

To: NCITS L2, Unicode Technical Committee
From: Edwin Hart (edwin.hart@jhuapl.edu, +1-443-778-6926)
Subject: Need For a Standard Name for the Collection of Planes 0 to 16

The Concern

What should we call the full repertoire of characters contained in planes 0 to 16, that is, the full repertoire addressable using UTF-16? I request that L2 and the UTC review this document, decide on the preferred name for the collection, and then forward a request to SC 2/WG 2 to formally standardize the name for the collection.

As we get closer to approving part 2 of ISO/IEC 10646 and assigning characters in planes 1 to 16, it is becoming increasingly important to have a name for the full collection in the same way that "BMP" names all of the characters in plane 0. In addition, the Unicode Consortium is in the process of approving a Technical Report that would describe support for the UCS-4 mode for just the repertoire of UTF-16. It would be valuable to have a standard name for the collection to be used in the Unicode TR. In summary, having a standard name for this repertoire would have value to both ISO and Unicode.

Code Positions Included in the Collection

Before discussing the name, one of the issues is: Which characters should be included in the collection? The collection includes all characters in planes 0 to 16, except the following code positions:

- Code positions 0xD800 through 0xDFFF are reserved in plane 0 (the BMP).
- Code positions 0xFFFE and 0xFFFF are reserved on all planes.

Criteria for Naming the Collection

Here are some criteria for the name:

- The name should be short.
- The name should describe the repertoire.
- The name should be independent of the coding so that the name does not imply how the repertoire is coded. (Since UTF-16, UTF-8 and UCS-4 can all address the repertoire, the name should not imply that the coding is UTF-16).
- A clever name would be nice but the name need not necessarily be clever.

Name for the Collection of Characters in Planes 0 to 16

In response to my request for feedback on this concern, people have responded with several possible names. Although most names were seriously proposed, a few are facetious and are included for completeness and amusement. (I will leave it to the reader to decide for whose

amusement some of the names were included.) I also hope that this does not precipitate an endless discussion to decide a name. Just pick the one that I like best and you'll be done.

- ~~1. APs: "Additional Planes"~~
- ~~2. AUC: "All Unicode Characters"~~
- ~~3. BMC: "Basic Multilingual Collection"~~
- ~~4. CSS: "Complete SubSet"~~
- ~~5. EMP: "Expanded Multilingual Planes"~~
- ~~6. EMP: "Extended Multilingual Planes"~~
- ~~7. EUCS: "Extensive Universal Character Set"~~
- ~~8. FMC: "Full Multilingual Collection"~~
- ~~9. FMS: "Full Multilingual SubSet"~~
- ~~10. GCS: "Galactic Character Set"~~
- ~~11. GCS: "Global Character Set"~~
- ~~12. GCS: "Global Character SubSet"~~
- ~~13. GS: "Global SubSet"~~
- ~~14. GSS: "Global SubSet"~~
- ~~15. MMC: "Multiscript Multilingual Collection"~~
- ~~16. PAU: "Planes Accessible by UTF-16"~~
- ~~17. PAU: "Planes for All of Unicode"~~
- ~~18. SCC: "Standard Character Collection"~~
- ~~19. SCP: "Standard Character Planes"~~
- ~~20. TGC: "The Global Collection"~~
- ~~21. UGC: "Unified Global Collection"~~
- ~~22. UCSGCC: "UCS Global Character Collection"~~
- ~~23. UCSGCS: "UCS Global Character SubSet"~~
- ~~24. UMA: "Universal Multilingual Aggregate"~~
- ~~25. UMC: "Universal Multilingual Collection"~~
- ~~26. UMSS: "Universal Multilingual SubSet"~~
- ~~27. 1.1Meg: "1.1 Mega characters"~~
- ~~28. "Characters 0 to 10FFFF" or "Characters 0..10FFFF"~~
- ~~29. C17 or CP17: Collection of first 17 planes~~
- ~~30. "Decihex Collection"~~
- ~~31. "les dix sept" or "Les Dix Sept" (French for "the seventeen")~~
- ~~32. "Planes 0 to 16" or "Planes 0..16"~~
- ~~33. P17: "17 Planes from 0 to 16"~~
- ~~34. PC17: "Plane Collection of first 17 planes"~~
- ~~35. "Suite 16"~~
- ~~36. "UTF 16 collection"~~
- ~~37. "UnicodeX"~~

In the table, I crossed out the names that I rejected in this paragraph. Here are the reasons for my rejections. I consider number 1 to be too vague. For those using the words "Unicode" or "complete" or "expanded" or "extended" or "full" or "global" or "universal" (numbers 2, 4 to 9, 11 to 14, 17, 20 to 26, 37), I worry about the confusion should, at some future time, we decide that we need to code more characters than will fit into 15 planes plus 2 private-use planes. Moreover, what should we then call the new collection? I would compare this situation to the EMS and XMS memory specifications for the early PCs. I can never associate the "extended" and "expanded" name with the correct acronym. I consider number 10 to be too cute. Number 15, with both the words "multiscript" and "multilingual", seems redundant to me. Numbers 18 and 19 appear pretentious because of the use of the name "standard". All of the collection names are "standard" collection names and part of the ISO/IEC 10646 standard. Also, the collection may need to be extended beyond plane 17, and what would we call it? Numbers 27 to 28, 30 to 32, and 35 are either too long or appear to be too cute. (However, I know a French-Canadian who would likely approve of number 31.) I rejected number 36 because its name may imply the UTF-16 encoding.

My Personal Recommendation

If you agree so far, this leaves basically 3 classes of names: number 3, number 16, and numbers 29, 33 and 34 (which are similar enough to each other to be considered the same). Here they are:

- BMC: “Basic Multilingual Collection”
- PAU: “Planes Accessible by UTF16”
- C17 or CP17: Collection of first 17 planes
- P17: “17 Planes from 0 to 16”
- PC17: “Plane Collection of first 17 planes”

Although “Planes Accessible by UTF-16” is descriptive, I personally find the description somewhat long and awkward. I personally do not think that the name implies UTF-16 encoding but some people might want to reject it for using “UTF-16” in the name. If we chose this name, we could still define a collection beyond plane 17 and name the new collection to avoid confusing the collection names. However, a new name does not immediately come to mind. Finally, “P-A-U” in English has a crisp sound that people could pronounce. We must not underestimate the need for an easily pronounced acronym.

The last 3 names are short and somewhat mnemonic with the “17”. If we chose one of these names, we could still define a collection beyond plane 17 and name the new collection to avoid confusing the collection names. For example, if we needed 30 planes, we might name the extended collection “C30” or “P30” or “CP30” or “PC30”. Also “C-seventeen”, “P-seventeen”, “C-P-seventeen” and “P-C-seventeen” have a crisp sound in English.

Of these, I prefer the first one, “BMC” for “Basic Multilingual Collection”. I really like the historical reference with the words of “Basic” and “Multilingual”. Moreover, should we ever need to expand the coding beyond 17 planes, we could name the new collection to be the “EMP” or “EMC” for “Extended Multilingual Planes” or “Extended Multilingual Collection” yet not easily confuse the “EMP” or “EMC” collection with the “BMC” and “BMP” collections. Finally, although “Basic Multilingual Collection” does not sound quite as crisp to me as “Basic Multilingual Plane” (probably because of the multiple syllables in “Collection”), “B-M-C” and “Basic Multilingual Collection” to me still have a nice, pronounceable sound.

Action Requested for L2 and UTC

Please consider the need for a name for the collection of characters in planes 0 to 16, decide on a preferred name (one with which you can live to achieve consensus), and submit a request that SC 2/WG 2 formally name this new collection.

[END OF DOCUMENT]