ICS 01.020
A 22

中华人民共和国国家标准

GB/T 15625—2001
代替 GB/T 15625—1995

术语数据库技术评价指南

Guidelines for technical evaluation of terminological 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　术语库的技术特征与要求 …… 2
5　技术评价的一般原则和方法 …… 2
6　技术评价的主要内容 …… 3
7　评价的实施 …… 5
附录 A(规范性附录)　汉语语言文字规范文件 …… 7

前　言

本标准代替 GB/T 15625—1995《术语数据库技术评价指南》。

本标准对 GB/T 15625—1995《术语数据库技术评价指南》有如下主要修改：

——增加了规范性引用文件：

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

b) GB/T 16786—1997　术语工作　计算机应用　数据类目

c) GB/T 17532—1998　术语工作　计算机应用　词汇

——增加了术语和定义一章；

——修改了 4.1 条的内容；

——在第 4.2 条数据类型中增加了声音、图像等数据类型；

——修改了 7.1.2 条中第一条的内容；

——在 6.2.2 条 e)中的笔画排序、部首排序前加上“汉字”，并增加了“按汉语拼音排序”；

——在 7.4 条中增加了网络安全性一条；

——修改了 8.1 条的内容；

——在规范性附录 A 中增加了 GB/T 15834—1995　标点符号用法

——在规范性附录 A 中增加了 GB/T 16159—1996　汉语拼音正词法基本规则

——修改了文字使用不恰当的地方；

——修改和调整了 GB/T l5625—1995 标准的编排格式，符合 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。

本标准是术语数据库的系列国家标准之一。已经发布的系列标准有：

——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 15387.1—2001　术语数据库开发文件编制指南

——GB/T 15387.2—2001　术语数据库开发指南

——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

本标准的附录 A 是规范性附录。

本标准由全国术语标准化技术委员会提出。

本标准由中国标准研究中心归口。

本标准起草单位：中国标准研究中心、国家邮政局科学研究规划院。

本标准主要起草人：李明非、叶盛、张志云、肖玉敬、卢丽丽、徐俊荣。

本标准所代替标准的历次版本发布情况为：

——GB/T 15625—1995。

术语数据库技术评价指南

1 范围

本标准规定了术语数据库(简称术语库)技术评价的原则、方法、主要内容和实施程序等。

本标准适用于术语库的研究、开发、维护及有关管理工作。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标准的引用而成为本标准的条款。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标准,然而,鼓励根据本标准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标准。

GB/T 10112 术语工作 原则与方法(GB/T 10112—1999,ISO 704:1997,NEQ)

GB/T 13725—2001 建立术语数据库的一般原则与方法

GB/T 13745 学科分类与代码(GB/T 13745—1992)

GB/T 15237.1—2000 术语工作 词汇 第1部分:理论与应用(ISO 1087-1:2000,EQV)

GB/T 15387.1—2001 术语数据库开发文件编制指南

GB/T 15387.2—2001 术语数据库开发指南

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标准。为了便于使用,本标准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

术语数据库 terminological database

包含术语数据的数据库。

[GB/T 17532—1998,7.6]

3.3

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]

3.4

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]

3.5

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]

4 术语库的技术特征与要求

4.1 术语数据

术语数据具有相对的稳定性,术语库中术语数据应具有适时性、动态性和可追溯性。部分术语数据具有多义性和同义性,术语库应能通过概念的定义避免误解和歧义,促进术语的标准化,也应能提供各种多义词、同义词的检索和相互参考。

4.2 数据类型

术语库的系统环境应能支持处理字母、数字、汉字、图形、图像、声音、符号、公式、化学分子式、结构式等多种类型数据。

4.3 数据采集

术语数据来源具有多样性,选择术语数据应与有关的国家标准和国际标准的各项规定保持一致,并根据建库目的和需求,采集和收录数据(见 GB/T 15387.2—2001 第 6.1 条),入库数据应经规范化处理(见 GB/T 13725—2001 第 7.5.2 和 GB/T 15387.2—2001 第 7 章)。

4.4 数据类目

应按 GB/T 13725 的规定选择数据类目,并设置足够的便于检索的可检索字段。

4.5 分类法

术语的分类法应考虑科学性和实用性,以反映学科领域的结构。学科代码应能区分同音词和同形异义词,并确定概念的有效范围和边界。用户可灵活地运用分类代码,获得该类别中全部术语条目(见 GB/T 13745)。

4.6 语言文字

术语库使用汉语、汉字时应遵循国家相关的法律、标准和规定。在汉语的命名方式、语法结构分析、概念体系的建立以及与其他语种概念体系的对应等方面都应符合汉语的文字、词汇、语法和语义等规范(见附录 A)。

4.7 术语库的分类

根据建库目的和用途,术语库可分为面向概念型、面向翻译型、面向特定目标型和其他特殊用途型(见 GB/T 13725—2001 第 5.1 条)。

5 技术评价的一般原则和方法

技术评价是依据某种技术指标体系,采用一定的方法和程序,对术语库及其组成要素的功能、特性和运行效果进行评测或验证。

5.1 评价的原则

a) 力求准确、全面、公正;

b) 充分考察其社会效益和经济效益;

c) 考虑其发展和应用前景;

d) 认真考虑术语库的不同用途、类型和基本特点(见第 4 章);

e) 遵循有关的国家标准、国际标准(见 GB/T 15387.2—2001 的附录 A)以及相关技术规则和规范。

需要全本请访问 www.ertongbook.com

5.2 评价的主要类别

a) 宏观评价：评价术语库的宏观方面，例如术语的数量和质量、收录范围、数据源类型、系统结构、利用率、影响等；

b) 性能评价：评价术语库的使用效果和质量以及满足用户需求的程度；

c) 效益评价：评价术语库提供的服务使用户或库本身获得的社会效益和经济效益；

d) 性能费用评价：评价术语库达到某种性能水平与所需费用之间的关系；

e) 费用效益评价：评价术语库的服务效益与费用消耗之间的关系，以及效益与成本比的合理性。

5.3 评价的方法

5.3.1 专家评价

有关学科、领域的专家从科学的角度出发，根据规定的技术指标，对术语库的设计开发过程和所完成系统的各个方面进行评价。它由以下三个阶段组成。

5.3.1.1 现场测试

按规定的测试大纲和指标，对术语库的运行情况和各种性能进行测试。

5.3.1.2 定性评估

根据现场测试结果以及与术语库相关的基础理论和有关标准，对术语数据的数量和质量，系统的软硬件兼容程度、易学性、智能化程度、易用性、用户自主性、实用性等性能，按级别进行评分。

5.3.1.3 综合评价

根据当前的技术水平和发展趋势，进行纵向和横向比较，对术语库的各方面作出评价。

5.3.2 用户评价

用户在试用过程中，对于术语库所具有的功能、性能、可靠性、易用性、可维护性和效率等进行测试，并从实用的角度，对术语库的使用情况进行评价或与其他术语库进行比较。用户评价多为性能评价，也可以是性能费用评价、效益费用评价和宏观评价等。

6 技术评价的主要内容

术语库技术评价的主要内容如下，评价时可视具体情况适当增减。

6.1 系统环境

系统环境考虑到系统的体系结构及其配置的优化和协调。

6.1.1 硬件环境

a) 硬件配置；

b) 性能价格比；

c) 兼容性和通用性；

d) 可扩充性，需要时能方便地实现升级；

e) 安全性；

f) 可靠性(平均无故障时间)。

6.1.2 网络环境

a) 建构网络环境，能实现局域网、广域网和互联网的使用；

b) 在网络环境中可采用客户/服务器、浏览器/服务器等体系结构，实现对整个系统应用的支持；

c) 网络各结点间的数据传输与通信，能满足多个用户同时访问某个术语库，并保证迅速响应的要求；

d) 数据的分布与应用满足利用效率高、负载均衡合理、资源共享程度大等要求；

e) 网络安全性。

6.1.3 软件环境

a) 软件整体配置的先进性、通用性、易用性；

b) 软件的配套性(如系统软件、多文种处理软件、简繁体汉字兼容软件、汉字字符集和图像处理软件等);

c) 软件的灵活性和可移植性;

d) 软件的安全保密能力(如防病毒、防剽窃、防数据丢失等);

e) 软件的后续开发能力。

6.2 应用支持环境

6.2.1 术语库管理软件

a) 合理、灵活、独立和方便地组织记录与字段结构;

b) 分级的访问权限及安全管理功能;

c) 备份管理和多版本管理功能;

d) 数据结构与控制结构相分离,用户修改数据结构时不会导致改变术语库的程序,修改程序时也不会对数据产生影响;

e) 逻辑结构与物理结构相分离,能在逻辑结构上对系统进行增加、删除、修改等工作,而不需要考虑系统的物理结构,并能利用数据库概念模式,使系统不受信息的具体逻辑和物理结构的限制;

f) 采用一系列新的数据库结构与技术,提高系统性能;

g) 利用数据库的并行技术与分布管理方法,为统一管理不同的术语库及其信息提供便利。

6.2.2 术语库应用软件

a) 按词检索的功能(例如:上下文中的词检索、上下位词检索、相关词的检索、同义词的检索、多义词的检索、反义词检索等);

b) 能提供布尔逻辑运算、关系运算、全文检索等多种检索操作,各种检索符可进行有意义的自由组合;

c) 多层次、多括号的结构型术语数据应能进行逻辑组合和运算;

d) 浏览功能,可以多方位浏览并同时检索;

e) 排序功能,具有按字母排序、汉语拼音排序、汉字笔划排序、汉字部首排序和分类排序等多种排序功能;

f) 灵活方便的辅助编辑功能,可打印、排版、印刷,可制作软盘或光盘等电子出版物;

g) 统计和运算功能;

h) 可处理文本、图形、图像、声音等多媒体信息;

i) 具有友好的人机接口,可随时调用、检索和交换各种术语数据。

6.3 术语库数据

6.3.1 术语数据的数量和范围

术语库的数据量满足用户需求,能为用户提供必要、充分且分布均衡的信息。

术语数据的范围考虑以下方面:

a) 综合性术语库中术语的专业领域覆盖面合理,分布均衡,并注意跨学科的术语的处理;

b) 某专业领域的术语库中收录的本领域的术语数据全面和系统,并适量收录常用的相关领域的术语数据;

c) 多语种术语库中各语种的对应术语应齐备;

d) 可采用语料库技术进行术语数据的覆盖实验,判断其充分性和完备性。

6.3.2 术语数据的质量

术语库所收录的术语及其信息的权威性、可靠性、正确性和一致性是评价术语库的重要因素与内容,应符合 GB/T 10112、GB/T 13725—2001 的第 6.3 条和 6.4 条的规定。

6.3.3 数据类目的选择

数据类目满足术语库用户的要求和数据交换的易行性，符合 GB/T 13725—2001 中第 6.3.2 条的规定。

6.3.4 数据的采集

正确和规范地采集术语及其有关数据，并符合 GB/T 15387.2—2001 第 6 章的规定。

6.3.5 数据维护

a) 管理程序与数据相互独立，具有数据的增加、删除、修改等功能；

b) 用户访问权限有级别限制；

c) 具有迅速恢复数据和应急切换机制。

6.3.6 数据处理技术的先进性

考虑到人工智能和超文本等数据处理方法与技术的应用。

6.3.7 数据的交换与传递

术语库能方便、准确和快速地通过数据载体或网络实现数据的交换与传递。

6.4 用户界面

术语库面向用户，具有功能强、性能好、灵活、方便的人机界面，最大限度地满足用户的各项需求。

a) 根据用户的要求，采用键盘、文字识别和语音识别等适宜的输入方法；

b) 多窗口环境；

c) 界面清晰明了，操作简单，易于学习、掌握和使用。

6.5 开发过程与文件

术语库开发过程符合 GB/T 13725—2001 第 7 章的规定。术语库开发文件完整、正确、统一、清晰，并符合 GB/T 15387.1—2001 的规定。

7 评价的实施

7.1 评价的提出

a) 项目结束后和需申报成果时；

b) 用户需要购买或使用某个术语库时；

c) 开发单位要了解自己所设计术语库的实际水平时；

d) 管理部门或经销单位希望比较和鉴定若干术语库的优劣时；

e) 其他需要。

7.2 评价的程序

7.2.1 组织评价机构

该机构应由与术语库有关的各方面专家和用户组成。

7.2.2 拟定评价计划

a) 规定评价的目的、要求和范围；

b) 规定评价的具体准则、方法和工具；

c) 制定工作日程表；

d) 明确参与评价工作人员的分工；

e) 规定评价的管理和控制程序。

7.2.3 建立评价细则

根据技术评价的一般原则与方法，确定评价依据、办法和所遵循的标准、规范和规则等文件。

7.2.4 确定量化形式

为了使评价具体、明确，可以采用以下方法量化评价内容和指标；

a) 提问表型，如将指标列表，回答“是”或“非”；

b) 分级加权型，如将指标分成几级，规定权值；

c） 定量值型，如为指标设计定量计算公式，代入评测数据计算等。

7.2.5 测试系统

按照技术评价的类别，采用一定的量化形式，对术语库的主要技术内容与项目进行测试。

7.2.6 收集和分析数据

测试并收集所评价系统的各类数据，按照评价计划和细则的有关规定，研究、分析这些数据。

7.2.7 完成评价报告

根据评价细则和对数据的分析，评价系统，完成评价报告。

需要全本请访问 www.ertongbook.com

附 录 A
（规范性附录）
汉语语言文字规范文件

以下列出最常用的汉语语言文字规范文件目录，本标准中 4.6 规定的内容应与这些规范文件一致：

《中华人民共和国国家通用语言文字法》(2001 年 1 月 1 日实施)

GB/T 15834—1995 标点符号用法

GB/T 16159—1996 汉语拼音正词法基本规则

《简化字总表》 国家语言文字工作委员会 1986 年发布

《现代汉语常用字表》 国家语言文字工作委员会和国家教育委员会 1988 年联合发布

《现代汉语通用字表》 国家语言文字工作委员会和中华人民共和国新闻出版署 1988 年联合发布

《第一批异体字整理表》 中华人民共和国文化部和中国文字改革委员会 1955 年联合发布

《汉语拼音方案》 中华人民共和国第一届全国人民代表大会第五次会议 1958 年通过

《普通话异读词审音表》 国家语言文字工作委员会、国家教育委员会和广播电视部 1985 年联合发布

《中国人名汉语拼音字母拼写法》 中国文字改革委员会 1976 年修订

《中国地名汉语拼音字母拼写规则(汉语地名部分)》 中国地名委员会、中国文字改革委员会和国家测绘局 1984 年联合发布

《GB 13000.1 字符集汉字笔顺规范》 国家语言文字工作委员会

中 华 人 民 共 和 国
国 家 标 准
术语数据库技术评价指南
GB/T 15625—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 3/4 字数 18 千字
2002年4月第一版 2002年4月第一次印刷
印数 1—1 000
*
书号:155066·1-18329 定价 10.00 元
网址 www.bzcbs.com

版权专有 侵权必究
举报电话:(010)68533533

GB/T 15625-2001