



中华人民共和国国家标准

GB/T 18142—2017/ISO/IEC 14957:2010
代替 GB/T 18142—2000

信息技术 数据元素值表示 格式记法

Information technology—Representation of data elements values—
Notation of the format

(ISO/IEC 14957:2010, IDT)

2017-11-01 发布

2018-05-01 实施

中华人民共和国国家质量监督检验检疫总局 发布
中国国家标准化管理委员会

目 次

前言	Ⅲ
引言	Ⅳ
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 与数据元素表示的字符类型和长度有关的记法	1
附录 A (资料性附录) 用于数据元素描述的 EBNF 语法	4
参考文献	5

前 言

本标准按照 GB/T 1.1—2009 给出的规则起草。

本标准代替了 GB/T 18142—2000《信息技术 数据元素值格式记法》，与 GB/T 18142—2000 相比，除编辑性修改外主要技术变化如下：

- 改变了原来传统简单的格式记法，采用了 C 语言的方式来表达；
- 增加了编程语言(GB/T 15272)为使用该记法的标准(见引言)；
- 修改了规范性引用文件(见第 2 章)；
- 修改了原来传统简单的格式记法，采用了 C 语言的方式来表达(见第 4 章)；
- 删除了“数字的附加记法”(见 GB/T 18142—2000 第 5 章)。

本标准使用翻译法等同采用 ISO/IEC 14957:2010《信息技术 数据元素值表示 格式记法》。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本标准由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本标准起草单位：中国电子技术标准化研究院、国富通信息技术发展有限公司。

本标准主要起草人：王静、张展新、高爽、刘薇娜、赵菁华、樊星。

本标准所代替标准的历次版本分布情况为：

- GB/T 18142—2000。

引 言

在商业、技术和公共领域中,数据交换正迅速扩展。它使得不同团体间相互工作不断增多;而这些团体为满足特定需求往往已经独立开发了信息处理应用和电信网络。因此,全局一致性差。

为了补救这种情况,有必要特别针对数据元素表示开展急需的标准化工作。

对于数据元素的表示,首先考虑对格式(即数据元素表示中使用的字符的类型和该表示的长度)加以规范。为了使这些规范对所涉及的每个人都有同样的意义,有必要按照标准化约定来表达这些规范。

这些规则可能消除含糊、不确切和差错;也便于各种数据元素字典的比较、信息系统的设计和创建、及电子数据交换(EDI)。

在其他一些标准中按照特定背景对这些记法中的一部分做了不同表达,例如:EDIFACT (ISO 9735)、金融标准(如,ISO 7982-1)、字符集(GB/T 15273)、信息处理(GB/T 15273)、编程语言(GB/T 15272)。

本标准的目的是为所有那些利用与环境无关的记法类型的标准提供关于此主题的唯一引用源。

信息技术 数据元素值表示 格式记法

1 范围

本标准规定了用于叙述格式(即数据元素表示中使用的字符类型和这些表示的长度)的记法,也规定了有关数字表示的附加记法。例如,这种格式化技术可能用于数据元素的元数据部分。

本标准的适用范围仅限于图形字符,如数字、字母和特殊字符。且仅限于字符、字符串、整数、实数和指针这些基本数据类型。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 13000—2010 信息技术 通用多八位编码字符集(UCS)(ISO/IEC 10646:2003, IDT)

3 术语和定义

下列术语和定义适用于本文件。

3.1

数据元素 data element

在上下文中,被认为是不可分割的数据单元。

[GB/T 5271.4—2000, 定义 04.07.01]

3.2

字符集 character set

字符的一个有限集合,且它对于给定目的是完整的。

[GB/T 5271.4—2000, 定义 04.01.02]

3.3

字符类型 character type

同类的或具有相同用途的字符的集合。

示例:字母、数字、特殊字符等。

3.4

长度 length

表示的长度 length of representation

用于表示一个数据元素的字符数。

4 与数据元素表示的字符类型和长度有关的记法

格式应为字符串序列。该格式由零个或多个指示符组成,包括一个或多个间隔符(GB/T 13000—2010 中第 20 条定义),一个普通字符(既不是%,也不是间隔符),或一个转换规范。每个转换说明由字符%开始。

注:此格式规范以 C 编程语言(ISO/IEC 9899:1999)中的函数“fscanf()”为基础。

从概念上说,格式字符串相当于一种使用句法单位(用格式指示符描述)的句法处理器。本标准不指定任何服务接口,因此,不针对那些没有匹配格式化指示符的字符串规定差错处理。

在%之后逐一出现下列元素:

- 可选的禁止赋值字符 * ;
- 可选的非零十进制整数,用以指定最大字段宽度(以字符为单位);
- 可选的长度修饰符,用以指定接收对象的大小;
- 转换说明符字符,用以说明所要进行的转换的类型。

依次处理格式的每个指示符。

由一个或多个间隔符组成的指示符,在读取输入时予以执行,直到出现第一个非间隔符(保持未读取状态),或直到没有可读取的字符为止。

由普通字符构成的指示符,在读取字符流中的下一个字符时予以执行。如果这些字符中有任何一个与组成该指示符的字符不同,则该指示符无效,并且这些不同的字符和其后的字符保持未读取状态。

类似,如果在文件的结尾出现编码错误或造成阻止字符读取的读取错误,则该指示符无效。

指示符,作为转换规范,它定义一组匹配输入序列。下面描述指示符的每个说明符。转换规范按以下步骤执行:

- 跳过输入的间隔符,除非该规范中包括说明符[,c 或 n。¹⁾
- 读取字符流中的输入项,除非该规范中包括说明符 n。输入项定义为最长输入字符序列,它不超过指定的字段宽度并且是一个匹配的输入序列或匹配的输入序列的前缀。
- 输入项之后第一个字符(如果出现)保持未读取状态。如果输入项的长度为零,则该指示符的执行失败。此状况是匹配失败(除非是文件的结尾)、编码错误或造成阻止字符流输入的读取错误(在此情况下,是输入失败)。
- 把输入项(如果指示符是%n,则是把输入字符的计数)转换为与该转换说明符对应的类型;但对于说明符%的情况例外。如果输入项不是匹配序列,该指示符的执行失败;此状况是匹配失败。除非用*指出禁止赋值,转换结果放置在一个变元指出的对象中;这个变元是紧跟在还没有收到转换结果的格式变元之后的第一个变元。如果此对象没有一个与之对应的类型,或者转换的结果不能在此对象中表示,则该行为是未定义行为。

转换说明符及其含义如下:

- d 匹配一个可选带符号十进制整数,其格式预期与基数变元值为 10 的 strtol()函数的主体序列相同。例如:“%d”匹配“17”,而不是“1.0”“17H”“ABC”;“%03d”匹配“017”“000”“−017”,而不是“1 000”。

注:“%d0”不匹配任何字符串。

- i 匹配一个可选带符号的整数,其格式预期与基数变元值为 0 的 strtol()函数的主体序列相同。例如:“%i”匹配“+17”“17”“−17”,而不是“1.0”“+17H”。
- o 匹配一个可选带符号八进制整数,其格式预期与基数变元值为 8 的 strtol()函数的主体对象序列相同。例如:“%o”匹配“0123”“−456”,而不是“0789”。
- u 匹配一个可选不带符号的,其格式预期与基数变元值为 10 的 strtol()函数的主体序列相同。例如:“%u”匹配“17”,而不是“+17”“−17”。
- x 匹配一个可选带符号十六进制整数,其格式预期与基数变元值为 16 的 strtol()函数的主体序列相同。例如:“%x”匹配“013EF”“−013ef”,而不是“013EFG”。
- a,e,f,g 匹配²⁾一个可选带符号浮点数、无穷大或非数值,其格式预期与 strtod()函数的主体序

1) 这些间隔符不计入规定的字段宽度不可冲突。

2) a,e,f 和 g 格式之间没有差别。

- 列相同。对应的变元应是一个浮点指针。例如：“%f”匹配“1.2E10”“+1.2E+10”“-1.2e-10”，“无穷大”“非数值”，而不是“1.2F10”。
- c 匹配一个由字段宽度具体规定字符数的字符序列（如果该指示符中没有字段宽度，此数为 1）。如果有长度修饰符，那么该输入应是一个字符序列。例如：“%c”匹配“X”，而不是“XY”；“%2c”匹配“XY”。
 - s 匹配一个无间隔符的字符序列。如果有长度修饰符，那么该输入应是一个字符序列。例如：“%s”匹配“123abc”“123abcd”，而不是“123 abc”。
 - [] 匹配取自预期字符集合（扫描集）的非空字符序列。转换说明符包含该格式字符串中的所有后续字符，直到且包含匹配的右括号（]）。括号之间的字符（扫描列表）组成扫描集。但是，如果左括号后的字符是一个长音符号（ $\bar{\quad}$ ），那么，扫描集包含所有那些没有出现在长音符号与右括号之间的扫描列表中的字符。如果转换说明符以 [或 [$\bar{\quad}$ 开始，那么，右括号字符在扫描列表里并且右括号字符后面的下一个字符是结束该规范的匹配的右括号；否则，右括号后的第一个字符是结束该规范的字符。如果字符 - 在扫描列表中，并且它既不是第一个字符也不是最后一个字符，在以字符 $\bar{\quad}$ 为首字符的情况下也不是第二个字符，那么，此行为是定义实现的。例如：“abc[ad- $\bar{\quad}$]”匹配“abca”“abcd”“abce”“abcf”，而不是“abc”“abeg”；“abc[$\bar{\quad}$ d- $\bar{\quad}$]”匹配“abcg”，而不是“abc”“abcd”。
 - % 匹配单一字符 %；不发生转换。完整的转换规范应为 %%。

附录 A
(资料性附录)

用于数据元素描述的 EBNF 语法

以下句法基于 ISO/IEC 14977 信息技术—句法元语言—扩展 BNF(巴科斯范式)：

```

format_description = { [non_format_component] , format_component } ,
                    [non_format_component] ;
non_format_component = { non_format_characters } ;
non_format_characters = (ascii "%" ) · "%%" ;
format_component = signed_decimal_format |
                    signed_integer_format |
                    signed_octal_format |
                    unsigned_integer_format |
                    signed_hexadecimal_format |
                    character_array_format |
                    character_string_format |
                    defined_character_set_format |
                    real_number_format ;

decimal_with_optional_leading_zero = decimal_numeral · { decimal_numeral } ;
decimal_numeral = "0" | "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9" ;
signed_decimal_format = "%" , [ "-" | "." ] ,
                        [ "*" decimal_with_optional_leading_zero ] , "d" ;
signed_integer_format = "%" , [ "+" | "-" ] ,
                        [ "*" decimal_with_optional_leading_zero ] , "i" ;
signed_octal_format = "%" , [ "+" | "-" ] ,
                     [ "*" decimal_with_optional_leading_zero ] , "o" ;
unsigned_integer_format = "%" , [ "+" | "-" ] ,
                          [ "*" decimal_with_optional_leading_zero ] , "u" ;
signed_hexadecimal_format = "%" , [ "+" | "-" ] ,
                           [ "*" decimal_with_optional_leading_zero ] , "x" ;
character_array_format = "%c" ,
                        [ "*" decimal_with_optional_leading_zero ] , "c" ;
character_string_format = "%s" ,
                        [ "*" decimal_with_optional_leading_zero ] , "s" ;
defined_character_set_format = "%" ,
                              [ "*" | decimal_with_optional_leading_zero ] ,
                              "[ , { ascii - " } , "]" ;
real_number_format = "%" , [ "+" | "-" ] ,
                    [ "*" | decimal_with_optional_leading_zero ] ,
                    ( "a" | "e" | "f" | "g" ) ;
    
```

参 考 文 献

- [1] GB/T 5271.4—2000 信息技术 词汇 第4部分:数据的组织
 - [2] GB/T 14805.4—2007 行政、商业和运输业电子数据交换(EDIFACT)应用级语法规则(语法版本号:4,语法发布号:1) 第4部分:批式电子数据交换语法和服务报告报文(报文类型为 CONTROL)
 - [3] GB/T 15272—1994 程序设计C语言
 - [4] GB/T 15948.1—2007 银行电讯 资金转账报文 第1部分:电子资金转账报文数据段和数据元术语及通用集
 - [5] GB/T 18221 2000 信息技术 程序设计语言、环境与系统软件接口 独立于语言的数据类型
 - [6] ISO 6093 Information processing Representation of numerical values in character strings for information interchange
 - [7] ISO/IEC 8859 (all parts) Information technology—8-bit single-byte coded graphic character sets
 - [8] ISO/IEC 14977 Information technology— Syntactic metalanguage— Extended BNF
-

中华人民共和国
国家标准
信息技术 数据元素值表示 格式记法
GB/T 18142—2017/ISO/IEC 14957:2010

中国标准出版社出版发行
北京市朝阳区和平里西街甲2号(100029)
北京市西城区三里河北街16号(100045)

网址: www.spc.org.cn

服务热线: 400-168-0010

2017年11月第一版

书号: 155066·1-57759

版权专有 侵权必究



GB/T 18142-2017