

ICS 03.080.01
CCS A 10



中国语言服务行业规范

Specifications for the Language Service Industry in China

ZYF 00x.2-2021

中国特色话语翻译 高端语料库建设 第2 部分：系统架构

Translations of Chinese key terms and expressions — Corpus
construction — Part 2: system architecture

20**-**-**发布

20**-**-**实施

中国翻译协会 发布

目次

前 言	III
引 言	IV
1 范围	1
2 规范性引用文件.....	1
3 术语与定义.....	1
4 缩略语.....	2
5 基本建设原则.....	2
6 系统架构.....	2
7 应用服务.....	3
7.1 信息检索.....	3
7.2 统一门户.....	3
7.3 用户登录.....	3
7.4 结果展示.....	3
8 知识分析.....	4
8.1 知识库关联.....	4
8.2 查重分析.....	4
8.3 报表生成.....	4
9 知识库管理.....	4
9.1 语料库管理.....	4
9.2 术语库管理.....	5
9.3 元数据管理.....	6
10 加工处理.....	7
10.1 翻译管理.....	7
10.2 对齐标引.....	9
11 系统管理.....	10
11.1 用户管理.....	10
11.2 角色管理.....	10
11.3 审批管理.....	10
11.4 权限管理.....	10
11.5 其他管理.....	10
12 安全管理.....	11
12.1 用户安全管理.....	11
12.2 加解密功能.....	11
12.3 泄密溯源追踪.....	11
12.4 操作行为管控.....	11
12.5 预览安全防护.....	11
13 存储管理.....	11

13.1 存储机制选择.....	11
13.2 存储空间管理.....	11
13.3 存储方式管理.....	11

学位论文答辩稿

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。

本文件是 ZYF 00X《中国特色话语翻译 高端语料库建设》的第2部分。ZYF 00X 已经发布了以下部分：

——第1部分：基本要求

——第3部分：抽样检验

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。。。提出。

本文件由。。。归口。

本文件起草单位：

本文件主要起草人：

文件待审稿

引 言

讲好中国故事，传播好中国声音，展示真实、立体、全面的中国，是加强我国国际传播能力建设的重要任务。加强国际传播能力建设，有助于形成同我国综合国力和国际地位相匹配的国际话语权，为我国改革发展稳定营造有利外部舆论环境，有助于推动构建人类命运共同体。

利用信息化和语料库技术构建中国特色话语翻译语料库，有助于中国特色话语的阐释与传播，进一步拓展中国特色话语传播途径、拓宽传播方式、创新传播机制，增进国际社会对中国的理解和认同，争取更多国际话语权，提升中国软实力。中国特色话语翻译高端语料库建设系列标准旨在通过规范语料库建设过程、系统平台、质量控制方法等，提高语料库建设效率和质量，拟由三个部分组成：

——第1部分：基本要求。目的在于给出语料库建设基本原则、基本流程、步骤、管理要求。

——第2部分：系统架构。目的在于明确语料库建设系统的主要架构和功能。

——第3部分：抽样检验。目的在于给出对话料库质量合格情况进行检验评估的程序、流程、判定规则等。

中国特色话语翻译 高端语料库建设 第2部分：系统架构

1 范围

本文件规定了中国特色话语翻译高端语料库（以下简称“语料库”）建设管理系统的基本建设原则、系统架构，规定了语料库管理系统的应用服务、知识分析、知识库管理、加工处理、系统管理、安全管理、存储管理等功能。

本文件适用于语料库建设管理系统的设计、开发、维护、使用等。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

- GB/T 32905-2016 信息安全技术 SM3 密码杂凑算法
- GB/T 32907-2016 信息安全技术 SM4 分组密码算法
- GB/T 32918.1-2016 信息安全技术 SM2 椭圆曲线公钥密码算法 第1部分：总则
- GB/T 32918.2-2016 信息安全技术 SM2 椭圆曲线公钥密码算法 第2部分：数字签名算法
- GB/T 32918.3-2016 信息安全技术 SM2 椭圆曲线公钥密码算法 第3部分：密钥交换协议
- GB/T 32918.4-2016 信息安全技术 SM2 椭圆曲线公钥密码算法 第4部分：公钥加密算法
- GB/T 32918.5-2017 信息安全技术 SM2 椭圆曲线公钥密码算法 第5部分：参数定义
- GB/T 33133.1-2016 信息安全技术 祖冲之序列密码算法 第1部分：算法描述
- GB/T 35276-2017 信息安全技术 SM2 密码算法使用规范
- GM/T 0044-2016 SM9 标识密码算法
- ZYF 00x.1-2021 中国特色话语翻译 高端语料库建设 第1部分：基本要求

3 术语与定义

ZYF 00x.1-2021 界定的以及下列术语和定义适用于本文件。

3.1

术语 term

在特定专业领域中一般概念的词语指称。

[来源：GB/T 15237.1-2000, 3.4.3]

3.2

术语库 term base

术语数据库 terminological data base

含有术语数据的数据库集。

[来源：GB/T 17532-2005, 7.8, 有修改]

3.3

元数据 metadata

定义和描述其他数据的数据。

[来源：GB/T 18391.1-2009, 3.2.16]

4 缩略语

AES: 高级加密标准 (Advanced Encryption Standard)

CSV: 逗号分隔值 (Comma-Separated Values)

IM: 即时通讯 (Instant Messaging)

PDF: 便携式文档格式 (Portable Document Format)

QA: 智能问答 (Question Answering)

SQL: 结构化查询语言 (Structured Query Language)

TMX: 翻译记忆交换 (Translation Memory eXchange)

XML: 可扩展标记语言 (eXtensible Markup Language)

5 基本建设原则

应对各方面的用户需求进行调查分析, 满足大多数用户对语料库功能、性能、数据等方面的要求, 在充分考虑社会效益和经济效益的前提下开发系统, 并遵循以下基本建设原则:

- a) 易用性: 系统应简单易学, 使用方便;
- b) 经济性: 应选择技术上先进、经济上合理的设计方案;
- c) 可靠性: 硬件配置、软件架构、功能逻辑应稳定可靠;
- d) 兼容性: 新版本系统应兼容老版本的文件和数据格式;
- e) 易维护性: 便于维护和完善性能, 使系统保持良好工作状态;
- f) 安全性: 应确保各类用户能在允许权限内执行操作, 防止数据泄露、丢失、损坏, 避免计算机病毒侵入, 并具备安全防护的有效措施;
- g) 易扩展压缩性: 应能根据需求的变化, 增加、减少、优化系统功能。

6 系统架构

中国特色话语翻译高端语料库建设管理系统的架构如图1所示。

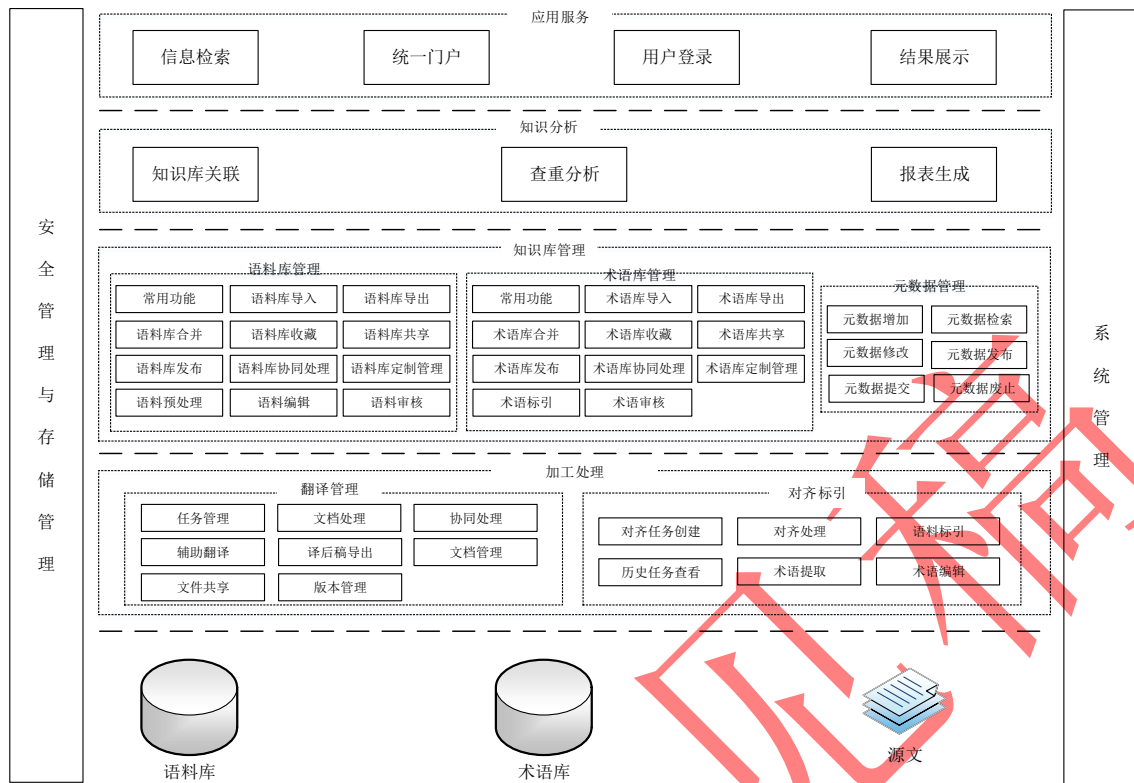


图 1 中国特色话语翻译高端语料库建设管理系统架构

7 应用服务

7.1 信息检索

信息检索包含但不限于以下功能：

- 统一检索：使用同一界面进行多语种检索，并展示检索结果；
- 模糊检索：针对任意内容，系统自动识别关键词的语言种类，并对检索结果进行展示；
- 限定检索：设置多种检索条件，检索条件包括语言种类、领域方向、作者、译者、审核人等。

7.2 统一门户

由统一入口进行登录，登录后根据用户角色，分别跳转到不同的界面。

7.3 用户登录

多种用户登录方式；不同用户账号之间相互切换；应能进行密码修改；密码失效后，显示重新登录窗口。

7.4 结果展示

分类展示检索结果。

8 知识分析

8.1 知识库关联

检索结果与原译文直接映射。

8.2 查重分析

对文件内容进行查重，分析重复率，生成展示页面。

8.3 报表生成

统计分析术语使用、语料使用、任务分配等情况。生成、下载统计分析报表。

9 知识库管理

9.1 语料库管理

9.1.1 常用功能

新建、删除、修改、查找语料库。

9.1.2 语料库导入

批量导入语料到语料库。

9.1.3 语料库导出

导出语料库中指定语料为各类型数据文件。

示例：将对齐语料导出 TMX、CSV 等文件。

9.1.4 语料库合并

合并多个语料库为一个语料库。

9.1.5 语料库收藏

标记语料、语料库，快速访问语料、语料库。

9.1.6 语料库共享

指定共享对象、共享范围、共享方式。

9.1.7 语料库发布

指定语料库的发布形式、发布范围、发布内容、发布格式等。

9.1.8 语料库协同处理

协同是多方相互配合、共同完成同一任务的一种工作方式。多人协同新建、删除、修改、查找、导入导出、收藏、共享发布语料库等。

9.1.9 语料库定制管理

创建多种类型的语料库。

示例：个人语料库、项目语料库。

9.1.10 语料预处理

对采集的语料进行语料清洗、语料查重。

9.1.11 语料编辑

以在线、离线形式对语料内容进行编辑、修改、批注等。

9.1.12 语料审核

多次审核指定语料，查看审核痕迹。

9.2 术语库管理

9.2.1 常用功能

新建、删除、修改、查找术语库。

9.2.2 术语库导入

批量导入术语到术语库。

9.2.3 术语库导出

导出术语库中指定术语为各类型数据文件。

9.2.4 术语库合并

多个术语库合并成一个术语库。

9.2.5 术语库收藏

标记术语、术语库，快速访问术语、术语库。

9.2.6 术语库共享

指定共享对象、共享范围、共享方式。

9.2.7 术语库发布

指定术语库的发布形式、发布范围、发布内容、发布格式等。

9.2.8 术语库协同处理

多人协同新建、删除、修改、查找、导入导出、收藏、共享发布术语库。

9.2.9 术语库定制管理

创建多种类型的术语库。

示例：个人术语库、项目术语库。

9.2.10 术语标引

标记术语来源、译者等术语属性。

9.2.11 术语审核

多次审核指定术语，查看审核痕迹。

9.3 元数据管理

9.3.1 元数据增加

当提交的增加元数据申请经过审批后，按以下步骤增加元数据：

- a) 进行元数据的正式注册，将元数据状态标注为“标准”状态，并赋予元数据完整属性；
- b) 对元数据备案；
- c) 更新元数据注册系统，并对外发布更新信息。

9.3.2 元数据修改

当提交的修改元数据申请经过审批后，按以下步骤修改元数据：

- a) 保留当前版本的元数据，对后续版本的元数据进行元数据的正式注册，除待更新的属性外，后续版本元数据的其他属性宜继承当前版本元数据的属性；
- b) 更新元数据注册系统，并对外发布更新信息。

9.3.3 元数据提交

按ZYP 00x.1-2021中7.6.2规定的元数据格式编写元数据注册表单，以纸质和电子文件的形式提出元数据增加、修改或废止的申请。

9.3.4 元数据发布

完成对元数据的维护工作后，标准的元数据由元数据管理部门发布。

9.3.5 元数据废止

当提交的废止元数据申请经过审批后，按以下步骤废止元数据：

- a) 对所要废止的元数据进行备份；
- b) 从注册系统中删除要废止的元数据；
- c) 更新元数据注册系统，并对外发布更新信息。

9.3.6 元数据检索

元数据检索基本要求如下：

- a) 支持对元数据的关键词检索，以及在结果中进行二次检索；
- b) 支持检索操作的“与”“或”“非”等各类逻辑运算；
- c) 支持对元数据的模糊检索；
- d) 支持对元数据的组合检索；
- e) 根据需要支持对元数据的批量检索；
- f) 支持对元数据进行SQL、XML Query等检索；
- g) 支持基于内容的检索；
- h) 提供外部系统检索接口；
- i) 支持检索结果的屏幕显示与用户直接下载。

10 加工处理

10.1 翻译管理

10.1.1 任务管理

任务管理包括以下功能：

- a) 任务派发：指定翻译者、审校者，指派多级译审任务；
- b) 进度监控：监控任务进度；
- c) 离线任务导入导出：导出离线翻译、审校任务，导入离线翻译、审校后译文，保留审校痕迹。

10.1.2 文档处理

文档处理包括以下功能：

- a) 预览文档：跨终端同步访问同一文档，同时预览多个文档，分页预览同一文档；
- b) 文档编辑：在线对文档进行编辑、修订、批注，自动消除文档中的修订、批注痕迹；
- c) 查找替换：在文档中检索相关内容，批量替换查找到内容；
- d) 文档标密：标写文档密级；
- e) 离线访问：包括离线权限控制和逾期访问自毁；
- f) 译后编辑：对预翻译后文档内容进行再编辑。

10.1.3 协同处理

10.1.3.1 多人预览

多人、多终端实时在线查看同一和多个文档。多人查看同一份文档时，通过权限、筛选等控制操作查看文档不同的内容。

10.1.3.2 多人编辑

多人、多终端实时在线编辑、保存同一文档，查询参与编辑者，进行编辑锁控制。

10.1.3.3 多人修订批注

多人修订、批注同一文档，浏览、保存修订记录，删除、回复批注，查看不同修订批注版本，显示不同版本差异。

10.1.3.4 稿件拆分与合并

多人协作时，拆分指定稿件，合并指定翻译稿。

10.1.3.5 协同译审

在线多人协同审校。

10.1.3.6 多人历史版本查看

记录每个拥有文档编辑权限的多个成员对文档做的任何更新处理记录。多人同时预览文档的历史版本。

10.1.3.7 消息通知

消息通知宜包括但不限于以下类型。

- a) 文件协同通知：多人协同进行文件的修改、更新、删除、移动、共享、恢复历史版本等操作时，通过消息发送通知到对应成员，告知文件的变动更新情况。宜支持第三方应用，通过订阅指定文件的动态，获取文件的变更通知。
- b) 团队变更通知：团队成员的退出、增加、移出等，发送通知给相关人员。
- c) 系统通知：团队存储空间不足时推送空间预警等通知；审核待处理流程到指定节点时通知相关人员。

10.1.4 辅助翻译

10.1.4.1 预翻译

对文档进行整体匹配翻译，匹配到的内容自动变为译文。宜支持与机器翻译引擎的集成。

10.1.4.2 智能匹配

智能匹配语料库中语料，自动推送最适合的语料。

10.1.4.3 质量检查

质量检查包括但不限于以下功能：

- a) 低错检查：检查数字、标点、电话号码等非译元素；
- b) 漏译检索：批量检索未被翻译的语句，直接定位原文功能；
- c) 术语一致性检查：译文中未使用指定术语的相应翻译时，提示术语不一致；
- d) 拼写及语法检查：检查不同语种的拼写错误、语法错误。

10.1.4.4 术语提示

翻译过程中，匹配到的术语以突出显示的方式进行提醒。

10.1.4.5 术语定义

翻译过程中实时添加术语及术语定义。

10.1.4.6 原译文切换

以不同方式显示原文、译文、原译文，各种显示方式能进行切换。

10.1.4.7 配置管理

设置匹配率、文档默认存储路径、默认语言、时间格式、字体格式等；自定义翻译结果的显示方式；设置审校修订标识的显示或隐藏。

10.1.5 译后稿导出

导出各种文件类型的译文。

10.1.6 文档管理

文档管理包括以下功能：

- a) 新建、删除、移动、复制、重命名文档；
- b) 上传文档至服务器或云端、下载文档至本地；
- c) 标记文档，对文档进行分类索引、管理和快速查找；

-
- d) 跟踪管理文档全生命周期,文档全生命周期视图是对文档生成、修改、复制、授权、流转、销毁等全部过程中的关键信息的记录及相应的表现形式。全生命周期视图以时间轴的形式进行展示。生成指定文档的全生命周期视图。

10.1.7 文档分享

分享文档给其它用户,包括但不限于文档访问链接生成、文档访问链接访问控制的方式:

- a) 文档访问链接生成:为每一文档生成相应的文档访问链接,通过文档访问链接对文档进行访问,复制文档访问连接分享给其他用户进行访问;
- b) 文档访问链接访问控制:
 - 1) 生成链接提取密码;
 - 2) 关闭访问链接;
 - 3) 指定外发对象;
 - 4) 指定多种外发渠道;
 - 5) 设定外发范围、操作权限及有效期。

10.1.8 版本管理

版本管理主要包括历史版本生成、预览、删除及还原:

- a) 历史版本生成:生成单个文件的历史版本、同时生成多个文件的历史版本、按用户权限生成指定文件的历史版本、设置文件备份期限;
- b) 历史版本预览:查看文件的指定历史版本、同时打开同一文件的多个历史版本;
- c) 删除及还原:还原文件到指定历史版本、删除文件的指定历史版本、设定不同级别的回收站。

10.2 对齐标引

10.2.1 对齐任务创建

创建语料对齐任务,对对齐文件中的相应语料设置篇章级标引。选择原文、译文语种,上传对齐后的双语文件。

10.2.2 对齐处理

对齐操作包括但不限于上移行、下移行、删除行、拆分句、合并句等。对齐方式包括自动对齐和人工对齐。

10.2.3 语料标引

在线标引语料句对。

10.2.4 历史任务查看

查看历史语料对齐任务。

10.2.5 术语提取

从语料中提取术语。

10.2.6 术语编辑

以在线、离线形式对术语进行编辑、修改、批注等。

11 系统管理

11.1 用户管理

用户管理包括但不限于以下功能：

- a) 创建、编辑、查询、启用、禁用、删除用户；
- b) 增加、查询、删除团队中成员；
- c) 团队中其个成员离开时交接账号所有资源，包括文档、团队及权限。

11.2 角色管理

创建不同用户角色，应符合 ZYF 00x. 1-2021 的 A. 2 中译员团队中对各角色的规定。添加、编辑、移除用户角色。

配置超级管理员角色，该角色具有查看及管理所有资源的权限。

11.3 审批管理

审批管理宜包括但不限于以下4项内容：

- a) 审批管理新创建团队；
- b) 定义审批流程模版；
示例：模版内容包括文件权限、团队组成、人员邀请等。
- c) 自定义事件的审批流程；
- d) 多节点审批。

11.4 权限管理

11.4.1 用户权限

依据用户角色分配权限。用户拥有个人私有空间，具备对应操作权限，具备用户目录管理功能。在权限范围内查看目录，按人、团队、项目设置目录权限。

11.4.2 知识库权限

分配用户对语料库、术语库的使用权限，如配置查看、编辑、删除等权限。

11.4.3 文档权限

文档权限宜满足但不限于以下要求：

- a) 设置文档浏览、编辑、打印、另存、复制等操作的细粒度权限；
- b) 按人、团队授权文档权限；
- c) 界定文档权限，在权限范围内操作文档；
- d) 设置文档的申请、审批、回收权限；
- e) 设置权限的授权对象、范围、时限；
- f) 管理权限模板包括模板的创建、编辑、删除、导入、导出；
- g) 自行设定目录权限、目录权限继承，独立设置文档内容权限、目录权限。

11.5 其他管理

系统宜包括但不限于以下管理功能：

- a) 日志管理：记录用户登录、访问等核心操作；
- b) 字典管理：用户根据需要，自行配置字典；

- c) 系统工具：字数在线统计、不同格式文档之间的相互转化；
- d) 其他功能：包括消息服务、即时通讯、电子签名签章等服务。

12 安全管理

12.1 用户安全管理

为接入应用提供用户身份认证。对用户的非正常行为进行设置、监控，提供不同的预警方式。

12.2 加解密功能

对用户指定文件类型的文档加解密。对文档进行多种形式、多层加密。加密信息显示在文档属性中供用户检索、查看。通过账号对文档加密、密钥二次加密、接入第三方算法。加密算法应符合 GB/T 32905-2016、GB/T 32907-2016、GB/T 32918.1-2016、GB/T 32918.2-2016、GB/T 32918.3-2016、GB/T 32918.4-2016、GB/T 32918.5-2017、GB/T 33133.1-2016、GB/T 35276-2017、GM/T 0044-2016 的规定。

12.3 泄密溯源追踪

通过打印、截屏、拍摄等多种泄密渠道溯源文档，具有文档水印信息隐写、水印信息提取、打印分发插件、编码库管理等功能。

示例：利用屏幕浮水印或文档阅读浮水印设置，进行文档来源的追溯。

12.4 操作行为管控

对文档信息防泄露进行事中、事后管控处理。

12.5 预览安全防护

对内容输出进行安全管控。

示例：在文档预览时，禁止内容复制、禁止内容打印等。

13 存储管理

13.1 存储机制选择

具有独立部署、云文档集成部署、服务器集群方式部署等多种部署方式。支行数据分块存储、加密存储（AES或接入第三方加密工具）、多节点存储等多种存储方式。

13.2 存储空间管理

对文档进行云存储时，不同类型存储空间的初始化、停用、动态调整，实时生效。自动清理、停用存储空间中符合条件的文档。

13.3 存储方式管理

本地或远程存储文档，文档权限在远程与本地保持一致。迁移不同存储空间之中的数据，包括历史数据的关系接管、本地目录权限与云文档目录权限映射、本地目录结构与云文档目录结构对应。

学位论文初稿