Chairperson of Korea JTC1/SC2
Date: 2013.05.17.
Status: NB position
Subject: How to reduce # candidate chars to 4000 for IRG review

1. In general, the total number of characters submitted for each project (e.g., ExtG) from all MBs/IOs can exceed 4000.

- According to IRG PnP, IRG needs to reduce the number of candidate characters to 4000 for IRG review.

- A method is suggested below on reducing # of candidate characters to 4000.

2. How to reduce # candidate chars to 4000 for IRG review

- It is suggested that each MB/IO can submit at most 1,000 chars for each project (e.g., ExtG). Based on this assumption, a method to reduce the number of candidate chars to 4000 is explained below step by step.

2.1 Calculate the total number of chars submitted

- Suppose that, as an example, the number of characters submitted by MBs/IOs is as follows (see column S). Assume that MBs are sorted in ascending order of number of chars submitted.

	subm	step 1	step 2	step 3	step4	[final step]
		(200)	*(600)*	*(700)*	*(900)*	*(900)*
MB1:	200	200	200	200	200	200
MB2:	600	200	600	600	600	600
MB3:	700	200	600	700	700	700
MB4:	900	200	600	700	900	833
MB5:	1000	200	600	700	900	833
MB6:	1000	200	600	700	900	833
total	4400	1200	3200	3600	4200	3999

2.2 determining the max. number of chars for each MB

- Principle: Assign one char to each of six MBs at each round while the total number of chars assigned to 6 MBs is less than or equal to 4,000.

- **step 1**: Try to assign 200 (the number of chars submitted by MB1) chars for all six MBs. It is shown in "step 1" column. The total number of chars assigned to 6 MBs is 1,200. It is less than 4,000. We are done with MB1.

- **step 2**: Try to assign 600 (the number of chars submitted by MB2) chars for the remaining five MBs (i.e., MB2 ~ MB6). It is shown in "step 2" column. The total number of chars assigned to 6 MBs is now 3,200. It is less than 4,000. We are done with MB2 (and MB1 in step 1).

- **step 3**: Try to assign 700 (the number of chars submitted by MB3) chars for the remaining four MBs (i.e., MB3 ~ MB6). It is shown in "step 3" column. The total number of chars assigned to 6 MBs is now 3,200. It is less than 4,000. We are done with MB3 (and MB1 and MB2 in previous steps).

- **step 4**: Try to assign 900 (the number of chars submitted by MB4) chars for the remaining three MBs (i.e., MB4 ~ MB6). It is shown in "step 4" column. The total number of chars assigned to 6 MBs is now 4,200. It is greater than 4,000. We need to step back and take a final step.

- **final step**: Assign 400 (= 4,000 - 3,600, which is the total number of chars assigned to 6 MBs in step 3) chars equally to the remaining three MBs (MB4 ~ MB6) in addition to 700 chars already assigned in step 3. In other words, assign 833 (= 700 + 400 / 3) chars to MB4, MB5, and MB6.
  It is shown in "final step" column. The total number of chars assigned to 6 MBs is now 3,999. We are finally done.

3. Parameters

3.1 There is ONE parameter that can be adjusted/determined by IRG in the above steps.

- The number of chars that each MB/IO can submit (1000 in the above example). This number seems reasonable; however, IRG can decide the value.

3.2 To implement the above method, each MB will assign a priority number for each char in the submission form (we need to modify IRG PnP by adding one more column to submission form). A priority number will start with 1 and the max. value is the same as the number of chars from each MB.

- In the above example, MB4 submitted 900 chars. Each char is allocated a priority number whose value is between 1 and 900 inclusive. Priority numbers will be assigned by MB/IO based on the importance of each char.

- According to the above method, only 833 chars from MB4 can become candidate chars for IRG review.

- Then 833 chars whose priority numbers are between 1 and 833 inclusive will become candidate chars for IRG review and the remaining 67 chars whose priority numbers are between 834 and 900 inclusive will be excluded from IRG review for this project (e.g. ExtG). Those 67 chars could be possibly submitted for ExtH later.

- Comments are welcome. Thanks.

＊＊＊

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2 IRG N1915 Feedback
Date: 2013-05-19

Source:	UK
Meeting:	IRG#40, Hong Kong SAR, China
Title:	The Submission Process for Future Extensions
Actions required:	For discussion
Distribution:	IRG
Medium:	Electronic
Pages:	2

Reflection on the WG2 submission process suggests ways of improving the IRG submission process. If IRG concludes that improvements are desirable, IRG could submit a request to WG2 for approval of changes to IRG procedure. Some possible improvements are suggested below.

1) Establish a fixed time frame for receiving new submissions.

The current IRG practice is that the times for receiving new submissions are decided on an ad hoc basis. The WG2 practice, on the other hand, is that the times for receiving submissions are clearly fixed and known well in advance. It would be desirable for the IRG to implement a system whereby new submissions would be accepted at a fixed time in a fixed cycle, say every 9th IRG meeting or every 10th IRG meeting, for example. This would be helpful to many members and should lead to better submissions. If, for example, new submissions were accepted at every 10th IRG meeting, this would correspond to the time frame required in IRG's recent experience. Fixing new submissions for every 8th or every 9th IRG meeting would also be possible. At the end of a cycle, only those characters fully checked would sent to WG2. This would ensure quality, and unresolved characters could be resubmitted if so desired to the next extension. Working to such a timetable would have many benefits.

2) Add a rewrite step instead of accepting or rejecting at the same meeting that submissions are made available.

In accordance with the current P&P, for extension F the specific content of submissions was made available at IRG 39 and submissions were either accepted or rejected at the same meeting. At WG2 it is common for submitters, especially inexperienced ones, to be asked to rewrite a

submission for the next meeting. It would be desirable for IRG to adopt this practice. This would mean adding a step to the submission process in which individual members would rewrite their submission for the next meeting after receiving feedback on IDS and comments from others. If submitters where required to either respond in writing for each queried character before the next meeting or withdraw that character there would be a considerable saving of time (questions relating to unification between submissions should not be included in this step).

3) There should be no restriction on the number of characters. Instead the restriction should be based on quality.

While in practice most submissions to WG2 are smaller than those to IRG, the WG2 does not limit the number of characters in a submission. The IRG, on the other hand, does. Current P&P guidelines on size are: "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." and there is also the retrospective 5% rule. An emphasis on quality rather than the size of a submission would streamline the process. For example, for Extension F there was an initial set with 8,511 characters. But if quality had been the main criterion for choosing submissions, the largest submission may well not have been included and the initial collection would have been only 5,096 ideographs. Such a collection would have been both of a better quality and of a more manageable size.