

Title: Considerations concerning ISO/IEC 10646 Code charts

Source: Michel Suignard, Project editor

Status: Individual Contribution

Distribution: WG2

Reference: WG2 N5253

Executive Summary:

The document WG2 N5253 'Consideration concerning ISO/IEC 10646 7th edition' was discussed at the last WG2 meeting but was not formally endorsed through a recommendation. This new document is requesting such a recommendation. In addition, additional consideration concerning the sharing of code charts between ISO/IEC 10646 and the Unicode Standard should be explored and possibly resulting in a recommendation to the project editor to create a common code chart resource between the two standards.

Development

First, it is important to remind us of what was considered in WG2 N5253 'Consideration concerning ISO/IEC 10646 7th edition'. That document recognized the large number of data files shared between the Unicode Standard and ISO/IEC 10646 and the duplication of clauses between these two standards. It recommended a simplification of the source descriptions (clause 24 to 28) by referring to Unicode Standard Annexes and using references to Unicode equivalent data sets where they were equivalent. A discussion happened on the topic at the last WG2 meeting in June 2024 and is described in WG2 5251 (WG2 meeting #71 minutes, page 13); it was however not recorded as a recommendation by the WG at the same meeting. In addition, some delays appeared in the creation of the required new Unicode Standard Annex covering some of these source descriptions (UAX#60), partly due to ongoing discussion concerning the exact nature of these script related properties and their syntax. For all these reasons, these suggestions have not been yet implemented in the current Committee Draft of the seventh edition.

However now, with UAX#60 being a draft Unicode Standard, it becomes possible to refer to it as its target completion (Unicode V18.0) is compatible with ISO/IEC 10646 7th edition target date.

The next step that should now be seriously considered is the sharing of the code charts between the two standards. For the same rationale calling to share references to the data set common to the standards, it makes sense to simplify the code chart context by simply referring to a common code chart resource between the two standards. Because Unicode iterates more often than ISO/IEC 10646 and its amendments, some code charts cannot be shared such as the 'versioned' version of the Unicode code chart which refers to addition to the previous version. But the Unicode Archived version of the code chart is in fact identical to the ISO/IEC 10646 format (not considering differences in header and footer). However, the Unicode version is also bookmarked which simplifies greatly the navigation through 3000+ pages of this resource. In addition, it would then be possible to explore the availability of code chart at the block level as done for the Unicode Standard (see for example: <https://www.unicode.org/charts/>). This would simplify greatly the document management of the ISO/IEC 10646 standard as its size would shrink from over 3300 pages to just over 100 pages.

It should be noted that the Unicode Standard does not include the code chart in its Core Specification and provides the related code chart as separate resource. Another consideration is that the code chart work for both standards is done by the Unicode Code chart editor and avoiding the duplication would simplify the task. In these times of shrinking resources, any streamlining of the process would be valuable.

The copyright status of these code charts should nevertheless be considered. On one hand, the Unicode Consortium is bearing most of the cost of the development of these code charts in terms of tooling and editing resources; on the other hand, the Unicode Consortium has made very clear the status of their own code charts for that specific usage (see Terms of Use shown below).

We all recognize that contributions to these standard are made by many parties, notably in terms of fonts. These contributions are made either by National Bodies, Liaison Organizations, other commercial or non-commercial entities or individual experts. The Unicode Consortium has spent significant resources streamlining the terms of use of these code charts. The current 'Terms of Use' attached to each Unicode code chart are duplicated below (the emphasis on 'ISO use' is only added here in this document, and that emphasis is not part of the regular version):

Terms of Use

© 1991–2025 Unicode, Inc. This publication is protected by copyright, and permission must be obtained from Unicode, Inc. prior to any reproduction, modification, or other use not permitted by the Terms of Use (<https://www.unicode.org/copyright.html>). Specifically, you may make copies of this publication and may annotate and translate it solely for personal or internal business purposes and not for public distribution, provided that any such permitted copies and modifications fully reproduce all copyright and other legal notices contained in the original. You may not make copies of or modifications to this publication for public distribution, or incorporate it in whole or in part into any product or publication without the express written permission of Unicode.

The Unicode Consortium specifically grants ISO a license to produce such code charts with their associated character names list to show the repertoire of characters for that standard, as a normatively referenced, integral part of that standard.

Unicode uses most fonts under restricted license from the original font owner. You may not extract, copy, modify, or distribute fonts or font data from any Unicode Products, including this publication, without license from the font owner.

Use of all Unicode Products, including this publication, is governed by the Unicode Terms of Use (<https://www.unicode.org/copyright.html>). The authors, contributors, and publishers have taken care in the preparation of this publication, but make no express or implied representation or warranty of any kind and assume no responsibility or liability for errors or omissions or for consequential or incidental damages that may arise therefrom. This publication is provided "AS-IS" without charge as a convenience to users.

Unicode and the Unicode Logo are registered trademarks of Unicode, Inc., in the United States and other countries.

Conclusion

Making these simplifications in the standard development process would greatly simplify the work of the Project Editor and this is the reason why the editor is requesting that such recommendations be done by the WG (that is endorsing WG2 N5253 and this new document). It would also facilitate any succession plan which is becoming an acute issue.

■ end of document-