ICS 01.020
A 22

中华人民共和国国家标准

GB/T 15387.2—2001
代替 GB/T 15387.2—1994

术语数据库开发指南

Guidelines for the development of terminology database

2001-11-14 发布　　2002-06-01 实施

中华人民共和国
国家质量监督检验检疫总局 发布

需要全本请访问 www.ertongbook.com

目　　次

前言 …… I
1　范围 …… 1
2　规范性引用文件 …… 1
3　术语和定义 …… 1
4　术语条目结构 …… 2
5　术语库的功能 …… 4
6　数据规范化 …… 6
7　其他要求 …… 7
8　术语库的扩充 …… 7
附录 A(资料性附录)　建立术语数据库有关的国家标准和国际标准 …… 8
A.1　术语工作 …… 8
A.2　字符集 …… 8
A.3　代码与数据元的表示 …… 9
A.4　字符转写 …… 9
A.5　系统开发 …… 10
A.6　其他相关标准 …… 10

前　言

GB/T 15387 分为两个部分：

——第 1 部分：术语数据库开发文件编制指南；

——第 2 部分：术语数据库开发指南。

本部分代替 GB/T 15387.2—1994《术语数据库开发指南》。

本部分与 GB/T 15387.2—1994 相比主要变化如下：

——修改了术语和定义一章中的一部分术语，并更新了所有术语的定义。

——第 4 章的标题改为“术语条目结构”；

——“字段的强制性”改为“字段的强制级别”(见 4.7)；

——术语库的功能一章中增加“网络支持”的要求(见 5.10)；

——附录 A 中增加了新制定的相关标准；

——根据 GB/T 1.1—2000《标准化工作导则　第 1 部分：标准的结构和编写规则》的要求，对标准编排的格式进行了修改和调整。

GB/T 15837 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—2001　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 15625—2001　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 16786—1997　术语工作　计算机应用　数据类目

GB/T 17532—1998　术语工作　计算机应用　词汇

GB/T 18155—2000　术语工作　计算机应用　机器可读术语交换格式(MARTIF)协商交换

本部分的附录 A 为资料性附录。

本部分由全国术语标准化技术委员会提出。

本部分由中国标准研究中心归口。

本部分由中国标准研究中心负责起草。

本部分由全国术语标准化技术委员会负责解释。

本部分主要起草人：叶盛、李明非、张志云、肖玉敬、卢丽丽、徐俊荣。

术语数据库开发指南

1 范围

本部分规定了术语数据库(简称“术语库”)术语条目的结构、数据库的功能、扩充以及数据规范化的要求。

本部分适用于术语库的研究、开发、维护及其管理工作。其他涉及术语数据处理的工作也可参照使用。

2 规范性引用文件

下列文件中的条款通过本部分的引用而成为本部分的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本部分,然而,鼓励根据本部分达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本部分。

GB/T 3792.1 **文献著录总则**(GB/T 3792.1—1983)

GB/T 3860 **文献叙词标引规则**(GB/T 3860—1995)

GB/T 8566 **信息技术 软件生存周期过程**(GB/T 8566—2001,ISO/IEC 12207:1995,IDT)

GB/T 13418 **文字条目通用排序规则**(GB/T 13418—1992,ISO 7154:1983,NEQ)

GB/T 13725—2001 **建立术语数据库的一般原则与方法**

GB/T 13726—1992 **术语与辞书条目和记录交换用磁带格式**(ISO 6156:1987,NEQ)

GB/T 13745 **学科分类与代码**(GB/T 13745—1992)

GB/T 15387.1—2001 **术语数据库开发文件编制指南**

GB/T 15237.1—2000 **术语工作 词汇 第1部分:理论与应用**

GB/T 16786—1997 **术语工作 计算机应用 数据类目**(ISO/DIS 12620—1996,EQV)

GB/T 17532—1998 **术语工作 计算机应用 词汇**(ISO/DIS 1087-2-2:1996,EQV)

3 术语和定义

GB/T 15237.1—2000、GB/T 16786—1997、GB/T 17532—1998 确立的术语和定义适用于本部分。为了便于使用,本部分重复了其中的部分术语和定义。

3.1

术语 term

在特定专业领域中一般概念的词语指称。

[GB/T 15237.1—2000,3.4.3]

3.2

数据库 database

长期储存在计算机中,有组织、可共享的数据集合。

[GB/T 17532—1998,7.5]

3.3

术语数据库 terminological database

包含术语数据的数据库。

[GB/T 17532—1998,7.6]

3.4

记录 record

作为一个单位来对待的数据元的集合。

[GB/T 17532—1998,7.9]

3.5

数据元 data element

在一定的上下文中具有区别特征的数据单元。

[GB/T 17532—1998,7.11]

3.6

数据字段 data field

为特定的数据元而保存在一个记录中的变长或定长的部分。

[GB/T 17532—1998,7.12]

3.7

数据类目 data category

数据元类型 data element type

关于给定数据字段的类型说明。

[GB/T 17532—1998,7.14]

3.8

标识符 identifier

用于标识或命名一个数据类目的一个或多个字符。

[GB/T 17532—1998,7.15]

3.9

术语条目 terminological entry

术语数据集合中所包含的关于一个概念的术语数据。

[GB/T 17532—1998,3.22]

4 术语条目结构

在术语库中,根据不同的使用目的,每一条术语及其相关的一组数据元构成一个术语单位。每一数据元占用一个字段(或子字段),具有相同属性的一类数据元构成一个数据字段。

4.1 字段数目

应根据术语库应用的需求确定字段数目,既要按照规范化的要求,最大程度地满足用户需要,又要避免浪费。

4.2 字段名称

应使用易记、易识别的字符串。

例 1:

字段名称:序号(xuhao/number)

缩写:xh(汉语拼音)或 Num(英语)

例 2:

字段名称:定义(dingyi/definition)

缩写:dy(汉语拼音)或 def(英语缩写)

4.3 字段标识符

字段标识符可使用字母或数字,如果使用数字,应符合 GB/T 13726—1992 附录 A 的要求。字段标

需要全本请访问 www.ertongbook.com

识符及其编制原则与方法，应在提供给用户的文件中注明。

4.4 字段数据类型

4.4.1 字符型

可以汉字、字母、数字、特殊符号等。术语、定义等应使用字符型字段。

4.4.2 数字型

可以是整数、小数、正数和负数等。序号应使用数字型字段。

4.4.3 日期型

日期型数据字段结构为：CCYYMMDD。

例：20010515 表示 2001 年 5 月 15 日。

术语源的生成日期（例如标准发布日期，辞书出版日期等）、术语记录的生成日期等应使用日期型字段。

4.4.4 其他

术语数据还可能涉及其他字段数据类型，例如逻辑数据类型等。

4.5 字段长

术语库应尽可能使用变长字段，使用定长字段时应考虑使用方便及节约存储空间等因素。

4.6 字段间关系

根据术语数据元之间关系的特点，字段间的关系可以是一对一（1∶1）、一对多（1∶N）或多对多关系（M∶N）。

4.7 字段的强制级别

字段的强制级别应根据该术语库的工作目的而确定。

4.7.1 必选字段

当术语数据类目是构成术语条目所必需的最小信息时，使用此类字段。例如，面向概念的术语库中的“术语”“定义”；面向翻译的术语库中的“术语”“外文对应词”等。必选字段必须设立，不允许空缺。

4.7.2 应选字段

当术语数据类目是构成术语条目所必要的信息时，使用此类字段，但允许空缺。例如，面向翻译的术语库中的同义词、术语的缩略形式等。

4.7.3 可选字段

当术语数据类目是构成术语条目的可选信息时，使用此类字段。例如，语法信息、语境等。

4.8 字段的可重复性

当属于相同的术语数据类目的术语数据元（如同义词）需要在一个记录中出现几次时，可重复使用同一字段。

4.9 字段的组合性

几乎所有术语数据类目的数据元（字段）均可以与一个或若干术语数据类目的数据元（字段）组合，构成一个新的术语条目。

4.10 子字段

当确定的某一字段可能存在几个值，或一个字段需反映几项内容时，可使用子字段。

a) 一条术语可能有若干个同义词，可根据使用状态进一步划分子字段，例如优先、许用、拒用。

b) 同义性（程度）字段需划分子字段，例如恒等于（≡）、大于（＞）、小于等于（＜＝）、交叉（×）、近义（≈）、反义（←→）等。

4.11 相互参照

不同记录之间（如一条术语参照另一条术语）；不同的文件之间（如主文件与倒排文件）；不同数据库之间（如图形数据库与文本数据库之间，不同语种数据库之间）均存在相互参照。

5 术语库的功能

5.1 用户界面

应能提供良好的汉字处理环境和友好的用户界面。

设计的屏幕应具备以下功能：

a）人机对话。

使用者通过键盘或其他输入设备向机器输入提问或其他有关信息，机器通过显示屏幕或其他输出设备给使用者提供解答、提示或其他有关信息。

b）选单驱动。

应根据需要设置系统管理和用户查询使用的不同选单，应设计一个选单系统为用户显示多重选择，根据需要可设计多级选单。

例如：

一级选单：主功能选择。例如数据输入、数据维护与更新、数据检索、排序和打印等；

二级选单：子功能选择，单项功能中的子项，条件等。例如数据检索功能中的浏览、设定检索式、显示检索结果等；

三级选单：需要设计的下一层细化功能选择。

c）提示功能。

例如：

出错提示：指示错误的性质和类型，以至于如何改正；

拒绝接收提示：指示、告警无效或错误的操作；

操作步骤提示。

d）求助功能。

例如：

正确操作的具体、详细的说明；

有关内容的资料，包括语种代码、学科代码、源机构代码、责任者代码等常用代码及其含义说明；

有关屏幕显示内容的解释，参数值的含义与范围，命令或功能选择描述。

求助功能可使用窗口技术或调用辅助库、辅助屏予以实现，此状态应易于进入和退出、返回原操作状态。

e）用户需要的其他功能。

例如：浏览、窗口等。

5.2 数据检索

术语库除应能检索单一术语及相应信息（定义、外文对应词等）外，还应根据需要设计以下几种类型的检索。

a）内部属性检索。

根据分类语言、叙词语言进行检索。分类语言检索得出的是适合某一分类的某一领域内成体系的全部术语；叙词语言检索出的是某一术语可能涉及的所有术语，包括其他领域内的相关术语。

b）外部属性检索。

例如：

按术语源书目编目数据检索出该文献所包含的所有术语；

按源机构（或责任者）检索出该机构所编纂（制定）的所有术语；

按年代检索出术语编纂的年度变化情况。

c）组合属性检索。

同属性或不同属性的组合检索。

检索技术应满足 GB/T 13725—2001 中 8.2 的各项要求。

5.3 查重与数据校验

应设计查重功能，数据校验功能应根据 GB/T 13725—2001 中 8.1.2 的要求设计。

5.4 术语库重组与重构

a) 根据需要，设计数据库重组功能，自动的回收已删除记录所占的存储空间，并重新加以利用。

b) 当需要时，可改变存取控制、硬件配置、数据库分布以及修改数据结构等。

5.5 恢复、备份和重新启动

应支持术语库的恢复、备份和重新启动。由于误操作、程序出错或其他意外原因(例如掉电)需要重新启动时，使用备份；

由于人为或意外原因造成术语库的破坏时，应能：

a) 查明错误的原因；

b) 查明错误的影响；

c) 纠正错误；

d) 纠正后继续工作。

5.6 定义默认值

应按照 GB/T 13725—2001 中 8.1.3 的要求，根据需要定义默认值，便于字段内容的自动或半自动生成与修改。

5.7 数值处理

应根据需要设计数值处理功能，例如术语和定义的平均长度，某一长度的术语数量及百分比，年术语生成量、使用量、某一用户使用术语库的情况，某一机构编纂术语的情况等。

5.8 排序和打印

字顺排序应符合 GB/T 13418 的规定。

根据需要设计排序和打印功能以满足下列要求。

a) 应能生成和打印按概念体系排序的术语词汇或词典；

b) 应能生成和打印按外文字母顺序或汉语拼音排序的英—汉、汉—英对照词汇或词典，或其他外语与汉语对照词汇或词典；

c) 应能生成和打印同义(近义)词汇或词典；

d) 应能生成和打印反义词汇或词典；

e) 如果需要，还应能生成语料的有关信息，例如按术语使用频率排序、按术语与其他术语组配频率排序等。

f) 其他功能。

5.9 程序设计语言接口

应能提供某种常用程序设计语言的接口，允许使用该程序设计语言编写自己的应用程序，开发某些特殊功能。

5.10 网络支持

应能实现网络查询、数据搜索和数据下载等网络应用的基本功能。

5.11 其他

a) 数据输入应符合 GB/T 13725—2001 中 8.1 的要求；

b) 数据显示应符合 GB/T 13725—2001 中 8.3 的要求；

c) 数据安全性和保密应符合 GB/T 13725—2001 中 6.1.7 和 8.6 的要求；

d) 数据传递和数据交换应符合 GB/T 13725—2001 中 7.7 和第 10 章的要求。

6 数据规范化

6.1 数据采集

应从需要和可能两个方面，并根据 GB/T 13725—2001 中 7.2.1 确定数据采集的要求。

6.1.1 数据需求

a) 拟收录术语的范围及总数量(可分期安排)；

b) 拟使用的语种及使用程度(是仅收录相应语种的外文对应词，还是也收录该术语的定义、同义词等其他数据)；

c) 拟确定的数据类目。

6.1.2 数据保障

a) 术语源的类型，例如标准、百科全书、词典、专业文献等；

b) 术语源的权威性、充足性，例如，多语种的术语库就需要有多语种的文献；

c) 专家构成，包括专业领域的专家、术语学家、语言学家等；

d) 工作人员水平，包括专业水平、外语水平等；

e) 应根据 GB/T 15387.1—2001 中 5.14～5.16 的要求编写数据加工技术规则；

f) 应根据 GB/T 15387.1—2001 中 5.18 的要求编写数据加工工作规程；

g) 应规定数据验收标准和验收方式，包括对各项数据和各加工过程和质量要求、检验方法、依据和程序等。

6.1.3 数据的采集方法

根据 6.1.1～6.1.2 条确定数据采集原则以后，规定数据采集的统一格式(采集卡片或工作单)，根据该统一格式收集、挑选、登录数据。该格式应满足术语库对数据的要求，应与术语库结构一致。

6.2 标引

6.2.1 分类法

为了便于标引工作和国内外术语数据的交换，应尽量采用通用的分类法。

例如：

a) 中国标准文献分类法(CCS)；

b) 国际标准分类法(ICS)；

c) GB/T 13745 学科分类与代码。

6.2.2 叙词表

为了表示某一概念所属的学科领域或主题特性，可使用叙词检索语言。

例如：

a) 汉语主题词表及与其兼容的有关专业性主题词表；

b) 标准文献主题词表(ROOT THESAURUS)。

6.2.3 分类标引和叙词标引

a) 对于通用、基础性术语或选自综合性文献的术语应根据该术语的分类和定义的主题特性进行标引。

b) 某些术语条目的分类标引和叙词标引也可以采取与术语源一致的方法。

某一术语条目取自某一术语标准，则该条目的分类标引和叙词标引与该标准一致。

c) 多重标引。

某一概念属于若干个专业领域时，或术语库需要同时使用多种分类法和叙词表时，需重复标引。

6.2.4 标引规则与标引内容

应遵照 GB/T 3860 的规定，分类标引的内容包括分类法的源码，该术语(或术语源)所属专业领域的分类代码及专业领域名称。

需要全本请访问 www.ertongbook.com

例 1：GB/T 13745(源码)520.4050(分类代码)×××数据库(专业领域名称)。

例 2：UDC(源码)681.3.01(分类代码)×××数据处理原理(专业领域名称)。

叙词标引的内容包括叙词表的源码、叙词代码及叙词。

例：ISO-ROOT(源码)MJG.JM(叙词代码)×××数据库(叙词)

应向用户提供所使用的分类法和叙词表以及标引规则说明，必要时，还应提供不同分类法和叙词表的对照(或转换)说明。

6.3 代码

术语库中所使用的代码，例如：语种代码、学科代码、源码、责任者机构代码等应符合相应的国家标准；如无国家标准可参照时，可自行定义，但必须在向用户提供的技术文件中详细说明。必要时，应建立辅助的代码数据库，供用户随时查询。

6.4 著录

应规定统一的著录项目、规则和格式以及编写著录细则。

术语条目著录应符合 GB/T 3792.1 及相关国家标准的规定。

7 其他要求

a) 建库过程与工作步骤应符合 GB/T 13725—2001 第 7 章的要求。

b) 术语库计算机系统应符合 GB/T 13725—2001 中 6.2 的要求。

c) 术语库有关的软件开发，应遵照 GB/T 8566 的各项规定。

d) 术语库开发过程中需编制有关技术文件，文件内容及编写要求应遵照 GB/T 15387.1 的各项规定。

e) 少数民族语言文字的术语库应尽可能利用拉丁文转写，以便于使用新引入的术语库软件。

8 术语库的扩充

下述情况可能需要扩充术语库：

a) 增加新的专业领域；

b) 增加新语种；

c) 增加新的数据类目；

d) 增加新条目；

e) 增加新的数据类型(如图、公式、特殊符号等)；

f) 增加新功能；

g) 改变软、硬件配置，变换机型；

h) 数据传递，数据转换等。

附 录 A
（资料性附录）
建立术语数据库有关的国家标准和国际标准

A.1 术语工作

A.1.1 国家标准

GB/T 3860 文献叙词标引规则
GB/T 3792.1 文献著录总则
GB/T 4880.2 语种名称代码 第2部分:3字母代码
GB/T 10112 术语工作 原则与方法
GB/T 13190 汉语叙词表编写规则
GB/T 13418 文字条目通用排序规则
GB/T 13715 信息处理用现代汉语分词规范
GB/T 13725 建立术语数据库的一般原则与方法
GB/T 13726 术语与辞书条目的记录交换用磁带格式
GB/T 15237.1 术语工作 词汇 第1部分:理论与应用
GB/T 15238.1 辞书编纂基本术语
GB/T 15387.1 术语数据库开发文件编制指南
GB/T 15417 文献工作 多语种叙词表编制规则
GB/T 15625 术语数据库技术评价指南
GB/T 15933 辞书编纂常用汉语缩略语
GB/T 16617 辞书编纂符号
GB/T 16785 术语工作 概念与术语的协调
GB/T 16786 术语工作 计算机应用 数据类目
GB/T 17532 术语工作 计算机应用 词汇
GB/T 18155 术语工作 计算机应用 机器可读术语交换格式(MARTIF)协商交换
GB/T 20001.1 标准编写规则 第1部分:术语

A.1.2 国际标准

ISO 12620 计算机辅助术语工作——术语交换格式(TIF)—数据元词典
ISO 12200 计算机辅助术语工作——术语交换格式(TIF)—SGML(标准通用置标语言)的应用
ISO/DTR 12618 计算机辅助术语工作—术语数据库和文本数据库的建立和使用
ISO 10241 国际术语标准的制定与编排
ISO/R 860 概念和术语的国际协调
ISO/R 1149 多语种分类词汇的编排
ISO 2788 文献工作 单语种叙词表编制规则
ISO/R 919 分类词汇表编制规则

A.2 字符集

A.2.1 国家标准

GB/T 1988 信息处理 信息交换用七位编码字符集
GB/T 2311 信息处理 七位和八位编码字符集代码扩充技术

GB/T 11383 信息处理 信息交换用八位代码结构和编制规则
GB/T 8565 信息处理 文本通信用编码字符集
GB 2312 信息交换用汉字编码字符集 基本集
GB/T 7589 信息交换用汉字编码字符集 第二辅助集
GB/T 7590 信息交换用汉字编码字符集 第四辅助集
GB/T 12345 信息交换用汉字编码字符集 辅助集
GB/T 13131 信息交换用汉字编码字符集 第三辅助集
GB/T 13132 信息交换用汉字编码字符集 第五辅助集
GB 13000.1 信息技术通用多八位编码字符集(UCS)
GB/T 6513 文献书目信息交换用数学字符编码字符集
GB 8045 信息处理 信息交换用蒙古文七位和八位编码图形字符集
GB 12050 信息处理 信息交换用维吾尔文编码图形字符集
GB 12052 信息交换用朝鲜文字编码字符集
GB 13134 信息交换用彝文编码字符集
GB/T 13141 书目信息交换用希腊字母编码字符集
GB/T 13142 书目信息交换用拉丁字母代码字符扩充集

A.2.2 国际标准

ISO 5427 书目信息交换用基里尔字母扩展字符集
ISO 6438 文献工作 书目信息交换用非洲字母编码字符集
ISO 8859 信息处理 八位单字节编码图形字符集
ISO/DIS 6861 基里尔字母编码字符集
ISO 6630 文献工作 书目控制字符集
ISO/DP 8957.2 希伯莱字母编码字符集

A.3 代码与数据元的表示

A.3.1 国家标准

GB/T 2659 世界各国和地区名称代码
GB/T 3304 中国各民族名称的罗马字母拼写法和代码
GB/T 3469 文献类型与文献载体代码
GB/T 4880 语种名称代码
GB/T 4881 中国语种代码
GB/T 7408 数据元和交换方式 信息交换 日期和时间表示法
GB/T 7156 文献保密等级代码
GB/T 13745 学科分类代码

A.3.2 国际标准

ISO/DIS 6093 信息交换用字符串中的数值表示法
ISO 6523 数据交换 机构标示法的结构
ISO 7064 信息处理 校验码系统
ISO/DP 7352 数据交换用数据元的组织结构及其表示法的导则
ISO/DIS 7826 用于交换的代码表示法的一般结构

A.4 字符转写

A.4.1 国际标准

ISO 9 文献工作 斯拉夫语系基里尔字符转写成拉丁字符
ISO 233 文献工作 阿拉伯字符转写成拉丁字符
ISO 259 文献工作 希伯莱字符转写成拉丁字符
ISO/DP 2805 文献工作 苏联非斯拉夫语言的基里尔字符转写成拉丁字符
ISO 843 文献工作 希腊字符转写成拉丁字符
ISO/DIS 3602 文献工作 日文拉丁化
ISO 7098 文献工作 中文拉丁化
ISO/DP 9984 文献工作 格鲁吉亚字符转写成拉丁字符
ISO/DP 9985 文献工作 亚美尼亚字符转写成拉丁字符

A.5 系统开发

A.5.1 国家标准

GB/T 1526 信息处理 数据流程图、程序流程图、系统流程图、程序网络图和系统资源图的文件编制符号及约定
GB/T 8566 计算机软件开发规范
GB/T 8567 计算机软件产品开发文件编制指南
GB/T 9385 计算机软件需求说明编制指南
GB/T 9386 计算机软件测试文件编制规范
GB/T 12504 计算机软件质量保证计划规范
GB/T 15205 计算机软件配置管理计划规范

A.5.2 国际标准

ISO 6592 计算机应用系统文件编制指南

A.6 其他相关标准

A.6.1 国家标准

GB/T 14814 信息处理 文本和办公系统标准通用置标语言(SGML)
GB/T 9387 信息处理系统 开放系统互连 基本参考模型(系统标准)
GB/T 5271 数据处理词汇(系列标准)

A.6.2 国际标准

ISO 9660 信息处理 信息交换用 CD-ROM 的卷和文件结构
ISO/IEC 10027 信息技术 信息源词典系统结构
ISO 8613,10166 信息处理 文本和办公文件系统(系列标准)
ISO 8632,9592,9593,9636 信息处理系统 计算机制图(系列标准)
ISO/IEC 9066,9072,100021 信息处理 文本通信(系列标准)
ISP/IEC/TR 9294 信息技术 软件管理守则
ISO/IEC 100032 数据管理参考模型

中华人民共和国
国家标准
术语数据库开发指南
GB/T 15387.2—2001
*
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:68523946 68517548
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
*
开本 880×1230 1/16 印张 1 字数 24 千字
2002年4月第一版 2002年4月第一次印刷
印数 1—1 500
*
书号:155066·1-18331 定价 12.00 元
网址 www.bzcbs.com

版权专有 侵权必究
举报电话:(010)68533533

GB/T 15387.2-2001