



中华人民共和国国家标准

GB/T 15936.10—1996
idt ISO 8613-10:1991

信息处理 文本和办公系统 办公文件体系结构(ODA)和交换格式 第10部分：形式规范

Information processing—Text and office systems—
Office Document Architecture(ODA) and interchange
format—Part 10: Formal specifications

1996-12-26 发布

1997-07-01 实施

国家技术监督局 发布

目 次

前言	I
ISO 前言	II
1 范围	1
2 引用标准	2
3 定义	2
4 规范语言的语法和语义	2
4.1 基本概念	2
4.2 规范语言的语法	3
4.3 含固有语义的谓词符号	6
4.4 含固有语义的算符符号	6
4.5 其他项	7
4.6 记号简化	8
5 形式规范的结构	9
6 常用定义	10
7 文件结构的形式规范	16
7.1 组元集	21
7.2 组元	31
7.3 属性	55
7.4 辅助定义	80
7.4.1 谓词	80
7.4.2 函数	87
7.5 补充术语定义	89
7.6 谓词符号、算符和属性名的索引	91
附录 A(提示的附录) 有关规范语言的介绍	101
A1 引言	101
A2 原子构造	101
A3 复合构造	102
A4 点	102
A5 谓词	103
A6 算符	104
A7 GB/T 15936 形式规范的谓词	106
A8 进一步举例	106

前 言

本标准等同采用国际标准 ISO 8613-10—1989《信息处理 文本和办公系统 办公文件体系结构 (ODA)和交换格式 第10部分:形式规范》。

通过制定这项国家标准,使信息处理文本和办公系统的文件便于交换。

本标准与 ISO 8613-10 的区别如下:

a) 正文和附录中引用其他标准时,用我国的标准编号代替相应的国际标准编号,其对应关系是:

GB 2311—1990 代替 ISO 2022:1986;

GB/T 7408—1994 代替 ISO 8601:1988;

GB 8565—1988 代替 ISO 6937:1983;

GB/T 15936 代替 ISO 8613:1989。

b) 在原文中提到了附录 B、附录 C 和附录 D,但并未给出这些附录,所以在本标准中将出现这些附录的字样删去了。

c) 根据编写国家标准的基本规定,本标准保留了被采用国际标准的前言,同时增加了“前言”。

GB/T 15936 在《信息处理 文本和办公系统 办公文件体系结构和交换格式》总标题下,目前包括以下 8 个部分:

第 1 部分(即 GB/T 15936.1):引言和导则;

第 2 部分(即 GB/T 15936.2):文件结构;

第 4 部分(即 GB/T 15936.4):文件轮廓;

第 5 部分(即 GB/T 15936.5):办公文件交换格式(ODIF);

第 6 部分(即 GB/T 15936.6):字符内容体系结构;

第 7 部分(即 GB/T 15936.7):光栅图形内容体系结构;

第 8 部分(即 GB/T 15936.8):几何图形内容体系结构;

第 10 部分(即 GB/T 15936.10):形式规范。

本标准的附录 A 是提示的附录。

本标准由中华人民共和国电子工业部提出。

本标准由电子工业部标准化研究所归口。

本标准起草单位:电子工业部标准化研究所。

本标准主要起草人:杨霖、罗韧鸿、段祥、李韵琴。

ISO 前言

ISO(国际标准化组织)是由各个国家标准化机构(ISO 的成员体)联合组成的一个世界性组织。该组织通过其各个技术委员会进行国际标准的制定工作。凡是对于已设有技术委员会的某一专业感兴趣的每一个成员体,都有权参加该技术委员会。与 ISO 有联系的官方和非官方国际组织也可参与国际标准的制定工作。ISO 与国际电工委员会(IEC)在电子技术标准化的所有方面都进行密切合作。

各个技术委员会提出的国际标准草案,须先分发给各成员体表决通过后,再由 ISO 理事会批准为国际标准。根据 ISO 工作导则,国际标准至少需要投票成员体的 75%赞成。

国际标准 ISO 8613.10 是由 ISO/TC97“信息处理系统”技术委员会制定的。

本国际标准是与下述标准并行制定的:

- ECMA-101(1985):办公文件体系结构;
- CCITT T. 73(1984):用于信息通信服务的文件交换协议;
- CCITT T. 410 系列(1988):开放文件体系结构(ODA)和交换格式。

目前,ISO 8613 由 8 个部分组成:

- 第 1 部分:引言和总则;
- 第 2 部分:文件结构;
- 第 4 部分:文件轮廓;
- 第 5 部分:办公文件交换格式(ODIF);
- 第 6 部分:字符内容体系结构;
- 第 7 部分:光栅图形内容体系结构;
- 第 8 部分:几何图形内容体系结构;
- 第 10 部分:形式规范。

注:目前没有第 3 部分和第 9 部分。

本国际标准可增补新的部分。

附录 A 只是提示的附录。

中华人民共和国国家标准

信息处理 文本和办公系统 办公文件体系结构(ODA)和交换格式 第10部分：形式规范

GB/T 15936.10—1996
idt ISO 8613-10:1991

Information processing—Text and office systems—
Office Document Architecture(ODA)and interchange
format —Part 10:Formal specifications

1 范围

GB/T 15936 的目的是为了便于文件的交换。

在 GB/T 15936 中,将文件看成是诸如备忘录、信件、清单、表格和报告的一些项,它可包括图形和表格材料。文件中使用的内 容元素可以包括图形字符、几何图形元素和光栅图形元素,它们都可包含在同一个文件中。

注: GB/T 15936 设计成允许扩充的,以便在其中包括印刷特征、颜色、传单及诸如声音内容的附加类型。

GB/T 15936 适用于通过数据通信或交换存储媒体进行的文件交换。

GB/T 15936 为下列两种目的或其中之一提供文件交换:

- 允许按始发者的意图进行显现;
- 允许进行诸如编辑和重新格式化处理。

交换中,文件的组成可有以下几种形式:

- 允许文件显现的格式化形式;
- 允许文件进行处理的可处理形式;
- 允许显现和进行处理的格式化可处理形式。

GB/T 15936 还提供了用于处理被交换文件的 ODA 信息结构的交换。

此外,GB/T 15936 还允许包括一种或多种不同类型内容的文件交换,诸如字符文本、图像、图形和声音。

本标准的主要内容如下:

- 规定一种适合描述文件结构(GB/T 15936.2)、文件轮廓(GB/T 15936.4)和内容体系结构(GB/T 15936.6,GB/T 15936.7,GB/T 15936.8)的技术规范的形式描述技术;

- 使用上述形式描述技术,给出文件结构、文件轮廓和内容体系结构形式规范。

使用开放文件体系结构形式规范(FODA)的用途是使用形式语法和形式语义,对 GB/T 15936 第2、4、6、7、8 部分的技术规范提供一个精确的且无歧义的解释。

FODA 能用来作为

- 实现 GB/T 15936 的基础;
- 验证一致性系统的确认工具;
- 检查 GB/T 15936 未来扩充和修订的基点。

如果在自然语言文本和形式规范之间发现存在矛盾,则在解决此矛盾之前,将自然语言文本看作此

国家标准的有效解释。

2 引用标准

下述标准中所包含的一些条文,通过在本标准中引用而构成了本标准的条文。本标准出版时,所示版本均为有效。所有标准都会被修订,使用对本标准的各方应探讨使用下列标准最新版本的可能性。

GB 2311—1990 信息处理 七位和八位编码字符集 代码扩充技术(idt ISO 2022:1986)

GB/T 7408—1994 数据元素和交换格式 信息交换 日期和时间表示法(idt ISO 8601:1988)

GB 8565.2—1988 信息处理 文本通信用编码字符集 第二部分:图形字符集

GB/T 15121.1—94 信息处理系统 计算机图形 存储和传递图片描述信息的元文卷 第1部分:功能描述(idt ISO 8632-1:1987)

GB/T 15121.3—1996 信息处理系统 计算机图形 存储和传递图片描述信息的元文卷 第3部分:二进制编码(idt ISO 8632-3:1992)

GB/T 15936.1—1995 信息处理 文本与办公系统 办公文件体系结构(ODA)和交换格式 第一部分:引言和总则(idt ISO 8613-1:1989)

GB/T 15936.2—1996 信息处理 文本与办公系统 办公文件体系结构(ODA)和交换格式 第二部分:文件结构(idt ISO 8613-2:1989)

GB/T 15936.4—1996 信息处理 文本与办公系统 办公文件体系结构(ODA)和交换格式 第四部分:文件轮廓(idt ISO 8613-4:1989)

GB/T 15936.6—1996 信息处理 文本与办公系统 办公文件体系结构(ODA)和交换格式 第六部分:字符内容体系结构(idt ISO 8613-6:1989)

GB/T 15936.7—1996 信息处理 文本与办公系统 办公文件体系结构(ODA)和交换格式 第七部分:光栅图形内容体系结构(idt ISO 8613-7:1989)

GB/T 15936.8—1995 信息处理 文本与办公系统 办公文件体系结构(ODA)和交换格式 第八部分:几何图形内容体系结构(idt ISO 8613-8:1989)

ISO 9541-5¹⁾ 信息技术 字体信息交换 第5部分:字体属性和字符模型

3 定义

GB/T 15936.1 中给出的定义适用于本标准。

4 规范语言的语法和语义

本章描述用于形式规范的形式描述技术。

注:附录 A 单独给出该形式描述技术的说明。

4.1 基本概念

GB/T 15936 通过抽象的信息构造描述了文件结构、文件轮廓和内容体系结构,这种信息构造是从下列结构类别得到的:

——一个 ODA 构造可以是一个原子构造,例如:客体标识符内的属性名或自然数;

——一个 ODA 构造可以是一个复合构造,即它可包含其他构造。至于它们之间的相互关系,可用下面的三种复合来区分。一个 ODA 构造可以是其他 ODA 构造的。

- a) 一个集;
- b) 一个函数(映射);
- c) 一个序列(表,串)。

1) 即将出版

例如：一个特定布局描述为各组元的一个集，一个组元是一个为函数或映射（从属性名到属性值）的指派，“下级”或“客体标识符”的属性值是一个（原子自然数的）序列。

用于 GB/T 15936 形式规范的语言正是利用了这些结构。该语言称作“复合信息模型化语言”（IM-CL）。此规范语言的语义包含下述抽象元素：

——全域，它是下列各种实体的非空集：

- a) 构造，
- b) 点，
- c) 点集（即多个点的集），
- d) 实体 UNDEF（“未定义”）；

——从全域到全域的函数，即全域实体上的算符；

——全域中的谓词，即全域实体上的谓词。

一个构造是下列信息客体之一：

——原子构造，或简称原子；

——复合构造，或简称复合体，它可以是：

- a) 一个聚合，该聚合是成分构造的一个无序集，

b) 一个指派，它是一个能看成有序对组成的无序集的函数，在此集中每一对都含一个名字和一个值，

- c) 一个连接，它是由成分构造组成的序列。

使用复合构造这一专门名词是为了与其他集、函数或序列区别开来。

为了能对任意复合结构的构造中的各个成分进行讨论，这里引入点的概念。这个概念是与指向一个在某位置上的信息结构并声明“在此处”相关的直观思维的抽象对应物。然而，一般“在此处”不能由成分构造来唯一地标识（例如，在某个字中，同一字母可多次出现），而要由其所出现的上下文来标识。为了从概念中处理“在此处”，需要有一种标识上下文的途径。

引入点的概念使得能明确区别所考虑的构造及其作为某个复合构造的成分在此复合构造中的位置。例如，字符串“data”（一种连接）含‘d’、‘a’和‘t’等成分构造。其中‘d’和‘t’各自只出现在一个点，‘a’出现在两个点，即出现在从头数起的第2和第4个位置上。因此，“data”具有4个成分点，但只含3个成分构造。如果认为某个构造不在上下文范围内，则称其为独点。

各个点通常是由选择准则来选定的。然而，某种选择准则不一定是唯一的。因此，处理起来最自然的客体并不是点，而是多个点的集或点集。从而没有对单点的表达式，只有对单点集的表达式，即只含一个点的点集（见4.3中 $\hat{c}$ 和4.4中 $\hat{^}$ ）。

应该注意：规范语言是建立在一阶谓词逻辑和数学集合论的基础之上的。

4.2 规范语言的语法

该条定义规范语言的语法，即形式规范中每一表达式均使用本条给出的语法规则来建立。语法规则中出现的终结符的语义在4.3至4.5中规定。

元语言摘要：

符号对 $\{\}$ 、 $[\]$ 和 $- \dots -$ 及符号 $::=$ 、 $|$ 、 $\equiv$ 、 $\dots$ 均属于元语言。其含义为：

$::=$ 将被定义的元变量（左边）与定义它的元语言表达式（右边）隔开。

$\{\}$ 限定一个语法单位。

$[\]$ 限定一个语法单位并指出该语法单位是可选的，即它也可以不出现。

注1：元语言符号 $[和]$ 不同于在空常量、显式复合项和扩充聚合项的产生式规则中所使用的特殊字符 $[和]$ 。

$- \dots -$ 限定元语言文本中的注释。

$|$ 分隔可替换的语法单位，即指明对多个语法单位中某一个的选择。

例如： $\{a|m|p\}x$ 的含义是 ax 或 mx 或 px 。

注2: 元语言符号|不同于表示集的项中使用的特殊字符|。

… 它是递归定义的方便记号,此符号跟在可能出现一次或多次的语法单位之后,即该语法单位可重复出现几次。

例如: $\{yf\}$ …的含义是 yf 或 $yfyf$ 或 $yfyfyf$ 等;

$i[so]$ …的含义是 iso 或 $isoso$ 或 $isososo$ 等。

≡ 语言定义中的一个空格符在定义的语言表达式中要求有一处或多处空白。正好相反,符号≡要求规范语言中相邻的串直接并置在一起。如果不影响语法唯一性的话,可以省去规范语言表达式中的空白(例如括号前和括号后)。

为了增强可读性,将符号“(和)”用作元变量(而不是如左定界符或右定界符的字符)。所有其他的元变量均为由小写字母组成的串(用连字符连接元变量的各个成分)。

产生式规则:

表达式::=

公式|项

公式::=

原公式|

not 公式|公式{and|or|impl|iff|xor}公式| $\exists$ 变量(公式)| $\forall$ 变量(公式)|(公式)

注3: 此产生式规则中所使用的终结符具有一阶谓词逻辑通常的语义;not 是逻辑非,and、or、xor(异或)、impl(蕴含)和iff(当且仅当)为通常的逻辑连接词, $\forall$ (所有)和 $\exists$ (存在)是逻辑量词。

原公式::=

[参数部分]谓词符号部分…[参数部分 谓词符号部分…]…[参数部分]

谓词符号部分::=

大写字母[≡字母|≡数字]…≡小写字母[≡字母|≡数字]…|≡|≠|<|≤|>|≥|∈|∉|⊂|
⊆|⊃|⊇

注4: 此产生式规则中终结符(=, ≠, …, ⊇)的语义在4.3中规定。

项::=

变量|常量|算符项|显式复合项|条件项|扩充聚合项|扩充点集项|点选择项|(项)

变量::=

小写字母[≡字母|≡数字]…[≡下标数字]…

常量::=

标准常量|非标准常量

标准常量::=

UNDEF|空常量|数原子常量

空常量::=

[]—空聚合.—|

[:]—空指派.—|

[→]—空连接.—|

<>—空点集.—

数原子常量|::=

[+≡|−≡]数字[≡数字]…[≡.≡数字[≡数字]…]

非标准常量::=

,≡字符[≡字符]…≡,—. 对撇号加限制.—

算符项::=

[参数部分]算符符号部分…

[参数部分 算符符号部分…]…[参数部分]

算符符号部分::=

大写字母[≡大写字母|≡数字|≡—]…|

^|+|-|*|/|U|∩|\\|/|.|'|↑|↓

注 5:本产生式规则中终结符(^,+,…↑)的语义在 4.4 中规定。

显式复合项::=

[项[;项]…]—. 聚合. —|

[项:项[;项:项]…]—. 指派. —|

[→项[→项]…→]—. 连接. —|

“≡字符[≡字符]…≡”—. 字符连接,对引号的限制. —

条件项::=

IF 公式 THEN 项 ELSE 项

注 6:该产生式规则中终结符(IF,THEN 和 ELSE)的语义在 4.5 中规定。

扩充聚合项::=

[变量|公式]—. 公式所适用构造的聚合. —

扩充点聚合项::=

<变量|公式>—. 公式所适用单点集的并集. —

点选择项::=

项 点选择子句|

椭圆点选择项

点选择子句::=

<变量||公式>|

<公式>—. x_s 假定为变量. —|

<名字规范[,名字规范]…>

椭圆点选择项::=

项{. '| ↓ | ↑ }名字规范

注 7:该产生式规则中终结符(. ,', ↓ 和 ↑)的语义在 4.5 中规定。

名字规范::=

非标准常量|变量

参数部分::=

项|(项[,项]…)

字符::=

字母|数字|下标数字|特殊字符

字母::=

大写字母|小写字母

大写字母::=

A|B|C|D|E|F|G|H|I|J|K|L|M|N|O|P|Q|R|S|T|U|V|W|X|Y|Z

小写字母::=

a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|v|w|x|y|z

数字::=

0|1|2|3|4|5|6|7|8|9

下标数字::=

0|1|2|3|4|5|6|7|8|9

特殊字符::=

.|,|;|+|-|~|.等.-

4.3 含固有语义的谓词符号

谓词符号部分序列即指一个谓词符号。对于每个 n 元谓词符号,都有一个在规范语言全域上 n 元谓词,即全域实体上的一种 n 元关系。某些谓词符号具有下面所介绍的固有语义。

注:谓词符号部分在语法上不同于算符符号部分和变量。

True	表示有效事实(所陈述的事物为真)
False	表示无效事实(所陈述的事物为假)
IsAtom(t)	表示 t 是一个原子构造,或简称原子
IsNat(t)	表示 t 是一个自然数(除零外,1,2,...)
IsInt(t)	表示 t 是一个整数(...,-1,-2,0,1,2,...)
IsReal(t)	表示 t 是一个实数
IsCol(t)	表示 t 是一个聚合
IsNom(t)	表示 t 是一个指派
IsCat(t)	表示 t 是一个连接
IsSpotset(t)	表示 t 是一个点集
IsSingle(t)	表示 t 是一个单点集
$t_1 = t_2$	表示 t_1 等于 t_2 (所有实体)
$t_1 \neq t_2$	表示 $t_1 = t_2$ 的非(t_1 与 t_2 不等)
$t_1 < t_2$	表示 t_1 小于 t_2 (数)
$t_1 \leq t_2$	表示 t_1 小于或等于 t_2
$t_1 > t_2$	表示 t_1 大于 t_2
$t_1 \geq t_2$	表示 t_1 大于或等于 t_2
$t_1 \in t_2$	表示 t_1 是 t_2 (聚合)的元素
$t_1 \notin t_2$	表示 t_1 不是 t_2 的元素
$t_1 \hat{\in} t_2$	表示 t_1 是单点集并且是 t_2 的子集(点集)
$t_1 \subset t_2$	表示 t_1 是 t_2 的子集(聚合或点集)
$t_1 \subseteq t_2$	表示 t_1 是 t_2 的子集或者等于 t_2 (聚合或点集)
$t_1 \supset t_2$	表示 t_2 是 t_1 的子集(聚合或点集)
$t_1 \supseteq t_2$	表示 t_2 是 t_1 的子集或者与 t_1 相等(聚合或点集)

一元谓词符号表示的是表达某个实体属于某一特定类别或类型实体的谓词,这种谓词表达某个实体保持某一特定的性质。二元谓词符号所指的谓词指明两个实体间是否具有特别的关系。

4.4 含固有语义的算符符号

一算符符号部分序列指一个算符符号。对于每个 n 元算符符号,都有一个从规范语言全域至全域的 n 元算符或函数,即从 n 元组实体至全域实体的映射。某些算符符号具有下面所介绍固有语义。

注:算符符号部分在语法上不同于谓词符号部分和变量。对于所有的算符来说,如果某一参数项不满足下面的所述的要求,则结果均是未定义的(UNDEF)。

Ct	如果 t 表示一个单点集,则 Ct 表示在 t 所给点的成分构造。
Nt	如果 t 是一个点的单点集且该点是一个指派的成分(形式术语为指派的“直接内含”),则 Nt 表示在指派中该成分的名字构造。
Ft	如果 t 表示仅由一个点组成的集,且它是一个连接点的“直接内含”,则 Ft 表示连接中 t 的前面部分,并且它与 t 连但不包含 t 所给成分(位置号小于 t 的组件的连接)。

CARD t	如果 t 是一个聚合或点集,则 CARD t 表示它的基数,即成分构造个数或点数。
R t	如果 t 表示仅由一个点组成的集,且它是一个连接的“直接内含”,则 R t 表示 t 所给成分之后的(不含 t)连接的后面部分(位置号大于 t 的组件的连接)。
$t_1 + t_2$ $t_1 - t_2$ $t_1 * t_2$ $t_1 \setminus t_2$	如果 t_1 和 t_2 都是数,则它们表示算术运算的数。通常的算术算符的优先规则适用于此处。
$t_1 \cup t_2$ $t_1 \cap t_2$ t_1 / t_2	如果 t_1 和 t_2 表示聚合或点集,则它们分别表示聚合或点集的并、交、差。
$t_1 // t_2$	如果 t_1 和 t_2 都是连接,则 $t_1 // t_2$ 表示按所给的次序并置这两个连接所获得的最终连接。
$\hat{t}$	如果 t 表示一个构造,则 $\hat{t}$ (t 的独点)表示含 t 独点的单点集。
$t \cdot$	如果 t 表示不含原子点(原子构成点)的点集,则 $t \cdot$ (“下级内含”)表示这样一个集,该集中所有点的均是点集 t 的各点的直接内含。
$t \downarrow$	如果 t 表示一个点集,则 $t \downarrow$ 表示 t 给定点所有内含终结点组成的集,或者对原子和空构造, $t \downarrow$ 表示与 t 所给点相等的所有终结点组成的集(称为“最低内含”)。
t°	如果 t 表示不含独点的点集, t° 表示这样一个集,其所有点均是 t 所给各点的直接外部(称为“上级外部”)。
$t \uparrow$	如果 t 表示一个点集,则 $t \uparrow$ 表示 t 给定点所有外部独点组成的集,或者对独点, $t \uparrow$ 表示与 t 所给点相等的所有独点组成的集(称为“最高外部”)。
$t \cdot$ $t \downarrow$ t° $t \uparrow$	如果 t 表示空点集,则算符项也表示空点集。

表达式的正常求值顺序为从左至右,但有以下例外情况:

- 如果某一项在括号内,则首先对该项求值;
- 算符优先于谓词;
- 各个算符之间的优先顺序为:

1. : $\hat{}$ (独点)
2. : $\cdot$ $\downarrow$ $\uparrow$ (下级内含,上级外部,最低内含,最高外部)
3. : 点选择条款
4. : 点选择项
5. : 其他算符

4.5 其他项

除算符项以外,还有其他产生构造或点集的复合项(或“未定义”UNDEF)。下面介绍它们的固有语义。

$[t_1; t_2; t_3]$	如果 t_i 表示构造,则整个项表示以构造 t_i 为成分的聚合。(它是显式复合项的一个例子)
$[\quad]$	表示空聚合。(它是空常量的例子)
$[n_1; c_1; n_2; c_2]$	如果 n_i 和 c_i 都表示构造,且所有 n_i 都不同,则整个项表示以(唯一)名字 n_i 下的构造 c_i 为成分的指派。(这是显式复合项的例子)
$[:]$	表示空指派。(这是空常量的例子)

$[\rightarrow m_1 \rightarrow m_2 \rightarrow m_3 \rightarrow]$	如果 m_i 表示构造,则整个项表示包含 m_i 作为指派序列中成分(亦可是成员)的连接。(这是显式复合项的例子)
“ODA Part ₂ ”	表示连接 $[\rightarrow 'O' \rightarrow 'D' \rightarrow 'A' \rightarrow " \rightarrow 'P' \rightarrow 'a' \rightarrow 'r' \rightarrow 't' \rightarrow " \rightarrow '2' \rightarrow]$ 。括在引号内的一串字符表示这些字符的连接。串中的引号对代表连接中的单个字符。(这是显式复合项的例子)
$[\rightarrow]$	表示空连接。(这是空常量的例子)
$\langle \rangle$	表示空点集。(这是空常量的例子)
IF 公式 THEN t_1 ELSE t_2	如果 t_1, t_2 都是项,则上面的整个项为 t_1 或 t_2 ,它分别取决于公式计算结果为真或假。(这是条件项的例子)
$[\text{变量} \text{公式}]$	表示满足公式的所有构造变量的聚合。(这是扩充聚合项例子)
$\langle \text{变量} \text{公式} \rangle$	表示点集,该点集是满足公式的所有单点集变量的并。(这是扩充点集项的例子)
$t \langle \text{变量} \parallel \text{公式} \rangle$	如果 t 表示一个(可能为空的)点集,则整个项表示所有下述单点集的并:这些单点集含取自 t 的一个点,并且公式对于它们为真。(这是点选择项的例子),下面是为经常出现的点选择条款提供的三种椭圆记号:
$t \langle \text{公式} \rangle$	如果没有明显地引入变量 var ,则按标准变量 xs (检查的单点集,或简称为检查点)对简写项 $t \langle \text{公式} \rangle$ 求值。(这是点选择项的例子)
$t \langle n_1, n_2, \dots \rangle$	如果公式的结构为 $N \text{ 变量} = n_1$ 或 $\text{变量} = n_2$ 或 $\dots$ 其中 n_i 是名字规范,则该公式可以简写为一个名字规范表。(这是点选择项的例子)
$t.n \quad t \downarrow n \quad t^*n \quad t \uparrow n$	如果只有一个名字规范 n 用于点选择,则给出一个椭圆点选择项作为特殊点选择项的简写(以 $\cdot, \downarrow, ^*, \uparrow$ 等结束)。 n 代表 $\langle NXS = n \rangle$ (见程序设计语言中的名字限定)。(这是椭圆点选择项的例子)

4.6 记号简化

对相邻逻辑限定条件可使用通常的记号简化方法。下述例子解释这些常用于一阶谓词逻辑的“简化”记号:

表达式 $\forall x(\forall y(\exists z(\text{公式})))$

可写成 $\forall x \forall y \exists z(\text{公式})$

或甚至写成 $\forall x, y \exists z(\text{公式})$

更进一步的简写有助于强调一个被限定公式的主要部分:

表达式 $\forall x(x \in m \text{ impl 公式})$

可写成 $\forall x \in m(\text{公式})$

表达式 $\exists x(x \in m \text{ and 公式})$

可写成 $\exists x \in m(\text{公式})$

上述两种记号可结合起来使用:

表达式 $\forall x(x \in m \text{ impl } \forall y(\exists z(z \in p \text{ and 公式))))$

可写成 $\forall x \in m, y \exists z \in p(\text{公式})$

但是,要注意表达式

$\forall y, x \in m \exists z \in m(\text{公式})$

包括了 $y \in m$ 的限制,因此该表达式要扩展为

$\forall y(y \in m \text{ impl } \forall x(x \in m \text{ impl } \exists z(x \in m(\text{公式})))$ 。

对 $\in$ 可作同样的简写。

5 形式规范的结构

本章概述形式规范的一般概念。在整个形式规范中多处使用的项在第 6 章给出。

形式地讲,每一形式规范都是一阶谓词逻辑的单个公式。此公式称作总体公式,在靠近每一形式规范的开始处给出。它包含由“and”所连接的其他公式:

公式 1 and 公式 2 and 公式 3 and...公式 n

在本标准中,将每个公式称作一个“定义”并用一个唯一的参考号来标识。每个定义规定 GB/T 15936 其他描述部分所使用的一个概念;或者规定在整个定义中具有辅助作用的概念,即它被单独列出以使其他定义更加易读。

所有的定义被分成几个组。例如,在文件结构的形式规范中,有关组元集的定义包含在 7.1 内,有关组元的定义在 7.2 内,而有关属性的定义在 7.3 内。另外,形式规范中未使用但 GB/T 15936.2 的文本中使用过的概念在 7.5 中定义。

定义的分解只是为了方便规范制定者和读者;它决不会对规范的形式严密性带来任何损害。

定义中出现的变量都用全域量词($\forall$)或存在量词($\exists$)来限定。因此一旦为某个变量选定了值,则不管变量在何处出现,在整个量词范围内该值均被保留。

不从属于规范语言的谓词均用同一种格式来定义。对一元谓词,其格式为:

$\forall$ 变量(谓词符号(变量)iff 公式)

n 元谓词使用类似的格式,各个变量按“自然”的方式置于谓词符号的两旁,例如:(id)IsIdContIn(b)将理解为“id 是基本客体 b 中一个内容部分的标识符”。

一阶谓词逻辑中的一个谓词是一种命题模式,其中某几个位置留空以插入各个实体。“...大于...”是该模式的一个例子。一旦各个空位由各实体所占据,该模式就成为一个可对其求“真值”的命题。命题“5 大于 2”的求值结果为真,而命题“3 大于 5”和“Mercedes 大于 SC18”的最终求值结果为假。注意:一个命题决不能求得一个未经定义的真值。

一元谓词是仅含一个留空参数位置的一种形式。其定义由用单个实体代替其参数来求值。它在 iff 的左边产生一个命题,并将括号内的整个表达式转换为一个命题。

对大多数算符的定义也采用了一种统一的格式。它依赖于规范语言的条件项:

$\forall$ 变量(算符符号(变量))=IF 公式 THEN 项 ELSE UNDEF

一个“算符”是将某些位置留空以插入各个实体的一种项模式。“...加...”是该模式的一个例子。一旦各个空位由各个实体所占据,该模式就成为一个项,即对其求值结果为一个实体或表示一个实体的表达式。项“5 加 2”表示实体 7,而项“WG3 加 SC18”表示由形式规范语言中 UNDEF(未定义)所表达的实体。因此,如果对插入的实体未定义该算符,则项的求值结果为“未定义的”。

应该注意,形式规范不能直接应用于办公文件交换格式(GB/T 15936.5),并且读者不应希望能得到任何直接的对应关系。形式规范是基于 GB/T 15936 概念的文字描述,并不一定反映了 ODA 文件的 ODIF 编码。GB/T 15936 中规定 ODIF 编码的各条款通常对交换的数据流的构成还要施加补充规则。可以认为这些补充规则不在形式规范的范畴内:ODIF 编码正式规定了这些补充规则,因此没必要在 FODA 中重复这些补充规则。

此外,尽管不要求可缺省的属性或参数出现在交换的数据流中,但对形式规范可以认为它们出现,因为它们不同于非必备性属性,它们的值必须通过 GB/T 15936.2 的 5.1.2.4 中定义的缺省机制得到。可缺省属性或参数的概念由谓词 IsPlaceholder 作出形式描述(见第 6 章)。

在整个形式规范中,只要 v 是 GB/T 15936 之外的其他国际标准或国家标准中所定义的一个实体,就要用谓词“IsAtom(v)”。例如

- 1 $\forall v$
- 2 $({}_0\text{IsGB8565MSCharacter}^{1,21}(v))\text{iff}$
- 3 $\text{IsAtom}(v)_0)$

用来指明实体 v 在 GB8565 中定义,它是 GB8565.2 最小子集中一个字符。该实体的定义不在 GB/T 15936 的范围内。

在第 7 章中,对已在某个附录中给出精确定义的实体 v ,当需要该实体的“初始”定义时,也使用谓词“ $\text{IsAtom}(v)$ ”。

便利时可使用“引入定义”来代替 $\text{IsAtom}(v)$ 。一个引入的定义是由 GB/T 15936 之外的其他国际标准或国家标准规定其语义的一个函数或谓词。在本标准的范围内,只用半形式化描述并参引合适的国际标准来表达“引入的定义”的语义。这些定义的形式规范不在本标准的范围内。

下面的提示信息将有助于读者理解形式规范。注意:除了应用为逻辑符号和算符符号(以及括号)所规定的优先级规则而产生的顺序外,在一个公式中不存在阅读或求值顺序。总公式并不打算包含定义的顺序,也无需按所示的顺序来阅读。相反,此顺序只是为了支持规范的系统化。对于依次由公式组成的任何定义来说同样也是如此。

本部分中使用上标指示一个谓词符号或算符符号的定义号(例如, $\text{IsGB8565MSCharacter}^{1,21}(v)$)。如果该符号是规范语言中的部分(即它是一个预定义的符号)或是当前正在定义的符号,则不出现上标。这些上标(引用号)不是规范语言的一部分,它们只是为了方便读者,使读者能迅速找出一个被引用符号的定义。

此外,还使用行号和带下标的括号来构造公式并建立半形式化描述和形式定义之间的交叉引用(见上面)。这些行号和下标也不是形式规范的一部分,只是为了方便读者。

形式规范的每个定义都由一个所谓半形式化描述来加以介绍。这些半形式化描述充分利用规范方法的术语,给出各个定义的英文解释。半形式化描述本身不属于形式规范,但它们将有助于对形式定义理解。

在半形式化描述中,对行号的引用置于括号内,例如:(8)或(11,12)或(4—10);而对带下标的括号内的表达式的引用则需要冠以前缀“ expr ”,例如:(expr .5)。

6 常用定义

本章包含的定义不专门针对文件结构的形式规范、文件轮廓的形式规范或某个内容体系结构的形式规范作出规定。

半形式化描述 1.1

谓词“是一非空聚合”

如果实体 col 是不为空的一个聚合,则它是一个非空聚合(构造集)。

定义 1.1

- 1 $\forall col$
- 2 $({}_0\text{IsNeCol}(col))\text{iff}$
- 3 $\text{IsCol}(col)\text{and } col \neq []_0)$

半形式化描述 1.2

谓词“是一个非空指派”

如果实体 nom 是不为空的一个指派,则它是一个非空指派(名字到构造的映射)。

定义 1.2

- 1 $\forall nom$
- 2 $({}_0\text{IsNeNom}(nom) \text{ iff } \underline{\hspace{1cm}})$
- 3 $\text{IsNom}(nom) \text{ and } nom \neq [:]_0)$

半形式化描述 1.3

谓词“是一个非空连接”

如果实体 cat 是一个不为空的连接, 则它是一个非空连接(构造序列)。

定义 1.3

- 1 $\forall cat$
- 2 $({}_0\text{IsNeCat}(cat) \text{ iff } \underline{\hspace{1cm}})$
- 3 $\text{IsCat}(cat) \text{ and } cat \neq [\rightarrow]_0)$

半形式化描述 1.4

谓词“是一个空聚合”

如果实体 col 不含成分, 则它是一个空聚合。

定义 1.4

- 1 $\forall col$
- 2 $({}_0\text{IsEmptyCol}(col) \text{ iff } \underline{\hspace{1cm}})$
- 3 $col = []_0)$

半形式化描述 1.5

谓词“是一个空指派”

如果实体 nom 不含成分, 则它是一个空指派。

定义 1.5

- 1 $\forall nom$
- 2 $({}_0\text{IsEmptyNom}(nom) \text{ iff } \underline{\hspace{1cm}})$
- 3 $nom = [:]_0)$

半形式化描述 1.6

谓词“是一个空连接”

如果实体 cat 是不含成分, 则它是一个空连接。

定义 1.6

- 1 $\forall cat$
- 2 $({}_0\text{IsEmptyCat}(cat) \text{ iff } \underline{\hspace{1cm}})$

3 $cat = [\rightarrow]_0)$

半形式化描述 1.7

谓词“是一个非负整数”

如果实体 int 是一个自然数或零,则它是一个非负整数。

定义 1.7

- 1 $\forall int$
- 2 $({}_0IsNnInt(int)iff$
- 3 $IsNat(int)orint=0_0)$

半形式化描述 1.8

谓词“是一个非负整数对”

一非负整数对是两个非负整数的一个连接。

定义 1.8

- 1 $\forall q$
- 2 $({}_0IsPairOfNnInt(q)iff$
- 3 $\exists l,r$
- 4 $({}_1q=[\rightarrow l\rightarrow r\rightarrow]andIsNnInt^{1.7}(l)andIsNnInt^{1.7}(r)_1)_0)$

半形式化描述 1.9

谓词“是一正整数对”

一正整数对是两个正整数的一个连接。

定义 1.9

- 1 $\forall q$
- 2 $({}_0IsPairOfPosInt(q)iff$
- 3 $\exists l,r$
- 4 $({}_1q=[\rightarrow l\rightarrow r\rightarrow]andIsNat(l)andIsNat(r)_1)_0)$

半形式化描述 1.10

谓词“是一个八位位组串”

八位位组串是形式规范的一个原子构造。

定义 1.10

- 1 $\forall v$
- 2 $({}_0IsOctetString(v)iff$
- 3 $IsAtom(v)_0)$

半形式化描述 1.11

函数“相对前端的位置”

PF 返回指派成分的位置编号,计数从第一个成分开始,在所考虑的成分处结束且包含这一成分。要求操作数 p 表示一个连接内直接内含恰好一个点的点集。

定义 1.11

- 1 $\forall p$
- 2 $({}_0\text{PF}(p) =$
- 3 IF $\text{IsSingle}(p) \text{ and } \text{IsCat}(Cp^*)$
- 4 THEN $\text{LENGTH}^{1.16}(\text{F}p) + 1$
- 5 ELSE UNDEF_0)

半形式化描述 1.12

函数“相对后端的位置”

PR 返回指派成分的位置编号,计数从最后一个成分开始,在所考虑的成分处结束且包含这一成分。要求操作数 p 表示一个连接内直接内含恰好一个点的点集。

定义 1.12

- 1 $\forall p$
- 2 $({}_0\text{PR}(p) =$
- 3 IF $\text{IsSingle}(p) \text{ and } \text{IsCat}(Cp^*)$
- 4 THEN $\text{LENGTH}^{1.16}(\text{R}p) + 1$
- 5 ELSE UNDEF_0)

半形式化描述 1.13

函数“第一成分”

HEAD 返回一个非空连接 cat 的第一成分。

定义 1.13

- 1 $\forall cat$
- 2 $({}_0\text{HEAD}(cat) =$
- 3 IF $\text{IsNeCat}^{1.3}(cat)$
- 4 THEN $C(\wedge cat, \langle xs \parallel \text{PF}^{1.11}xs = 1 \rangle)$
- 5 ELSE UNDEF_0)

半形式化描述 1.14

函数“最后成分”

LASTC 返回一个连接(cat)的最后成分。

定义 1.14