ICS 03.080.99;35.240.30
CCS A 10

中华人民共和国国家标准

GB/T 40035—2021

双语平行语料加工服务基本要求

Basic requirements for bilingual parallel corpus processing service

2021-04-30 发布　　2021-11-01 实施

国家市场监督管理总局
国家标准化管理委员会　发布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

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专利的责任。

本文件由全国语言与术语标准化技术委员会(SAC/TC 62)提出并归口。

本文件起草单位：中国标准化研究院、中国翻译协会、上海一者信息科技有限公司、上海佑译信息科技有限公司、中译语通科技股份有限公司、北京悦尔信息技术有限公司、苏州联跃科技有限公司、四川语言桥信息技术有限公司、北京百度网讯科技有限公司、沈阳雅译网络技术有限公司、上海智膳合网络科技有限公司、北京语言大学、北京邮电大学。

本文件主要起草人：刘智洋、张井、叶剑、柴瑛、黄宝荣、罗慧芳、蒙永业、朱励、张雪涛、王海涛、朱宪超、韩林涛、郑春萍、何中军、于立梅、张春良、甘克勤、张宝林。

双语平行语料加工服务基本要求

1 范围

本文件规定了双语平行语料加工服务的基本要求、加工流程、服务内容和数据安全等内容。

本文件适用于以原文和译文为对象的、以文字为表达形式的数字化双语语料加工服务，其他数字化文本的语料加工也可参照使用，也适用于对语料对齐工具的评价。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

文本 text

以字符、符号、词、短语、段落、句子、表格或其他字符排列形成的数据，用于表达意义，其解释基本上取决于读者对于某种自然语言或人工语言的知识。

[来源：GB/T 4894—2009，4.1.1.2.4]

3.2

语料 corpus

语言材料或资料。

3.3

双语平行语料 bilingual parallel corpus

由两种语言构成，并在篇章、段落、句子或其他级别平行对齐的**语料**(3.2)。

3.4

原文 source language text

源语言**文本**(3.1)。

[来源：GB/T 19363.1—2008，3.4，有修改]

3.5

译文 target language text

目标语言**文本**(3.1)。

[来源：GB/T 19363.1—2008，3.5，有修改]

3.6

客户 client

接受按其要求提供产品或服务的个人或组织。

[来源：GB/T 19000—2016，3.2.4，有修改]

3.7

元数据 metadata

关于数据的内容、质量、状况和其他特性的描述性数据。

3.8

服务提供方　service provider

提供服务的个人或组织。

3.9

光学字符识别　optical character recognition;OCR

自动识别通过扫描仪、数码相机、摄像机等得到的图像中的字符，便于存储、编辑和检索。

[来源:GB/T 31219.2—2014,3.4]

3.10

TMX　Translation Memory eXchange

翻译记忆交换的标准格式。

3.11

语料对齐　corpus alignment

将双语**语料**(3.2)进行篇章、段落、句子或其他级别的对齐，构成平行对照的形式。

3.12

语料对齐工具　corpus alignment tool

用于将双语文本对齐，并能制作成**双语平行语料**(3.3)的工具。

3.13

纠正　correction

为消除已发现的不合格内容所采取的措施。

[来源:GB/T 19000—2016,3.12.3]

3.14

脱敏　de-identification

去除可确认个人或组织身份的数据与数据主体之间联系的过程。

[来源:ISO/TS 25237:2008,3.18]

3.15

敏感信息　sensitive information

如果公开或者滥用会造成潜在危害的信息。

[来源:GB/T 4894—2009,4.7.3.2.4,有修改]

3.16

匿名化数据　anonymized data

去除直接涉及数据主体的个人或组织数据。

[来源:GB/T 4894—2009,4.7.3.2.3,有修改]

4　总则

4.1　双语平行语料加工服务是将客户提供的原文和译文的文本内容按段落、句子或其他级别建立对应关系的一种服务。

4.2　双语平行语料加工服务的目的是获取双语对齐的文本资料，为计算机辅助翻译、机器翻译和语言学研究提供基础数据。

4.3　双语平行语料加工的对象包括原文、译文和加工文本的元数据。

4.4 双语平行语料加工服务提供方(以下简称“服务提供方”)对译文不进行审核,译文质量由客户保证。

4.5 双语平行语料加工服务可以采用多个工具完成,也可以在一个集成环境中完成。该环境应集成对齐、元数据采集等功能,以适应双语平行语料加工服务的需要。

5 基本要求

5.1 服务提供方

服务提供方应具备以下条件:

a) 建立完备的语料加工流程体系,包括但不限于数据预处理、语料对齐、项目管理、质量审核等;
b) 配备合格的语料加工人员;
c) 配备稳定可用的语料对齐工具及相关文字处理工具;
d) 配备可完成语料加工服务的场所。

5.2 语料加工人员

服务提供方应确保双语平行语料加工人员具备以下能力:

a) 阅读源语言和目标语言的能力:能理解源语言和目标语言,并能快速阅读原文和译文;
b) 研究和处理文本的能力:能拓展必要的文本处理及专业知识,并能制定策略来有效利用现有资源;
c) 技术能力:利用技术资源,包括使用工具和信息系统支撑整个语料加工过程,完成其中的各项技术任务。

注:双语平行语料加工人员的培训见附录A。

5.3 服务环境

服务提供方的服务环境应拥有完成双语语料加工所需的技术设备和办公设备,如光学识别工具、对齐工具等。客户可与服务提供方约定加工时使用的工具名称和版本。

服务提供方的保密环境及级别应符合客户对语料保密的要求,按客户的要求配备保密设备、进行安全加固、为语料加工人员开展保密培训等。

5.4 加工内容

双语语料应由客户提供,语料可来自正式出版物、公司内部资料、网站等。

双语语料的加工应优先选择数字化后的双语语料,尚未数字化的双语语料,可通过扫描或拍照等手段,后采用光学字符识别的方式转换成数字化形式,或直接通过键盘录入。

通过光学字符识别或键盘录入的双语语料应增加校对环节保证内容的质量。

5.5 加工结果

5.5.1 完整性

在符合客户数据处理要求的前提下,服务提供方的加工结果应保证原文、译文及元数据的完整性,确保加工结果无信息丢失。

注:双语语料加工的元数据见附录B。

5.5.2 准确性

在符合客户数据处理要求的前提下,服务提供方的加工结果应保证原文和译文对应关系的准确性

以及元数据的准确性，确保加工结果准确无误。

注：双语语料加工的元数据见附录B。

5.5.3 可用性

服务提供方应保证加工结果符合以下要求：

a) 能被语料检索、管理和生产工具解析；

b) 无乱码、多余标签等不可用信息；

c) 无格式混乱或原文译文不对应情况；

d) 无用户未要求的多余信息。

5.5.4 规范性

服务提供方的加工结果应符合客户的规范要求，加工结果的数据格式应包括TMX、TXT等，并符合以下要求：

a) TMX文件应符合翻译记忆库交换规范，包含留存版本号、编码格式、制作语料的工具名称、制作时间、双语语言编码等元数据信息；

b) TXT文件应采用一种常见的大字符集的编码格式，如UTF-8。

注：TXT文件常见编码格式见附录C，TMX格式规范见附录D。

5.6 语料加工工具

语料对齐是双语平行语料加工的关键环节，因此语料对齐工具作为语料加工工具的重要组成部分，应满足以下可靠性、易用性和兼容性三方面要求。

5.6.1 可靠性

语料对齐工具应在出现局部功能故障时，不影响其他功能的操作，仍能提供对齐功能。

语料对齐工具应提供对齐过程数据自动保存及恢复功能。

5.6.2 易用性

5.6.2.1 本地化界面

语料对齐工具应支持中文界面。

5.6.2.2 操作功能

语料对齐工具应支持对齐双语文本所需的操作功能：

a) 文字编辑：在允许文字输入的内容标识区域，支持文字修改、删除和添加等；

b) 合并：支持将分布在两行的文本合并成一行；

c) 拆分：支持将一行文本切分成两行；

d) 上移：支持将文本位置向上移动；

e) 下移：支持将文本位置向下移动；

f) 插入：支持在某一行文本上方或下方插入一行；

g) 删除：支持删除某行或多行文本；

h) 回退：支持回退至上一步操作，没有上一步时，停留在当前操作；

i) 对齐：支持文本调整完成后，执行段落或句子级别的对齐；

j) 导出：支持对齐完成后，导出对齐的双语文本；

k) 保存：支持将对齐过程中的文本进行保存。

5.6.2.3 帮助系统

语料对齐工具应提供：

a) 系统功能离线帮助文档或在线帮助支持，并与工具的功能保持一致，使用户在使用系统过程中遇到问题时能够快速获得相应的帮助；
b) 基本操作指引，使用户在使用系统过程中能够快速了解操作技巧；
c) 友好交互提示，能够帮助用户找到错误定位，提示错误原因。

5.6.2.4 效率

应从以下方面评价语料对齐工具的效率：

a) 响应时间：
   1) 工具启动时间；
   2) 自动对齐、拆分、合并、保存等基本操作的响应时间；
   3) 恢复作业时间：关闭后再次打开工具时，能快速定位上次作业位置的时间。
b) 便捷度：
   1) 支持快捷键操作；
   2) 支持右键菜单。

5.6.3 兼容性

语料对齐工具的兼容性要求如下：

a) 服务器端的语料对齐工具应说明能够支持的浏览器，并避免使用基于特定浏览器和特定操作系统功能的脚本和插件；
b) 服务器端的语料对齐工具应适应不同浏览器和分辨率的展示，应提供至少一种推荐的浏览器和分辨率，确保在该浏览器和分辨率下展示的网页布局和元素完整正确；
c) 本地的语料对齐工具应提供完整的安装文档，说明支持的操作系统、应用的配置信息和常见问题的提示信息等内容。

6 加工流程

6.1 预处理

6.1.1 语料准备

对于图片格式或扫描版的尚未数字化的语料，需先通过光学字符识别或直接通过键盘录入转成可编辑的电子文本语料。

6.1.2 清洗

对语料中的乱码及特殊字符等进行排查和纠正。

6.1.3 去重

对语料进行数据查重操作，检查已有的双语语料数据和元数据，尽量利用客户已有的数据，避免重复加工。

6.1.4 脱敏

按客户的脱敏要求对数据进行脱敏处理，去除语料中的身份信息和其他敏感信息，把语料转换成匿

名化数据。

6.2 语料对齐

语料加工人员利用语料对齐工具导入双语文档后，工具执行自动断句，结合工具自动对齐与人工手动调整对齐后，导出最终的双语平行语料库，导出时应确认源语言和目标语言、语料库名称以及语料库格式等信息。

注：文件命名规则、编码格式及文件格式见附录E。

6.3 语料审核

服务提供方应对加工结果进行抽样检查，抽样数不少于结果总条目数的10%，抽样数据的准确率不低于99%。

服务提供方应按照客户提供的规范，参照客户提供的示例对加工结果进行检查，确保加工结果符合客户的要求，检查结果应予以记录并归档。

7 服务内容

7.1 需求沟通

服务提供方应与客户建立完善的需求沟通机制，在接受客户的双语平行语料加工任务前，与客户沟通明确原文文本和译文文本的加工级别、采集元数据的范围、脱敏要求等及其可行性，因为双语平行语料加工服务的效率受加工级别、原文文本和译文文本是否已数字化、元数据是否容易采集以及脱敏程度等因素的影响较大。对于尚未数字化的语料，服务提供方应与客户就语料数字化的加工方式（光学字符识别或直接通过键盘录入）达成一致。

按照客户对语料的用途，双语平行语料加工可分为以下两种级别。

a) 标准级。对原文和译文执行段落或句子级别的对齐，采集基本的元数据。

b) 精标注级。按照客户的要求，除语料对齐和采集元数据之外，对语料进行分词、词性标注、句法标注、语义标注等。

7.2 客户协议

服务提供方应与客户达成协议，并予以记录。如果通过口头或电话达成协议，服务提供方应以书面形式（如信函、传真或电子邮件等）确认该协议及其条款。

客户与服务提供方应就语料加工级别（段落级、句子级等）达成一致。如果以句子为基本单位，双方应就句子拆分的断句规则、原文和译文无法对应的处理规则等情况达成一致。客户应将相关规范（如断句规则、用途等）连同样例发给服务提供方，并由其遵照执行。

客户与服务提供方可对语料的知识产权归属及数据保密等要求进行协商约定。

在协议执行过程中，如果出现了与协议不符的情况，各方应达成一致，对协议进行修订并予以记录和归档。

7.3 项目管理

服务提供方应安排项目经理对语料加工项目执行任务分配、进度管理、质量检查等工作。

7.4 加工环节

双语平行语料加工环节包括双语语料预处理、双语语料对齐和双语平行语料审核。

7.5 交付内容

双语平行语料加工的交付内容应包括双语平行语料和加工报告，交付要求如下：

a) 双语平行语料应通过移动存储介质或云存储形式交付，且应包含加工服务提供方名称、交付日期、语料总条目数等元数据信息；

b) 加工报告应通过移动存储介质或云存储形式交付，内容应包含客户提供语料概况、加工流程说明、实际加工交付语料条目数、未能加工语料说明、交付语料准确率说明、加工使用工具、服务完成所用时间等信息。

7.6 质量保证期

服务提供方应与客户约定质量保证期，未约定的应以一年为最短质量保证期。质量保证期内，服务提供方需修复客户提出的语料加工问题。

7.7 服务评价与改进

服务提供方应指定专人跟踪客户的反馈意见并进行记录和整理，采取相应的改进措施，优化语料加工流程。

对于分批交付的双语语料，服务提供方应在每批数据加工结果交付后安排专门客服人员进行质量跟踪，询问客户的反馈，采取相应的改进措施。

8 数据安全

8.1 数据备份

双语平行语料加工的各环节中，要确保双语语料的安全、有序，并及时做好数据多重备份。

8.2 文档管理与日志

双语语料的整个加工过程应记录操作日志，及时撰写和汇总加工过程中的技术和管理文档。

8.3 数据存储

服务提供方应对客户的需求文件、加工过程文件及最终交付文件按客户及日期分类存储，以便语料信息查询及客户跟踪。

附 录 A
（资料性）
双语平行语料加工人员的培训

A.1 对双语平行语料加工人员进行语料加工所需知识和技能的培训可以：

a） 为双语平行语料加工人员提供语料加工所需的技能；

b） 有助于满足逐渐增长的语料加工需求，提高效率；

c） 推动双语平行语料加工技术的发展和创新。

A.2 双语平行语料加工人员的培训可包括：

a） 高级文本处理技巧，使用脚本处理双语文本；

b） 脱敏和语料清洗（包括去除语料中的乱码，格式标记等）技术，以便能够更好地处理双语语料对齐的场景；

c） 使用质量工具在项目结束时执行质量检查，如检查格式的合法性等。

附 录 B
（资料性）
双语语料加工的元数据

双语平行语料加工过程中应记录与双语语料相关联的元数据，将其作为加工结果的一部分。元数据内容包括但不限于表 B.1 所列，每一元素均可选，且可重复。

表 B.1 元数据

内容名称	标签	定义	注释
identifier	标识符	双语语料的唯一识别符	一般是特定应用系统内具有唯一识别性的标识符号。可由标识应用系统的前缀（即标识符的类型）与一字符串（即标识符的值）组成，可由系统产生或由人工赋予
sourceLanguage	源语言	双语语料中的源语言	一般采用 ISO 639 中规定的语言代码和 ISO 3166 中规定的国家代码的组合标识双语语料中的源语言。例如，使用 ZH-CN 代表简体中文
targetLanguage	目标语言	双语语料中的目标语言	一般采用 ISO 639 中规定的语言代码和 ISO 3166 中规定的国家代码的组合标识双语语料中的目标语言
sourceTitle	原文标题	原文的标题	一般指原文公开的标题
targetTitle	译文标题	译文的标题	一般指译文公开的标题
sourceOfSourceText	原文来源	原文文本资源的来源	一般是用于确定提供所加工的原文文本的单位或个人
sourceOfTargetText	译文来源	译文文本资源的来源	一般是用于确定提供所加工的译文文本的单位或个人
sourcePublicationDate	原文出版日期	原文文本资源的发布日期	通常采用××××年××月××日的格式，如 2018 年 9 月 1 日
targetPublicationDate	译文出版日期	译文文本资源的发布日期	通常采用××××年××月××日的格式，如 2018 年 9 月 1 日
author	作者	原文的作者	编写原文文本的单位或个人
translator	译者	译文的作者	翻译原文文本的单位或个人
subjectField	领域	双语语料的领域	一般采用关键词或分类号来描述，建议使用受控词表
register	语域	双语语料的语域	一般采用关键词或分类号来描述，建议使用受控词表
format	格式	文本资源的数字表现形式	文本资源的内容形式，包括资源内容和元数据的类型

表 B.1　元数据（续）

内容名称	标签	定义	注释
collectionMode	采集方式	双语资源的采集方式	如：OCR，人工录入
OCRTool	识别工具	识别文本资源的软件	识别工具的名称
alignMode	对齐方式	双语资源的对齐方式	如：人工对齐，自动对齐或自动＋人工对齐
alignmentTool	对齐工具	对齐时使用的软件	对齐软件的名称
characterSet	字符集	双语平行语料加工结果采用的字符集的名称	通常采用的编码为 UTF-8，UTF-16，GB 18030 等
date	日期	完成双语平行语料加工的日期	通常采用××××年××月××日的格式，如 2018 年 9 月 1 日

附 录 C
（资料性）
TXT 文件常见编码格式

表 C.1 列出了 TXT 文件常见的编码格式。

表 C.1 TXT 文件常见编码格式

编码格式	说 明
ASCII	ASCII(American Standard Code for Information Interchange,美国信息交换标准代码)是一种标准的单字节字符编码方案,用于基于文本的数据。ASCII 由美国国家标准学会(American National Standard Institute,ANSI)制定,后被 ISO 646 标准所采用
ISO 8859-1	ISO 8859-1 是单字节编码,向下兼容 ASCII。ISO 8859-1 编码有时称 Latin-1
Unicode	Unicode(中文:万国码、国际码、统一码、单一码)是计算机科学领域里的一项业界标准。Unicode 是为了解决传统的字符编码方案的局限而产生,为世界上大多数的常用语言中的字符设定了统一并且唯一的二进制编码,以满足跨语言、跨平台进行文本转换、处理的要求
UTF-8	UTF-8 编码是一种针对 Unicode 的可变长度字符编码,也是一种前缀码。每一个字符的长度从 1～6 个字节不等。UTF-8 编码一个很重要的特性就是兼容 ISO 8859-1 编码
GBK/GB 2312	GBK/GB 2312 是汉字的国标码,专门用来表示汉字,是双字节编码

附 录 D
（资料性）
TMX 格式规范

表 D.1 和表 D.2 为 TMX 文件类型所需的元素和文件属性定义。

表 D.1 TMX 文件元素表

元素	说 明
〈TMX〉	〈TMX〉元素包含一个〈HEADER〉元素，后跟一个〈BODY〉元素。 〈TMX〉元素有一个必需属性：VERSION
〈HEADER〉	〈HEADER〉元素包含零个、一个或多个〈META/〉元素；零个、一个或多个〈NOTE〉元素；零个、一个或多个〈UDE〉元素；零个、一个或多个〈PROP〉元素。 〈HEADER〉元素有四个必需属性：CREATIONTOOL，SEGTYPE，O-TMF 和 DATATYPE。有七个可选属性：O-ENCODING，CREATIONDATE，CREATIONID，CHANGEDATE，CHANGEID，ADMINLANG 和 SRCLANG
〈PROP〉	〈PROP〉(Property)元素不包含其他元素。〈PROP〉元素有一个必需属性：NAME 和两个可选属性：LANG 和 O-ENCODING。〈PROP〉元素用于定义父元素(或在〈HEADER〉元素中使用〈PROP〉时的文件)的各种属性
〈UDE〉	〈UDE〉(用户定义的编码)元素包含一个或多个〈MAP〉元素。 〈UDE〉元素有一个必需属性：NAME。 用于指定一组用户定义的字符和/或从 Unicode 到用户定义的编码的映射
〈MAP/〉	〈MAP/〉元素为空(没有内容且没有结束标记)。〈MAP/〉元素有一个必需属性：UNICODE 和三个可选属性：CODE，ENT 和 SUBST，用于指定用户定义的字符及其某些属性
〈BODY〉	〈BODY〉元素包含主数据，即组成文件的〈TU〉集。无属性
〈TU〉	每个〈TU〉(翻译单元)元素包含零个、一个或多个〈NOTE〉元素，后跟零个、一个或多个〈PROP〉元素，后跟一个或多个〈TUV〉元素
〈TUV〉	每个〈TUV〉(翻译单元变体)指定给定语言的文本。包含零个、一个或多个〈NOTE〉元素，后跟零个、一个或多个〈PROP〉元素，后跟一个〈SEG〉元素。〈TUV〉有一个必需属性：LANG 和九个可选属性：O-ENCODING，DATATYPE，USAGECOUNT，LASTUSAGEDATE，CREATIONTOOL，CREATIONDATE，CREATIONID，CHANGEDATE 和 CHANGEID
〈SEG〉	每个〈SEG〉(段)包含〈TUV〉的文本。 无属性
〈NOTE〉	〈NOTE〉元素用于注释，不包含其他元素。〈NOTE〉元素有两个可选属性：O-ENCODING 和 LANG

表 D.2 TMX 文件属性定义

属性	定 义
CREATIONTOOL	CREATIONTOOL 属性标识创建 TMX 文档的工具
CREATIONDATE	CREATIONDATE 属性指定元素创建的日期

表 D.2　TMX 文件属性定义（续）

属性	定　义
CREATIONID	CREATIONID 属性指定创建元素的用户
CHANGEDATE	CHANGEDATE 属性指定元素修改的日期
CHANGEID	CHANGEID 属性指定修改元素的用户
O-ENCODING	O-ENCODING 属性指定元素数据的原始或首选代码集，以防在非 Unicode 代码集中进行编码
O-TMF	O-TMF(原始翻译记忆库格式)元素指定从中生成 TMX 文档的翻译记忆库文件的格式
LANG	LANG 属性指定元素数据的语言或区域设置
DATATYPE	DATATYPE 属性指定元素的数据类型
SRCLANG	SRCLANG 属性指定源语言文本的语言或区域设置
ADMINLANG	ADMINLANG 属性在〈HEADER〉元素中用于指定管理和信息元素〈NOTE〉,〈META/〉和〈PROP〉的默认语言
NAME	NAME 属性指定〈META/〉或〈PROP〉元素的信息类型，或〈UDE〉元素的名称
REF	REF 属性用于指定〈META/〉元素的外部参考文档
ID	ID 属性指定〈TU〉元素的标识符
USAGECOUNT	USAGECOUNT 属性指定〈TU〉或〈TUV〉的使用次数
LASTUSAGEDATE	LASTUSAGEDATE 属性指定上次使用〈TU〉或〈TUV〉的时间
VERSION	VERSION 属性指示文档的 TMX 格式版本
UNICODE	UNICODE 属性指定〈MAP/〉元素的 Unicode 字符值
CODE	CODE 属性指定用户定义编码中的代码点值，该编码对应于给定〈MAP/〉元素的 UNICODE 字符
ENT	ENT 属性指定由给定〈MAP/〉元素定义的字符的实体名称
SUBST	SUBST 属性允许为给定〈MAP/〉元素中定义的字符指定备用字符串
SEGTYPE	SEGTYPE 属性指定〈TU〉元素中使用的分段类型

附 录 E
（资料性）
文件的命名规则、编码格式及文件格式

E.1 双语平行语料加工结果的文件命名宜遵循以下规则。

a) 拥有唯一标识符，不能与其他文件标识符重复。

b) 文件名定义应清晰明确，以便于文件名的标准化与统一管理。

c) 具备长期可用性。文件命名方式不依赖于某种处理或者系统。文件名包含的信息不应随时间的推移而改变。

d) 严格遵守技术限制。符合计算机系统对文件名中特殊字符、空格、日期等字符使用的限制，以及文件名字符长度的限制。

e) 文件扩展名的字母用小写形式。

E.2 双语平行语料加工结果的文件编码应避免采用非标准的专有字体和格式。双语平行语料加工结果的文件字体宜转换为标准字体，编码宜采用下列格式中的一种：

a) GB 2312；

b) GB 18030；

c) GB 13000；

d) UTF-8；

e) UTF-16。

参 考 文 献

[1] GB/T 4894 信息与文献 术语

[2] GB/T 13000 信息技术 通用多八位编码字符集(UCS)

[3] GB 18030 信息技术 中文编码字符集

[4] GB/T 19000 质量管理体系 基础和术语

[5] GB/T 19363.1—2008 翻译服务规范 第1部分:笔译

[6] GB/T 25000.51 系统与软件工程 系统与软件质量要求和评价(SQuaRE) 第51部分:就绪可用软件产品(RUSP)的质量要求和测试细则

[7] GB/T 31219.2—2014 图书馆馆藏资源数字化加工规范 第2部分:文本资源

[8] ISO 639 Code for the representation of names of languages

[9] ISO/IEC 646 Information technology—ISO 7-bit coded character set for information interchange

[10] ISO 3166 Codes for the representation of names of countries and their subdivisions

[11] ISO 8601:2004 Data elements and interchange formats—Information interchange—Representation of dates and times

[12] ISO/TS 25237:2008 Health informatics—Pseudonymization