Proposal to include Group Mark symbol

Ken Shirriff

Feb 2015

Abstract

The group mark symbol was part of the character set of important computers of the 1960s and 1970s, such as the IBM 705, 7070, 1401 and 1620. Books about these computers and manuals for them used the symbol ǂ to represent the group mark in running text. Unfortunately, this symbol does not exist in Unicode, making it difficult to write about technical details of these historical computers.

Introduction

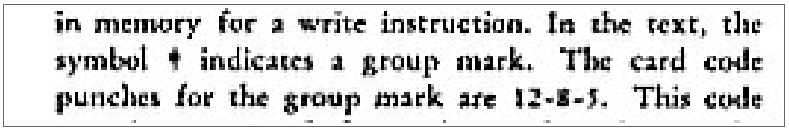
The group mark was introduced in the 1950s as a separator character for I/O operations. In written text, the group mark is indicated by the symbol: ǂ. Unicode doesn't include this symbol, which is inconvenient when writing about the group mark or the character set of these computers.

The group mark became part of IBM's Standard BCD Interchange Code (BCDIC) in 1962. [1, page 20 and figure 56]. This standard was used by the IBM 1401, 1440, 1410, 1460, 7040, and 7044 data processing systems [2]. The BCDIC standard provided consistent definitions of codes and the relation between these codes and printed symbols, including uniform graphics for publications.[2, page A-2]. Unicode can represent all the characters in BCDIC, except for the group mark.

This proposal is for a Unicode encoding of the group mark as a human-readable character that is used in running text to represent the group mark. This is somewhat different from the use of the group mark as a separator in data files. While the proposed group mark could be used in data files, that is not the focus of this proposal. In other words, this proposal is for a text symbol, not a control character.

Examples of the group mark in text

The group mark symbol ǂ has been used in running text from the 1950s up to the present, both in documentation and in books. This section gives examples of the use of the group mark symbol from a variety of sources.


in memory for a write instruction. In the text, the
symbol ǂ indicates a group mark. The card code
punches for the group mark are 12-8-1. This code

1 Eric Fischer, *The Evolution of Character Codes, 1874-1968*.

2 Jack Melnick, *IBM 1401, 1440 and 1460 Programming Techniques*, 1964.

Figure 1. Example of group mark symbol in use in text from 1956, in a manual for the IBM 705 data processing system. While the resolution of the symbol is bad, this example illustrates an early use of the group mark. [3, p20]

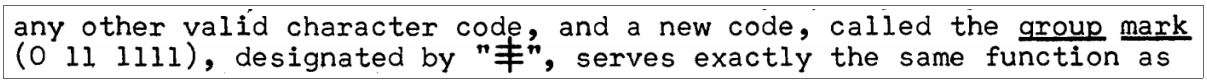

any other valid character code, and a new code, called the group mark
(0 11 1111), designated by "≢", serves exactly the same function as

Figure 2. Example of group mark symbol in use in text from a Univac manual from 1957. [4, p1103.6]

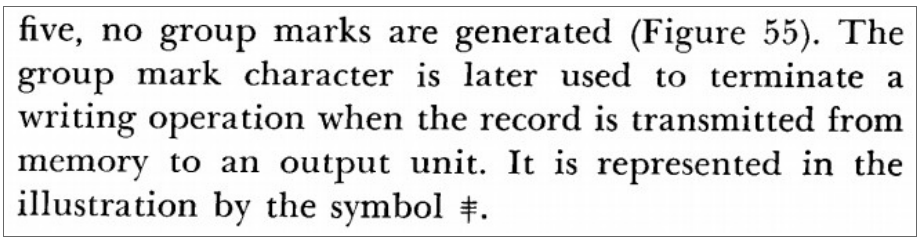

five, no group marks are generated (Figure 55). The group mark character is later used to terminate a writing operation when the record is transmitted from memory to an output unit. It is represented in the illustration by the symbol ≠.

Figure 3. Group mark explained and symbol used in running text in General Information Manual for the IBM 705. [5, p34]

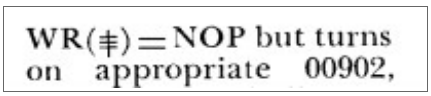

WR(≢) = NOP but turns
on appropriate 00902,

Figure 4. Group mark symbol in use in text from 1959, in a reference manual for the 705 data processing system [6, p157]

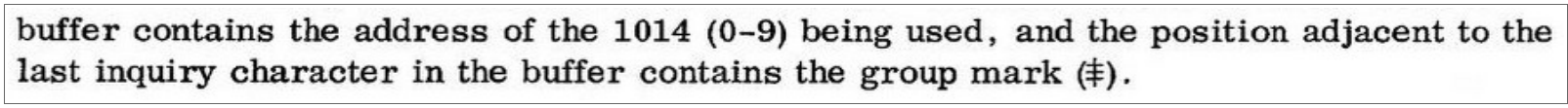
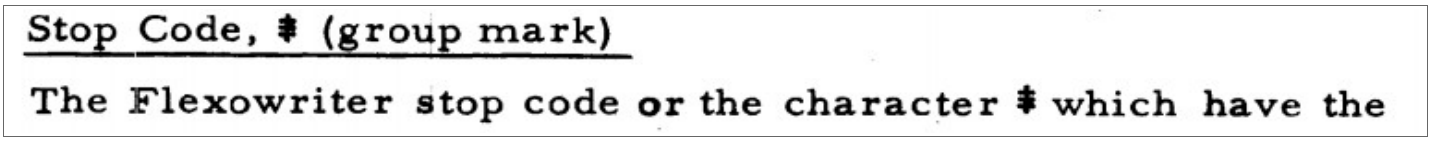

buffer contains the address of the 1014 (0-9) being used, and the position adjacent to the last inquiry character in the buffer contains the group mark (≢).

Figure 5. Group mark in the manual for an IBM I/O system. [7, p4]


Stop Code, ≢ (group mark)

The Flexowriter stop code or the character ≢ which have the

Figure 6. The group mark used in a programming document from Scientific Data Systems, Inc. [8, p3]

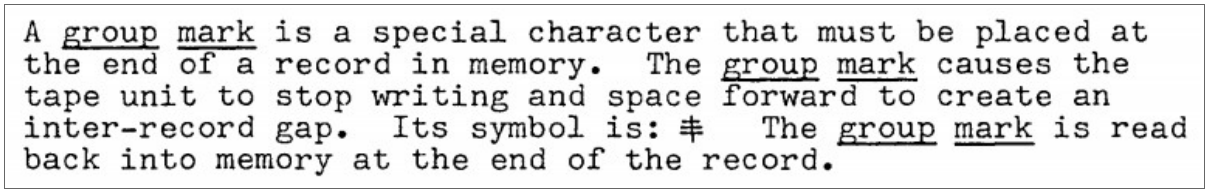

A group mark is a special character that must be placed at the end of a record in memory. The group mark causes the tape unit to stop writing and space forward to create an inter-record gap. Its symbol is: ≢ The group mark is read back into memory at the end of the record.

Figure 7. The group mark used in a book from 1962 about the IBM 1401. [9, p146]

---

3 *IBM electronic data-processing machines, type 705: Preliminary manual of operation*, 1956.
4 *Univac II Marketing Manual*, 1957.
5 IBM, *General Information Manual, 705 Data Processing System*, 1959.
6 IBM, *Reference Manual, 705 Data Processing System*, 1959.
7 *IBM 7070-7074 Data Processing System Bulletin*, 1962.
8 Scientific Data Systems, *SDS 900 Series Program Library*, Catalog No. 000011, 1962.
9 James A. Saxon and William S. Plette, *Programming the 1401*, Prentice Hall, 1962.

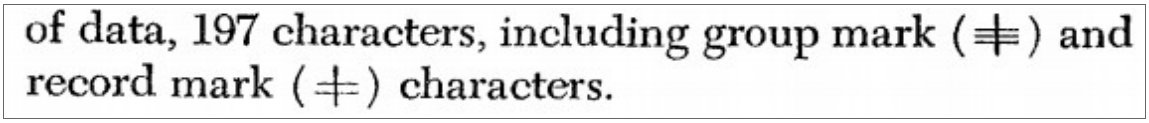


Figure 8. Group mark from a 1964 printer manual. Note the similar record mark character, with two horizontal lines instead of three. [10, p1]

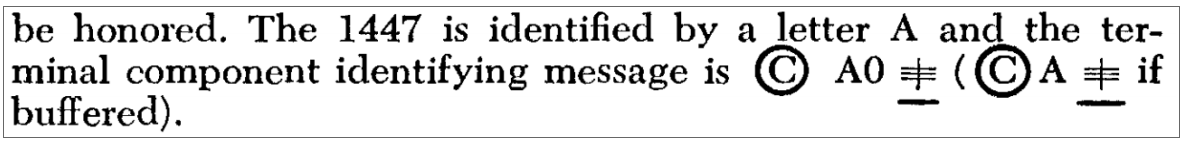


Figure 9. Group mark symbol in text in an IBM reference document from 1964 [11, p26]

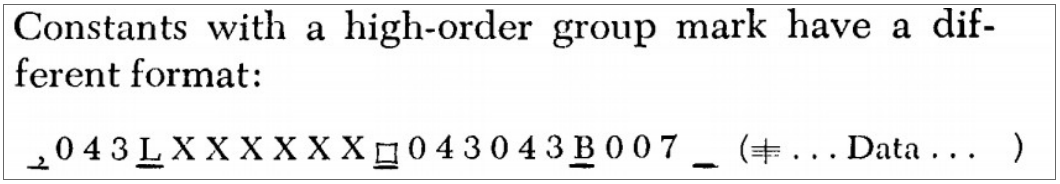


Figure 10. Group mark used in text in another IBM manual. [12, p51]

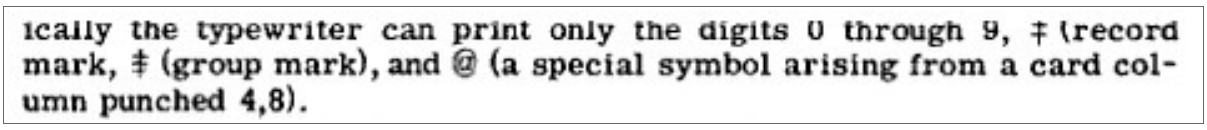


Figure 11. Group mark in a programming book for the IBM 1620 from 1965 [13, p130]

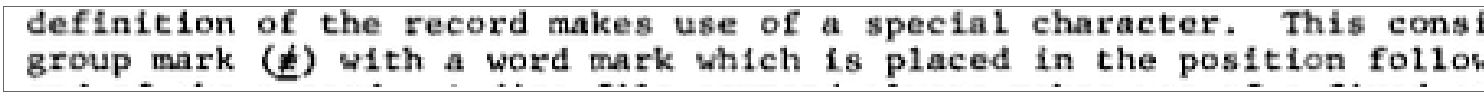


Figure 12. The group mark was also in use on the IBM 7070. This hand-drawn group mark has an slightly slanted vertical bar. The underline indicates the character has been combined with a word mark. [14, p84]

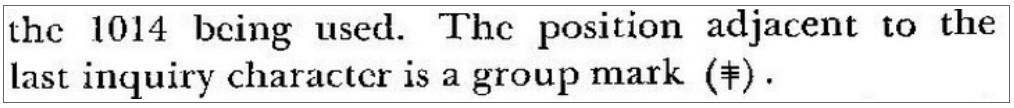
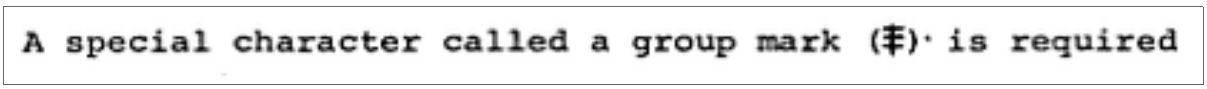


Figure 13. The group mark in an IBM 7094 manual from 1966. [15, p96]



10 *IBM 1443 Printer for 1620/1710 Systems,* 1964.

11 IBM Systems Reference Library, *IBM 1447 Console*, 1964.

12 *Autocoder (on Tape) Language Specifications and Operating Procedures IBM 1401 and 1460*, 1964.

13 Eric A Weiss, *Programming the IBM 1620: The Hands-on Approach,* McGraw-Hill, 1965.

14 Gordon Bitter Davis, *An introduction to the IBM system/1401 computer*, McGraw-Hill, 1965.

15 *IBM 7094 Principles of Operation*, 1966.

Figure 14. Group mark in a programming book from 1968 [16 ,p253]

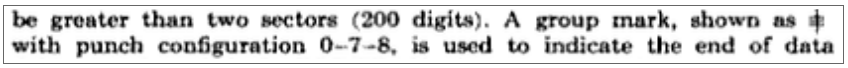

be greater than two sectors (200 digits). A group mark, shown as ⋕ with punch configuration 0–7–8, is used to indicate the end of data

Figure 15. Group mark in a FORTRAN book from 1969. [17, 79]

The figure below shows that books were using the group mark symbol into the 1980s.

Fig. 23-21. The IBM BCD interchange code.

Figure 16. Group mark in an 1982 engineering handbook showing its use in BCD Interchange Code. [18, p23-17]

The group mark symbol is in current use on the web, as seen in several Wikipedia pages and other pages. Because the symbol is missing from Unicode, the pages need to use a bitmap image for the group mark.

BCD文字	Print-A	Print-H	カード	BCD	操作	定義その他
空白				0		
.	.	.	12-3-8	BA8 21	Halt	
¤	¤	)	12-4-8	CBA84	Clear Word Mark	*Lozenge*
[			12-5-8	BA84 1		
<			12-6-8	BA842		*Less Than*
⋕			12-7-8	CBA8421		*Group Mark*

Figure 17. The Japanese Wikipedia page on IBM 1401 uses a bitmap image for the group mark. [19]

GM	⋕	(Stop)	(Stop)	0 8421	E (Stop)	0-7-8	1-2-4-8		8421		*Group Mark*

Figure 18. The Wikipedia page on IBM 1620 also uses a bitmap image for the group mark. [20]

Additional evidence that the group mark is considered a typographical character for use in text is found in a catalog for the Selectric typewriter. The group mark is listed as an available character and is available on a "88 character universal" type ball [21, p11,12]. In addition, page 114 shows the group mark character available for printer type bars.

---

16 William J. Claffey, *Principles of programming the IBM 1620 computer,* Dickenson Pub. Co., 1968.
17 D. K. Carver, *Introduction to Fortran II and Fortran IV programming,* Wiley, 1969.
18 Electronics engineers' handbook. Donald G. Fink, Donald Christiansen, McGraw-Hill, 1982.
19 Wikipedia (Japanese), *IBM 1401*, 2015.
20 Wikipedia (English), *IBM 1620* , 2015.
21 IBM ISG, *Type Catalog*, 1983.

Properties

The proposed properties for the group mark are:

```
23FA;GROUP MARK;So;0;ON;;;;;N;;;;;
```

Property	Suggested Value
Code point	23FB
Name	GROUP MARK
General Category	So
Canonical Combining Class	0
Bidirectional Class	ON
Decomposition Type / Decomposition Mapping	
Numeric Type	
Numeric Value	
Bidi Mirrored	N
Unicode 1 Name	
ISO Comment	
Simple Uppercase Mapping	
Simple Lowercase Mapping	
Simple Titlecase Mapping	

Table 1: Suggested character properties for group mark.

Range: The character seems suited for Miscellaneous Technical, with proposed code point 23FB. Another possibility would be Miscellaneous Symbols and Arrows, with code point 2BD2.

Collation: The group mark symbol is part of a well-defined collating sequence [22, page N-1] for Standard BCD Interchange Code. The group mark is near the beginning of the collating sequence, after "(" and "<", before "&" and "$". It would make most sense for the group mark symbol to have default Unicode collation in that area, but since the Standard BCD Interchange Code has a very different collation order, the collation doesn't matter too much.

The group mark is uncased and has no special line-breaking behavior. The character is not meant for use in identifiers. It is a standalone symbol. It is not a white-space character and has no numeric value. It is not a combining character or punctuation.

Potential issues

Can an existing Unicode symbol represent the group mark? The CJK character U+4E30 丰 looks kind of like the group mark. It's not really satisfactory since the lines are different length and have hooks. The character NOT IDENTICAL TO U+2262 ≢also somewhat resembles the group mark, except the vertical line is

22 *System Operation Reference Manual. IBM 1401 Data Processing System. IBM 1460 Data Processing System*, 1966.

slanted, which makes it unsatisfactory. One possibility from an earlier email thread [23] is an overstrike: IDENTICAL TO, COMBINING LONG VERTICAL LINE OVERLAY . This isn't a good representation, since the group separator is not a combination of two underlying characters. (See earlier email discussion [24]). As an aside, the group separator with word mark, shown in Figure 12 is clearly a combination of the group separator with an underline, and would not make sense as a separate character. (In the 1401, any character can have a word mark on its storage location and this is represented in text by underlining the character.)

Unicode has graphic Pictures for Control Codes including: U+241D “SYMBOL FOR GROUP SEPARATOR”. [25] with glyph: ␝. This raises the question of whether the group mark could be represented as the SYMBOL FOR GROUP SEPARATOR. However, a new group mark symbol is desired for both semantic and visual reasons.

The Pictures for Control Codes are intended for displaying control codes within a data stream, while the proposed group mark is intended as a symbol for running text, which is semantically very different. The group mark was used in the IBM 1401 and other computers as a delimiter for I/O operations. For instance, a tape inter-record gap was represented in the computer by a group mark, and a group mark indicated the end of an I/O buffer. However, this proposal does not deal with the group mark as a character in storage or a data stream, but as a symbol that appears in text. For storing legacy data in a file using Unicode, the SYMBOL FOR GROUP SEPARATOR would work fine. But for written text, the proposed group mark symbol fills a role that is different from SYMBOL FOR GROUP SEPARATOR.

Visually, the SYMBOL FOR GROUP SEPARATOR glyph looks totally different from the group mark, so it doesn't make sense to consider them glyph variants. The SYMBOL FOR GROUP SEPARATOR glyph ␝ can't be substituted in text for the group mark without totally changing the meaning to the reader.

One question is if the group mark symbol is sufficient, or additional symbols will be necessary to support discussions of the 1401 and related computers. The Standard BCD Interchange Code has several special characters in addition to the group mark:

- Mode change: Δ.
- Word separator: ˠ
- Tape segment mark: ⧺
- Blank: ␣
- Tape mark: √
- Record mark ‡

These characters can all be represented reasonably well with existing Unicode symbols. These are displayed above as GREEK CAPITAL LETTER DELTA, MODIFIER LETTER SMALL GAMMA, TRIPLE PLUS, BLANK SYMBOL, SQUARE ROOT, and DOUBLE DAGGER respectively. Adding the group mark is sufficient to complete the character set. While new characters could be added to Unicode for MODE CHANGE, etc. the new characters would look nearly identical to the existing characters. There would be

23 026 keypunch, Unicode mail archive, 2007.
24 Productive Glyph Design vs. Productive Character Representation (was: Re: Quick survey of Apple symbol fonts ...), Unicode mail archive, 2011.
25 Unicode Standard 7.0, *Unicode Control Pictures*, Range: 2400-243F, 1991-2014.

some semantic benefit of separating MODE CHANGE from GREEK CAPITAL LETTER DELTA, for instance. But I don't see it as being useful enough to merit introducing multiple new characters.

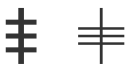
One question that could be raised is if the group mark symbol is of relevance only to IBM, as the majority of the usage references are IBM documents. However, the symbol has been used in books written by external authors [9, 13, 14, 16]. It has also been used by Univac [4] and Scientific Data Systems [8]. Finally, the symbol is currently in use on Wikipedia pages [19, 20]. Thus, the symbol is in general use, not just IBM.

One user community that could use the group code symbol is the IBM 1401 user community [26]. The author has contacted this community and received moderate interest.

In addition to being useful for written text, the group mark symbol could be used by software. IBM 1401 simulators [27] currently use ASCII, representing the group mark with a somewhat arbitrary character such as " or }. Providing the group mark in Unicode would permit simulators to go beyond ASCII and display characters as intended.

Drawing the glyph

The sample texts above show two different glyphs have been used for the group mark, summarized below:



The first has short horizontals that are widely spaced, with a line weight that matches the surrounding text; Figure 3 is an example. The second has the three horizontal lines thin, close together, and long (about the same length as the vertical); Figure 7 is an example. These are clearly variants of the same symbol - I've seen several documents that use both glyphs for the group mark in the same document. I recommend the first glyph - it harmonizes better with the DOUBLE DAGGER if that is used for the record mark. The first glyph also renders better than the second at small size.

Included with this proposal is a TrueType font called GroupMark that contains the proposed group mark character, the first one at U+E000 and the second at U+E002.

Conclusion

Looking at the Unicode symbol guidelines, the group mark ⧺ is a good fit for addition to Unicode. It is typically used as part of computer applications. It has a well defined user community. It occurs in running text. Being able to search for it in text would be useful. It has well-defined semantics that make it appropriate for computer processing. It completes a class of symbols already in the standard (the symbols in Standard BCD Interchange Code). Finally, it is letterlike in the sense that it should match the surrounding font style.

The group mark symbol has been used in running text since the 1950s. Providing the group mark symbol in Unicode would be beneficial.

Acknowledgments: the author benefited from the advice in Proposal to add IEC power symbols to Unicode.

26 The 1401_software mailing list is one place to find the relevant user community.
27 The ROPE system is probably the easiest to use if you're looking for a 1401 simulator. It is built on top of the SIMH simulator.