Universal Multiple-Octet Coded Character Set

International Organization for Standardization

Organisation Internationale de Normalisation

Международная организация по стандартизации

Doc Type:	Working Group Document
Title:	Proposal to add Block Elements and Symbols for Software Developing to the UCS
Source:	Renzhi Li
Authors:	Renzhi Li
Status:	Individual Contribution
Action:	For consideration by JTC1/SC2/WG2 and UTC
Date:	February 16, 2019

I. Introduction

This document proposes the addition of 8 new graphic characters to provide capability with various symbols widely used in modern programming society into the UCS.

II. Background

Software developing heavily used text editors, especially text editors in terminal environments that all user interface elements are formed with text characters. Box-drawing characters and block elements are heavily used in such text editors to build user interface for programmers.

Box-drawing characters, solid and shaded blocks, and similar graphic characters were encoded in the UCS since 1991 (Unicode 1.0) for compatibility with character sets in various computer systems, especially IBM PC. The set of block characters was augmented in 1999 (Unicode 3.0) and in 2002 (Unicode 3.2) to cover additional platforms, due largely to proposals by Frank da Cruz [1] [2] [3] [4]. In 2019, Ewell et al. introduced Graphics for Legacy Computing block to support the similar characters included in various vintage computers and terminal emulators, including Apple, Atari and Commodore systems [5].

However, these change does not cover recent motivations of block element extension. A popular extension to existing block element set is called *Powerline*, initially introduced for *vim* text editor, which is a popular text editor that runs in terminals under various operating systems. *Powerline* introduced 7 PUA characters: 3 of them are indicators for source control branch, file writability and cursor position, and the rest 4 are triangular or angular block elements forms the user interface.

Figure 1. Image of *Powerline* in *vim* [6]

Powerline PUA characters get rapidly widespread across modern mono-space typefaces. Many recent mono-space typefaces, noticeably *Pragmata Pro* [7], *Source Code Pro* [8] and *Iosevka* natively supported *Powerline* characters. On the other hand, these characters are also get used by other software, including *Vim-Airline*, *Agnoster* [9] shell theme for *zsh*, *SpaceVim* [10], and more.

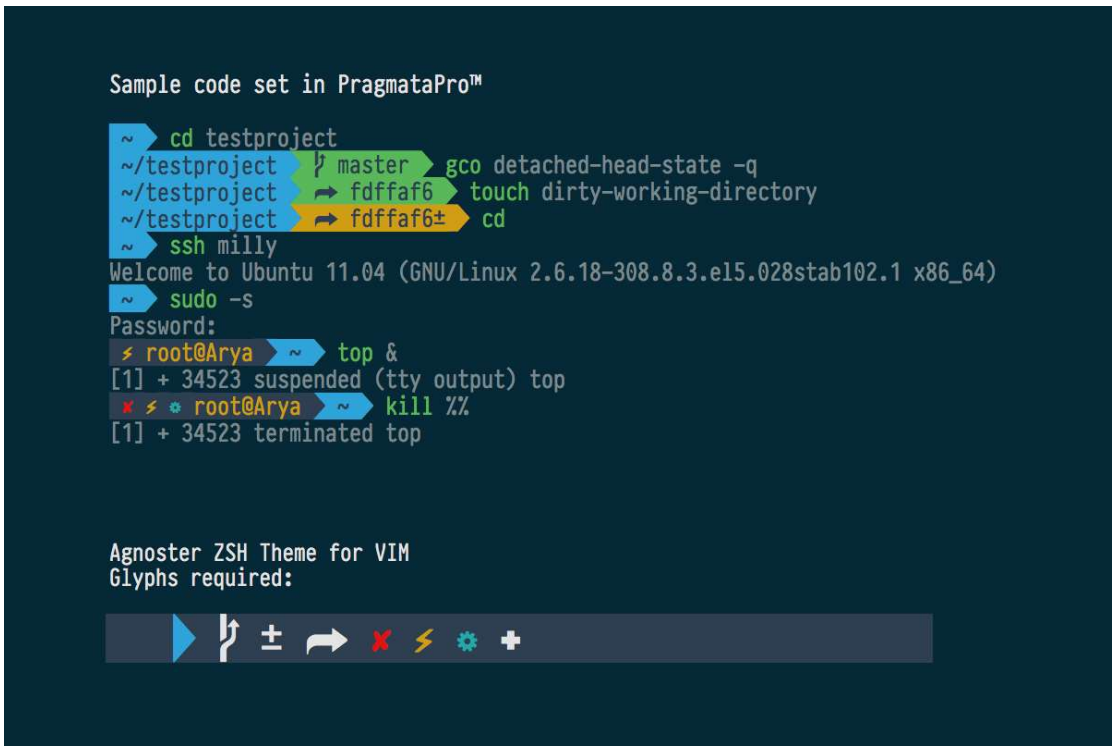

Figure 2. *zsh* theme *Agnoster* under Pragmata Pro font, heavily utilizing Powerline symbols

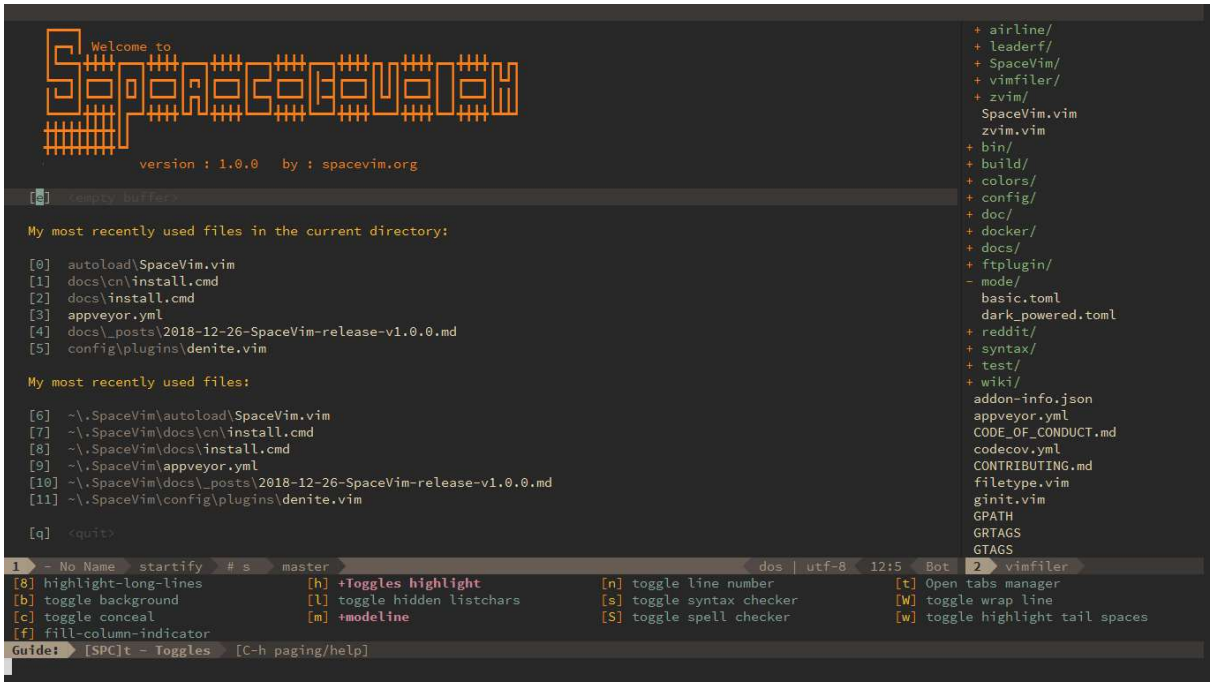

Figure 3. Image of *SpaceVim*, a popular *vim* configuration using *Powerline* characters

III. Character Set

Table 1 defines the purposed Unicode block "*Block Elements and Symbols for Software Developing*", encompassing 16 code points, while the first eight are assigned.

The preferred code point assignment for these symbols are U+2FE0 to U+2FEF. Arranging these characters into the almost-full BMP is because that most terminal emulators, consoles and programming-oriented text editors *do not have proper text shaping support*, therefore arranging them into BMP will minimize potential compatibility issues. The assignment U+1FXX0—U+1FXXF is the second preferred, arranging these symbols into SMP.

Block Elements and Symbols for Software Developing

	**U+2FE0** U+1FXX0	VERSION CONTROL BRANCH SYMBOL
	**U+2FE1** U+1FXX1	LINE NUMBER INDICATOR
	**U+2FE2** U+1FXX2	READ-ONLY SYMBOL
	**U+2FE3** U+1FXX3	COLUMN NUMBER INDICATOR
	**U+2FE4** U+1FXX4	RIGHT-POINTING TRIANGULAR BLOCK
	**U+2FE5** U+1FXX5	RIGHT-POINTING ANGULAR LINES SEPARATOR
	**U+2FE6** U+1FXX6	LEFT-POINTING TRIANGULAR BLOCK
	**U+2FE7** U+1FXX7	LEFT-POINTING ANGULAR LINES SEPARATOR
	U+2FE8 U+1FXX8	(RESERVED)
	U+2FE9 U+1FXX9	(RESERVED)
	U+2FEA U+1FXXA	(RESERVED)
	U+2FEB U+1FXXB	(RESERVED)
	U+2FEC U+1FXXC	(RESERVED)
	U+2FED U+1FXXD	(RESERVED)
	U+2FEE U+1FXXE	(RESERVED)
	U+2FEF U+1FXXF	(RESERVED)

Table 1. Block encoding chart

The UCD properties for all the characters listed above will follow this pattern, similar to other box-drawing characters and block elements: `<Code Point>;<Name>;So;0;ON;;;;;N;;;;;`.

IV. References

[1] F. d. Cruz, "L2/98-353: Additional Control Pictures for Unicode," 1998.

[2] F. d. Cruz, "L2/98-354: Terminal Graphics for Unicode," 1998.

[3] F. d. Cruz, "L2/98-355: Hex Byte Pictures for Unicode," 1998.

[4] F. d. Cruz, "L2/00-159: Supplemental Terminal Graphics for Unicode," 2000.

[5] D. Ewell, R. Bettencourt, R. Bánffy, M. Everson, E. M. Silva, E. Mårtenson, M. Shoulson, S. Steele and R. Turner, "L2/19-025: Proposal to add characters from legacy computers and teletext to the UCS," 2019.

[6] Powerline Team, "Powerline," [Online]. Available: https://github.com/powerline/powerline. [Accessed 16 2 2019].

[7] F. Schiavi, "Pragmata Pro — Agnoster," [Online]. Available: https://www.fsd.it/shop/fonts/pragmatapro/#tab-1449660161180-6-0. [Accessed 16 2 2019].

[8] Adobe Inc, "Source Code Pro issue #43 — Powerline symbols," [Online]. Available: https://github.com/adobe-fonts/source-code-pro/issues/43. [Accessed 16 2 2019].

[9] I. Wolkerstorfer, "Agnoster," [Online]. Available: https://github.com/agnoster/agnoster-zsh-theme. [Accessed 16 2 2019].

[10] SpaceVim, "SpaceVim," [Online]. Available: https://github.com/SpaceVim/SpaceVim. [Accessed 16 2 2019].