



中华人民共和国国家标准

GB/T 17532—1998
eqv ISO/DIS 1087-2-2:1996

术语工作 计算机应用 词汇

Terminology work—Computer applications—Vocabulary

1998-11-05 发布

1999-06-01 实施

国家质量技术监督局 发布

中 华 人 民 共 和 国
国 家 标 准
术 语 工 作 计 算 机 应 用 词 汇
GB/T 17532—1998

*

中国标准出版社出版
北京复兴门外三里河北街16号
邮政编码:100045

电 话:68522112

中国标准出版社秦皇岛印刷厂印刷
新华书店北京发行所发行 各地新华书店经售
版权专有 不得翻印

*

开本 880×1230 1/16 印张 1½ 字数 41 千字
1999年7月第一版 1999年7月第一次印刷
印数 1—1 200

*

书号: 155066 • 1-15943 定价 13.00 元

*

标 目 378—13

前 言

本标准等效采用 ISO/DIS 1087-2-2:1996《术语工作 词汇 第二部分 计算机应用》。

本标准根据我国的实际情况和汉语、汉字的特点,在等效采用 ISO/DIS 1087-2-2:1996 时,主要做了以下改动:

1. 根据我国术语工作标准的命名习惯,将标准名称改为《术语工作 计算机应用 词汇》。
2. 把某些术语注中的内容,直接引入到定义中,把定义中的内容改为注。
3. 增加了部分中文例。
4. 从 GB/T 12200.2—1994《汉语信息处理词汇 02 部分:汉语和汉字》中引入了 8 条有关汉字字符的术语。
5. 删去了附录 B 文献目录。

本标准是术语数据库的系列国家标准之一,已经发布的配套标准有:

GB/T 13725—1992 建立术语数据库的一般原则与方法

GB/T 13726—1992 术语与辞书条目的记录交换用磁带格式

GB/T 15387.2—1994 术语数据库开发指南

GB/T 15387.1—1994 术语数据库开发文件编制指南

GB/T 15625—1995 术语数据库技术评价指南

GB/T 16785—1997 术语工作 概念与术语的协调

GB/T 15786—1997 术语工作 计算机应用 数据类目

本标准由全国术语标准化技术委员会提出。

本标准由国家语言文字工作委员会语言文字应用研究所归口。

本标准由国家语言文字工作委员会、中国标准化与信息分类编码研究所、中国大百科全书出版社、北京大学计算语言学研究所、中国科技信息研究所等单位起草。

本标准由全国术语标准化技术委员会计算机辅助术语工作分技术委员会负责解释。

本标准的附录 A 和附录 B 是提示的附录。

本标准主要起草人:冯志伟、李莉、王渝丽、俞士汶、周智佑、宋正良、李竹等。

ISO 前言

国际标准化组织(ISO)是各国家标准化团体(ISO 成员团体)的一个世界范围的联盟。国际标准的制定工作通常是通过 ISO 的技术委员会进行的,对某个已成立了技术委员会的专业领域感兴趣的成员团体都有权参加该委员会。与 ISO 有联系的官方和非官方的国际组织也参与这项工作。ISO 在电工技术标准化方面与国际电工委员会(IEC)紧密合作。

委员会所采纳的国际标准草案需分发给各成员团体投票表决,作为国际标准发布时要求至少 75% 的成员团体投票批准。

国际标准 ISO 1087-2 是由 ISO/TC 37/SC 3“术语工作(原则与协调)技术委员会第 3 分委员会——计算机辅助术语工作分委员会”制定的。

作为 ISO 1087:1990 第 7 章的技术修订,本标准废止并同时取代该标准的这部分。ISO 1087:1990 主要章节的技术修订将发展为 ISO 1087-1。

附录 A 和附录 B 是提示性附录。

目 次

前言	I
ISO 前言	II
0 导言	1
1 范围	1
2 引用标准	1
3 一般概念	2
4 数据组织	4
5 术语数据的筛选	5
6 字符	6
7 数据存储	8
8 信息检索	9
9 数据管理	10
10 数据验证	11
11 操作	11
12 相关的应用领域	12
附录 A(提示的附录) 中文索引	14
附录 B(提示的附录) 英文索引	16

中华人民共和国国家标准

术语工作 计算机应用 词汇

GB/T 17532—1998
eqv ISO/DIS 1087-2-2:1996

Terminology work—Computer applications—Vocabulary

0 引言

本标准的条目按顺序给出,每部分设一个一般性的标题。在通常的情况下,后面的条目原则上应该由前面出现的条目来定义。

条目的格式按 GB/T 1.6—1997 的规定。

条目的内容顺序如下:

条目编号

优先术语(黑体)

缩写形式(黑体)

许用术语

拒用术语:括号中注明“(拒用)”

专业领域用尖括号< >

定义 引用的术语用黑体标出,并在后面用括号注明条目编号

例

注

除了条目编号之外,优先术语和定义仅在适当的地方出现。

1 范围

本标准规定了在术语工作和术语编纂中用于语言和信息处理的术语。

本标准适用于术语数据库的研究、开发、维护及管理工作,在其他涉及术语数据处理的工作中也可参考使用。

2 引用标准

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

GB/T 12200.2—1994 汉语信息处理词汇 02 部分:汉语和汉字

GB/T 12991—1991 信息处理系统 数据库语言 SQL(idt ISO/IEC 9075:1989)

GB/T 15237—1994 术语学基本词汇(neq ISO 1087:1990)

GB/T 5271.8—1993 数据处理 词汇 08 部分:控制、完整性和安全性(eqv ISO 2382-8:1986)

ISO/IEC 2382-1:1993 信息技术 词汇 第1部分:基本术语

ISO 2382-4:1987 信息处理系统 词汇 第4部分:数据的组织

ISO 2382-6:1987 信息处理系统 词汇 第6部分:数据的准备和处理

ISO/IEC 2382-9:1994 信息技术 词汇 第9部分:数据通信

ISO/IEC 2382-23:1994 信息技术 词汇 第23部分:文本处理

3 一般概念

3.1 信息 information

〈信息处理〉关于客体(如事实、概念、事件、思想、过程等)的知识,它在一定的上下文中具有特定的意义。

注

1 本条引自 ISO/IEC 2382-1。

2 在其他的应用领域,信息的定义不同。

3.2 数据 data

为进行通信、解释和处理而使用的信息(3.1)的形式化表现形式。

注:本条改自 ISO/IEC 2382-1。

3.3 数据处理 data processing

DP

对数据(3.2)进行的系统操作。

例:对数据进行算术运算或逻辑运算,数据的归并(9.5)或分类(9.4),程序的汇编或编译,以及对文本(3.6)的操作,如文本编辑(12.3)、分类、归并、存储、检索、显示(9.6)、打印等。

注:本术语不能作为信息处理(3.4)的同义术语。

3.4 信息处理 information processing

对信息(3.1)进行的系统操作,它包含数据处理(3.3)。

注

1 本条改自 ISO/IEC 2382-1。

2 本术语不能作为数据处理(3.3)的同义术语。

3.5 语言处理 language processing

对语言进行的系统操作,它包括数据处理(3.3)。

3.6 文本 text

以字符(6.1)、符号、词、短语、句子、段落、表格或其他的符号序列的构成的用于表达意义的结构化数据(3.2)。其解释主要根据阅读文本的人对于某种自然语言或人工语言的知识来进行。

注:本条引自 ISO/IEC 2382-1。

3.7 文本语料库 text corpus

语料库 corpus

〈自然语言处理〉根据预先确定的规则所准备、编码或存储的机器可读文本(参阅 7.4)或文本的某些部分的有组织集合。

注:文本语料库可按专业领域、容量或时代做不同的限定,例如,从 1986 年以来的某些特定的期刊、数学文本等。

文本语料库可用作进一步的语言分析或术语工作的原材料。

3.8 类型 type

〈自然语言处理〉文本(3.6)中代表一个确定类别的语言单位。

注:这种语言单位通常是未用定界符(3.10)隔开的不间断的字符串。

3.9 类例 token

〈自然语言处理〉类型(3.8)在文本中的具体表现形式。

例:在英语中,如果把 good 的所有词形定义为类型,那么 good, better 和 best 等都是词形 good 的类例。

3.10 定界符 delimiter

分隔符 separator

用于指明一个字符串(6.13)的开始或结尾的一个或多个字符(6.1)。

注

1 本条引自 ISO 2382-4。

2 空白或标点符号经常用作定界符。

3.11 词形 word form

给定词的任何形态句法变体。

例：在英语中，indicate；indicates，
visitor；visitor's。

注

1 在屈折语中，词形经常是屈折形式，例如，英语的 go，goes；以及 go，went，gone。

2 这个定义不包含正词法方面的变体。

3.12 词形变化范型 paradigm

〈自然语言处理〉属于某一给定词的所有词形(3.11)的类聚。

3.13 基本词形 base form

参照词形 reference form

根据词典编纂学的惯例选出的用以表示一个词形变化范型(3.12)中的各个形式的词形(3.11)。

例：在英语中，bind 是 bind，bound，binds，binding 等的基本词形。

注

1 本条改自 GB/T 15237—1994。

2 术语“基本词形”也可以应用于词组型术语。

3.14 词组型术语 multi-word term

包括两个以上词的术语。

3.15 压缩形 compressed form

规定形 normalized form(拒用)

缩减形 reduced form(拒用)

经过压缩(8.4)的字符串。

例：字符串“input/output-algorithm”变成其压缩形“inputoutputalgorithm”。

注：术语“缩减形”和“规定形”可能会导致误解，建议避免使用这两个术语。

3.16 取消屈折 deinflection

取消词形(3.11)中的屈折成分。

3.17 取消屈折的词形 deinflected word form

在取消屈折(3.16)之后余下的词段(3.18)。

3.18 词段 word part

词片 word segment

为了某种特殊用途从一个词形(3.11)中取出的字符串(6.13)。

3.19 词形还原 lemmatization

从某一给定词形(3.11)生成基本词形(3.13)的过程。

例：在英语中，“go”是“goes”通过取消屈折(3.16)而得到的基本词形(3.13)；而“go”是“went”通过不规则动词变换得到的基本词形(3.13)，这种变换不符合标准的屈折规则。

注：这样的结果也称为词形还原。

3.20 剖析 parsing

根据给定的算法，将给定的结构分解为其组成成分的操作。

例：在英语句子“all unsaturated fatty acids are not degradable by biological methods”中，“all unsaturated fatty acids”和“biological methods”可看成是词组型术语(3.14)，它们可以被抽取(8.9)。

注：剖析不一定必须提供对某一句子的完整分析。

3.21 术语数据集合 terminological data collection

包含特定专业领域有关各种概念的信息(3.1)数据(3.2)集合。

3.22 术语条目 terminological entry

术语数据集合(3.21)中所包含的关于一个概念的术语数据(3.2)。

注：一个术语条目可以包含两个以上的记录(7.9)。

3.23 同形词 homograph

两个以上的具有相同书写形式但表示不同的概念(语义同形)或不同句法功能(句法同形)的词形或词。

例：在英语中，lead(铅 Pb)和 lead(领导)；

bark(吠)和 bark(树皮)。

在汉语中，仪表(人的外表)和仪表(测量温度、压力等的仪器)；

杜鹃(布谷鸟)和杜鹃(映山红)。

注：具有不同书写形式但经过压缩(8.4)之后变为同形的词形(3.11)不算同形词。

3.24 歧义消解 disambiguation

通过赋予同形词贴切的概念或贴切的句法功能从而分化同形词(3.23)的过程，或者通过赋予同形词组以不同的语言解释来分化同形词组的过程。

例：在汉语中，分化“白跑”和“白纸”中的“白”分别为副词和形容词；分化“学习文件”的句法结构分别为动宾结构和偏正结构。

4 数据组织

4.1 分类值 sort value

排序值 sorting value

根据预先确定的顺序，字符集(6.2)中某一元素的位置。

例：在法语中，字母 A 具有比字母 B 较低的分类值。一个小写字母是否与它相应的大写字母具有相同的分类值，取决于实际应用的需要。带发音符号的字母有时按其相应的基本字母来处理，有时按不同的字母来处理。

4.2 分类关键字 sort key

排序键

用于满足分类(9.4)和归并(9.5)操作要求的字符串(6.13)。

例：当给图书数据分类以便产生作者目录时，作者的姓是第一分类关键字，作者的名是第二分类关键字。如果同一个作者有两个题目，则把出版年份或题目作为附加的分类关键字。

4.3 字母排序 alphabetical ordering

在组织字符串(6.13)时，表中的每一个串(6.12)的位置唯一地由从该串(6.12)头部开始的分类值(4.1)来确定。

注

1 带有发音符号和连音符号的字母以及带有数字、上标、下标的其他符号都可以作为特殊的分类值。

2 字母排序的规则可以因语言的不同而不同。

4.4 逆字母排序 reverse alphabetical ordering

在组织字符串(6.13)时，表中的每一个串(6.12)的位置唯一地由从该串(6.12)尾部开始的分类值(4.1)来确定。

4.5 频度顺序 frequency order

在特定的文本(3.6)或文本语料库(3.7)中，根据类例(3.9)出现频度上升或下降的顺序来排列的

类型(3.8)的顺序。

注：在通常情况下，表的类型是词形(3.11)频度表或原形词频度表。

4.6 轮排 permutation

使词组型术语中每个实词都作为关键词进行的排序。

例：在英语中，对字符串“millions of instructions per second”[MIPS]轮排时，“instructions, millions of per seconds”, “second, millions of instructions per”等形式分别出现在“instructions”和“second”等实词的排序表中，这样可以保证术语中任何想要的成分都可以出现在相应字母的排序位置。

注：本条改自 GB/T 15237—1994。

4.7 毗连 concatenation

两个以上的字符串(6.13)按特定的顺序合并，形成一个新的串(6.12)，其长度等于各个字符串(6.13)长度的和。

5 术语数据的筛选

5.1 非用词表 exclusion list

停用词表 stop word list

在数据处理(3.3)中任意选择的不予考虑的字符串(6.13)组成的表。

注

1 在术语工作中，产生忽略功能词(代词、冠词等)的词表可能是有益的。

2 有时，非用词表中的字符串可以被保留下来(例如在词语索引(5.4)中)，但是不注明其频度。

5.2 拟用词表 inclusion list

加用词表 plus word list

要保存或认为要进一步数据处理(3.3)的字符串(6.13)组成的表。

例：凡包含“bank”(银行)或“credit institution”(信用机构)的所有的句子都抽出；

凡以“M”起头的所有的条目在名字索引中都检索出来；

凡以“anti-”开头的所有的单词都选出来。

注：如果适合的话，拟用词表也可以包含词段(3.18)或其他的字符串(6.13)。

5.3 自由文本搜索 free-text search

在文本语料库(3.7)中进行的，能够检索任何类型(3.8)的搜索(8.7)。

5.4 词语索引 concordance

<自然语言处理>按字母顺序排列的词形(3.11)表，其词形(3.11)是从原文中抽取(8.9)出来的，包括要检索的词形(3.11)以及该词形(3.11)在原文中的前面部分和后面部分。

注：词语索引通常的形式是 KWIC(上下文关键词)词语索引和句子词语索引。

5.5 索引 index

<自然语言处理>从一个出处摘出的并属于同一类数据元的有序字符串(6.13)汇集表。

注：本条改自 GB/T 15237—1994。

5.6 术语索引 term index

参照词汇或术语数据库(7.6)中的术语条目(3.22)而编制的索引(5.5)。

注：术语索引中的术语可以按照字母来排序，也可以按照其他系统来排序。

5.7 字母索引 alphabetical index

按字母顺序排列的索引(5.5)。

5.8 系统索引 systematic index

所有的术语排列能反映特定专业领域中概念之间的系统关系的索引(5.5)。

5.9 轮排索引 permuted index

按轮排(4.6)方式列出的术语索引(5.6)。

5.10 逆序索引 reverse index

按逆字母排序(4.4)列出的索引(5.5)。

5.11 倒排索引 inverted index

列出文本(3.6)中的全部类型(3.8)以及全部对应类例(3.9)的索引(5.5)。

5.12 词形索引 word form index

词索引 word index

文本(3.6)中所有词形(3.11)的索引,其中在文本(3.6)中出现两次以上的词形(3.11)用一个单独的类型(3.8)来代表,并指出其出现次数(即相应的类例(3.9))以及它们的原文参照。

5.13 还原词索引 lemmatized word index

用词形(3.11)的基本词形(3.13)来代表的词形索引(5.12)。

6 字符

6.1 字符 character

用于表示、组织或控制数据(3.2)的元素集中的成员。

注

1 本条引自 ISO 2382-4。

2 字符可以分为图形字符和控制字符。

6.2 字符集 character set

不同字符(6.1)的有限集合,它对于一定的目的而言是完整的。

例:ISO 646 的国际参照本;

128 个 ASCII 字符。

注:本条引自 ISO 2382-4。

6.3 字母表 alphabet

按公认的线性次序排列的字母字符(6.6)集。

注

1 本条改自 ISO 2382-4。

2 这个定义也覆盖了由字符(6.1)组成的自然语言的字母表,这些字符(6.1)包含带发音符号的字母。

6.4 图形字符 graphic character

一种具有视觉表示的,通常可以书写、印刷或显示(9.6)的字符(6.1),它不同于控制字符(6.5)。

注

1 本条引自 ISO 2382-4。

2 图形字符可以是字母字符(6.6)、数字(6.7)或特殊字符(6.9)。

6.5 控制字符 control character

在特定上下文中并实现某种控制功能的字符(6.1)。

注

1 本条引自 ISO 2382-4。

2 控制字符(6.1)启动、修改和停止数据处理(3.3)的操作。

6.6 字母字符 alphabetic character

图形字符(6.4)中用于书面表示给定语言的字母。

注:它包括带发音符号的字母。

6.7 数字 digit

数字字符 numeric character

表示非负整数的字符(6.1)。

注：本条引自 ISO 2382-4。

6.8 字母数字字符 alphanumeric character

或者是字母字符(6.6),或者是数字(6.7)的任何字符(6.1)。

6.9 特殊字符 special character

不是字母数字字符(6.8)的图形字符(6.4)。

例：任何标点符号, and 符号[&], 百分符号[%], 加号[+]等都是特殊字符。

注：本条改自 ISO 2382-4。

6.10 间隔字符 space character

引起打印或显示位置沿着直线向前移动一个位置的**控制字符**(6.5)。

注：本条改自 ISO 2382-4。

6.11 空白 blank

在**图形字符**(6.4)串中表示一个空位置的**字符**(6.1)。

注：本条引自 ISO 2382-4。

6.12 串 string

由具有相同性质的元素组成的作为总体来看的序列。

例：字符(6.1)、二进制位或脉冲等组成的序列。

注：本条改自 ISO 2382-4。

6.13 字符串 character string

未被空白隔开的由多个**字符**(6.1)组成的序列。

注：一个字符串常常代表一个单词。

6.14 模样 pattern

当用组合**字符**(6.1)替换某些**字符**(6.1)时定义的**串**(6.12)。

注：本条参照 GB/T 12991。

6.15 汉字 Chinese character

记录书面汉语的文字。

6.16 汉字字形 character form

特指构成方块**汉字**(6.15)的二维图形。

注：本条引自 GB/T 12200.2。

6.17 笔画 stroke

构成楷书**汉字字形**(6.16)的最小连笔单位。汉字最基本笔画有横(一)、竖(丨)、点(丶)、撇(丿)、折(乙)等。

注：本条引自 GB/T 12200.2。

6.18 笔顺 stroke order

书写每个**汉字**(6.15)时的**笔画**(6.17)的次序和方向。

注：本条引自 GB/T 12200.2。

6.19 笔数 stroke count

构成一个**汉字**(6.15)或**汉字部件**(6.20)的**笔画**(6.17)数。

注：本条引自 GB/T 12200.2。

6.20 汉字部件 Chinese character component

由**笔画**(6.17)组成的具有组配**汉字**(6.15)功能的构字单位。现代汉字部件按是否成字可分为成字部件与非成字部件。

例：成字部件：木 心 口 也

非成字部件：讠 亻 刂 礻

注：本条引自 GB/T 12200.2。

6.21 部首 indexing component

一部分可以成批构字的部件。凡含有某一部件构成的字在字典中均排列在一起,该部件作为领头单位排在开头,成为查字的依据,称为部首。部首多为形旁构成。

例:木(松、柏、杨、架、杏、林、相……)

注:本条引自 GB/T 12200.2。

6.22 偏旁 radical

合体字的构字单位的传统称呼,旧称合体字(由两个或两个以上汉字部件构成),左为偏,右为旁,现代统称偏旁。偏旁本为独体字,在古代汉字(6.15)中偏旁一般与单独成字时形状相同,现代汉字(6.15)偏旁有一些变化,如“心”在左为“忄”;“刀”在右为“刂”;“火”在下为“灬”;左边的“埠”和右边的“邑”均作“阝”等。

注:本条引自 GB/T 12200.2。

7 数据存储**7.1 数据媒体 data medium**

用来记录数据(3.2)并且检索数据(3.2)的物理媒体。

例:磁盘,光盘,缩微胶片。

注

1 本条改自 ISO/IEC 2382-1。

2 在图书和文献工作的领域,数据媒体被看作数据载体。

7.2 格式 format

〈数据处理〉数据(3.2)预先确定的组织形式。

注:格式可进一步分为输入格式、存储格式和输出格式。

7.3 数据采集 data acquisition

收集数据(3.2)或将数据(3.2)装入、存储在数据处理系统中的过程。

7.4 机器可读数据 machine-readable data

可直接装入数据处理系统的数据(3.2),简称机读数据。

注:通常指磁盘、磁带和光盘中的数据,在一定环境下,采用光学字符识别(OCR)的办法,印刷物上的数据也可以被看成机器可读数据。

7.5 数据库 database

按照概念结构组成的数据(3.2)的集合。

注:本条改自 ISO/IEC 2382-1。

7.6 术语数据库 terminological database

包含术语数据的数据库(7.5)。

7.7 数据库集 data bank

包含管理框架的数据库(7.5)的集合。

7.8 术语库集 term bank

术语数据库集 terminological data bank

含有术语数据(3.2)的数据库集(7.7)。

7.9 记录 record

作为一个单位来对待的数据元(7.11)的集。

注:本条引自 ISO 2382-4。

7.10 文件 file

文卷

作为一个单位来存储或处理的命名记录(7.9)的集。

注：本条引自 ISO 2382-4。

7.11 数据元 data element

在一定的上下文具有区别特征的数据(3.2)单元。

注：该条改自 ISO 2382-4。

7.12 数据字段 data field

为特定的数据元(7.11)而保存在一个记录中的变长或定长的部分。

7.13 数据字段目录 data field directory

在一个记录(7.9)内的数据字段(7.12)的位置索引(5.5)，包括在该记录(7.9)之内的每一个数据字段(7.12)的标号(7.16)、长度和位置。

7.14 数据类目 data category

数据元类型 data element type

关于给定数据字段(7.12)的类型说明。

7.15 标识符 identifier

用于标识或命名一个数据类目(7.14)的一个或多个字符(6.1)。

注

1 该条改自 ISO 2382-4。

2 标识符的命名可表明该数据类目(7.14)的某些性质。

7.16 标号 label

数据元(7.11)集内或附加到数据元(7.11)集上的用于标识数据(3.2)的标识符(7.15)或关键字。

注：该条改自 ISO 2382-4。

7.17 可重复类目 repeatable category

在同一个记录(7.9)中能够出现两次以上的数据类目(7.14)。

例：在多语言术语条目(3.22)中，有些记录(7.9)的某些类目可以对每一种语言都重复出现。在单语言术语条目(3.22)中，诸如同义术语、出处、上下文等类目也可以重复出现。

8 信息检索

8.1 信息检索 information retrieval

IR

从存储的数据(3.2)中查找信息(3.1)的行动、方法和过程。

注

1 本条改自 ISO/IEC 2382-1。

2 信息检索可以在交互式处理或批处理中进行。

8.2 查询语言 query language

<信息检索>用户用来检索或者修改存储在数据库(7.5)中数据(3.2)的数据操纵语言。

8.3 截断 truncation

<信息检索>为了搜索或选择而将词形(3.11)变为词段(3.18)的过程。

注：截断过程可以自动地进行，或者让用户自行选择。截断的方向可以从左向右，也可以从右向左。

8.4 压缩 compression

<自然语言处理>为了数据处理(3.3)的目的，把词形(3.11)转换为简化形式的操作。

注：压缩一般是指去掉发音符号，消除空白(6.11)和各种特殊字符(6.9)以及不管大小写的区别。例如，当把某些拼写变体中立化以便生成索引(5.5)时，就需要使用压缩。

8.5 搜索词 search word

搜索关键字 search key

<信息检索>用于提供访问记录(7.9)的字符串(6.13)。

注：搜索词可以是术语、词、字等(英语中还包括单词或者单词的一部分)。

8.6 词干搜索 stem search

递降分析 degradation

按照预先确定的规则,逐次去掉字符串(6.13)中的字符(6.1),使用自动生成的搜索词(8.5)随意提问的过程。

注：根据不同语言中构词方式的不同,其顺序可以从右到左,也可以从左到右。

8.7 搜索 search

<信息检索>自动地检查数据(3.2)的集合,以匹配(8.8)或抽取(8.9)具有给定特性的元素。

例：在姓名的集合中查出所有以J开头的姓名。

8.8 匹配 match

<信息检索>通过比较来确定不同的数据(3.2)集之间项的同一性,以便选择或分离具有同一性的项。

8.9 抽取 extract

<信息检索>从一组项中,选择并取出某些符合预先确定的性质的项。

8.10 加权 ranking

<信息检索>根据搜索响应对于相应的搜索提问的相关程度的不同,给它们指派不同权值。

9 数据管理

9.1 读 read

从数据媒体(7.1)获取数据(3.2),并保持数据不变。

注：本条改自 ISO 2382-6。

9.2 写 write

将数据(3.2)长期或暂时记录在数据媒体(7.1)上。

注：本条改自 ISO 2382-6。

9.3 拷贝 copy

复写

将数据(3.2)从源数据媒体(7.1)读(9.1)出,保持源数据不变,然后写(9.2)到目标数据媒体上。

例：将一个文件从硬盘上拷贝到软磁盘上。

注：本条改自 ISO 2382-6。

9.4 分类 sort

按指定的准则将一些项分离成若干个组,对其加以排列整理的动作。

注：本条改自 ISO 2382-6。

9.5 归并 merge

将两个以上的具有相同给定顺序的数据(3.2)的集合中的项合并成具有该顺序的一个更大的集合。

注：本条引自 ISO 2382-6。

9.6 显示 display

在屏幕上对数据(3.2)进行可视的显现。

9.7 屏蔽码 mask

掩码

利用某一字符(6.1)模样(6.14)来控制另一字符(6.1)模样(6.14)中那些要保留或者要删除的部分。

例：所有的元音 a,e,i,o,u 都可用表示元音类别的字符 V 来替换。

注：本条改自 ISO 2382-6。

9.8 转换 convert

将数据(3.2)的表示从一种形式改变成另一种形式,而不改变数据所负载的信息(3.1)内容。

例:代码转换

模拟数字转换

媒体转换

注:本条引自 ISO 2382-6。

9.9 转写 transliterate

〈自然语言处理〉用一种字母表(6.3)中的字符(6.1)表示另一种字母表(6.3)中的相应字符。

注:ISO 2382-6 中有适合信息处理的定义。

10 数据验证

10.1 数据验证 data validation

用于确定数据在形式上是否准确、完备、合理的过程。

注:本条改自 GB/T 5271.8。

10.2 重复条目检查 double-entry check

确定数据条目是否重复存储过的数据验证(10.1)。

注:有时允许重复条目是有益的,例如,当一个术语可用于不同的专业领域时,可允许重复条目存在。

10.3 一致性检查 consistency check

确定彼此依赖的条目是否遵守预先确定的条件的数据验证(10.1)。

例:交叉引用的检查。

10.4 拼写检查 spelling check; spell check

确定所有的词是否遵守预先确定的拼写规则的数据验证(10.1)。

注:ISO/IEC 2382-23 在信息处理中,定义了文本编辑中上下文的拼写检查器。

10.5 完备性检查 completeness check

确定所要求的数据是否都出现的数据验证(10.1)。

注

1 本条改自 GB/T 5271.8。

2 数据必须在强制的数据字段(7.12)内出现。

10.6 格式检查 format check

确定数据(3.2)是否符合于指定格式(7.2)的数据验证(10.1)。

注:本条改自 GB/T 5271.8。

10.7 合理性检查 plausibility check; reasonableness check

确定数据(3.2)是否符合于特定准则的数据验证(10.1)。

注:本条改自 GB/T 5271.8。

11 操作

11.1 批处理 batch processing

数据(3.2)的处理或者作业的完成都以事先累积起来的方式进行,在工作过程中不再接受用户的进一步影响。

11.2 交互式处理 interactive processing

在用户和计算机系统之间的以对话方式进行的一系列输入和响应的操作。

11.3 数据保护 data protection

实施一套合适的管理、技术或物理等手段,以防止有意地或偶然地泄漏、修改或破坏数据(3.2)。

注:本条引自 ISO/IEC 2382-1。

11.4 更新 update

数据(3.2)的添加、删除或修改。

注：本条改自 ISO 2382-4。

11.5 文件维护 file maintenance

更新(11.4)或重组文件的活动。

注：本条引自 ISO 2382-4。

11.6 可移植性 portability

＜计算机程序设计＞同一程序在不同类型的数据处理系统中不必转换成另外的程序语言，或不必修改或者只作很少的修改就能执行的能力。

注：本条改自 ISO/IEC 2382-1。

11.7 兼容性 compatibility

功能部件不做可感觉到的修改而满足不同接口要求的能力。

注：本条引自 ISO/IEC 2382-1。

11.8 用户终端 user terminal

在一个系统或通信网络中，使用户能够与计算机进行通信的功能部件。

注：本条改自 ISO/IEC 2382-1。

11.9 选单 menu

数据处理系统显示(9.6)的项目表，用户从此表中能够选择将要进行的行动。

注：本条引自 ISO/IEC 2382-1。

11.10 数据通信 data communication

根据数据传输(11.11)协议和数据交换的协调规定，在功能单元之间进行的数据(3.2)传送。

注：本条引自 ISO/IEC 2382-9。

11.11 数据传输 data transmission

通过通信设施，从一个点到另一个或多个点的数据(3.2)传送。

注：本条引自 ISO/IEC 2382-9。

11.12 文件传送 file transfer

通过电子手段，一个或多个文件从一个数据处理系统到其他的数据处理系统的转移。

11.13 上载 upload

把程序或数据(3.2)从相连的计算机或数据媒体(7.1)转出到另外的计算机中。

11.14 下载 download

把程序或数据(3.2)从相连计算机转入到另外的计算机或数据媒体(7.1)中。

12 相关的应用领域**12.1 计算机辅助 computer-aided; computer-assisted**

机助 CA

借助计算机完成人的部分工作的有关技术或过程。

例：计算机辅助术语工作，计算机辅助翻译，计算机辅助辞书编纂。

12.2 文本处理 text processing

语词处理 word processing

对文本(3.6)进行的数据处理(3.3)操作。

注

1 本条改自 ISO/IEC 2382-23。

2 典型的操作有输入、编辑、分类、归并、检索、存储、显示或打印等。

12.3 文本编辑 text editing