N2688: Sterile characters

Javier A. Múgica - javier@aerotri.es
March 8th, 2021

Purpose

To clarify the conversion from the source character set to the execution character set.

Current text

In section **5.1.1.2 Translation phases**, it can be read:

5. Each source character set member and escape sequence in character constants and string literals is converted to the corresponding member of the execution character set; if there is no corresponding member, it is converted to an implementation-defined member other than the null (wide) character.[8]

Thus, they are converted to *a* member of the execution character set, and that member is implementation defined. But then we can read at the foot:

[8]An implementation need not convert all non-corresponding source characters to the same execution character.

Were it not for the footnote it would seem that all those characters (which I will call *sterile characters*) are to be converted to the same member, or at least that is one possible interpretation. While footnotes are intended to clarify the text, they should not be necessary in order for the text to acquire its intended meaning. If the intent of the committee is what is expressed in the footnote it seems best to make that clear in the main text.

And after reading the text and the footnote it is not clear whether all instances of the same sterile character have to be mapped to the same member of the execution character set, or they need not. I don't know what is the intended behaviour, so I cannot propose a single wording, but give two alternatives accordingly. Also, I provide wordings with and without the term *sterile character.*

Proposed wording 1

5. Each source character set member and escape sequence in character constants and string literals is converted to the corresponding member of the execution character set. Each instance of a source character or escape sequence for which there is no corresponding member is converted in an implementation-defined manner to some member of the execution character set other than the null (wide) character.[8]

---

[8]An implementation may convert each instance of the same non-corresponding source character to a different member of the execution character set.

Proposed wording 2

5. Each source character set member and escape sequence in character constants and string literals is converted to the corresponding member of the execution character set. Each source character or escape sequence for which there is no corresponding member is converted in an implementation-defined manner to some member of the execution character set other than the null (wide) character. All instances of the same non-corresponding source character or escape sequence are converted to the same member.

Proposed wording 1b

5. Each source character set member and escape sequence in character constants and string literals is converted to the corresponding member of the execution character set. If there is no corresponding member the character or escape sequence is called an *sterile character.* Each instance of an sterile character is converted in an implementation-defined manner to some member of the execution character set other than the null (wide) character.[8]

Proposed wording 2b

5. Each source character set member and escape sequence in character constants and string literals is converted to the corresponding member of the execution character set. If there is no corresponding member the character or escape sequence is called an *sterile character.* Each sterile character is converted in an implementation-defined manner to some member of the execution character set other than the null (wide) character. All instances of the same sterile character are converted to the same member.