



团 体 标 准

T/TAC 7.2—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

2022-04-01 发布

2022-05-01 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 加工处理	3
7.1 翻译管理	3
7.2 对齐标引	6
8 知识库管理	6
8.1 语料库管理	6
8.2 术语库管理	7
8.3 元数据管理	8
9 知识分析	8
9.1 知识关联	8
9.2 查重分析	8
9.3 报表生成	8
10 应用服务	9
10.1 信息检索	9
10.2 统一门户	9
10.3 用户登录	9
10.4 结果展示	9
11 系统管理	9
11.1 用户管理	9
11.2 角色管理	9
11.3 审批管理	9
11.4 权限管理	10
11.5 其他管理	10
12 安全管理	10
12.1 用户安全管理	10
12.2 加解密功能	10

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 T/TAC 7《中国特色话语翻译 高端语料库建设》的第 2 部分。T/TAC 7 已经发布了以下部分：

——第 1 部分：基本要求；

——第 2 部分：系统架构；

——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。

本文件由中国翻译研究院提出。

本文件由中国翻译协会归口。

本文件起草单位：中国标准化研究院、当代中国与世界研究院、传神语联网网络科技股份有限公司、中国翻译协会、对外经济贸易大学、北京悦尔信息技术有限公司、中国质量标准出版传媒有限公司。

本文件主要起草人：曹馨宇、于运全、杨平、范大祺、闫栗丽、张彤、赵庆、魏洁、刘强、王海涛、吴晓蕊、崔启亮、蒙永业、刘晓东。

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,以及语料库建设管理系统的加工处理、知识库管理、知识分析、应用服务、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(所有部分) SM9 标识密码算法
- T/TAC 7.1—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

T/TAC 7.1—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(3.1)数据的数据库集

[来源:GB/T 17532—2005,7.8]

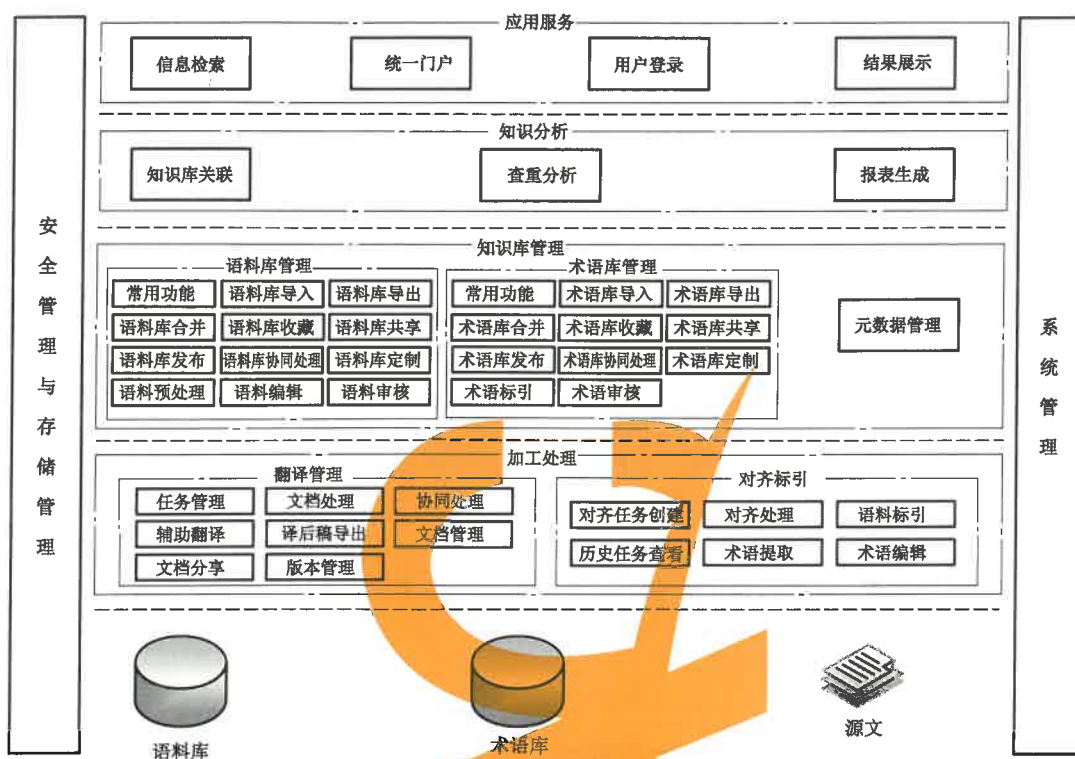


图 1 中国特色话语翻译高端语料库建设管理系统架构

7 加工处理

7.1 翻译管理

7.1.1 任务管理

任务管理包括以下功能：

- 任务派发：指定翻译者、审校者，指派多级译审任务；
- 进度监控：监控任务进度；
- 离线任务导入导出：导出译文以供离线翻译、编辑、审校，将离线翻译、编辑、审校的译文导入系统，导入导出时应保留审校痕迹。

7.1.2 文档处理

文档处理包括以下功能：

- 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；
- 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；
- 查找替换：在文档中检索相关内容，批量替换查到内容；
- 文档标密：标写文档密级；
- 离线访问：支持离线权限控制和逾期访问自毁；
- 译后编辑：对预翻译后文档内容进行再编辑。

- c) 术语一致性检查:原文中术语未按照规定译法进行翻译时,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。

7.1.4.4 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。

7.1.4.5 术语及定义

翻译过程中实时添加术语及其定义。

7.1.4.6 原译文切换

以不同方式显示原文、译文、原译文,且能在各种显示方式间进行切换。

7.1.4.7 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。

7.1.5 译后稿导出

将译文导出成各种类型的文件。

7.1.6 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) 对文档进行全生命周期跟踪管理,生成指定文档的全生命周期视图。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。

7.1.7 文档分享

能将文档分享给其他用户,包括但不限于以下功能:

- 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。

7.1.8 版本管理

版本管理主要包括历史版本的生成、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;

8.1.5 语料库收藏

标记语料或语料库,以便快速访问。

8.1.6 语料库共享

指定语料库的共享对象、共享范围、共享方式。

8.1.7 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。

8.1.8 语料库协同处理

多人协同工作、相互配合,共同完成语料库的建设管理等工作。

8.1.9 语料库定制

创建多种类型的语料库。

示例:个人语料库、项目语料库。

8.1.10 语料预处理

对通过其他方式采集到的语料进行清洗、查重等。

8.1.11 语料编辑

对语料内容进行编辑、修改、批注等。

8.1.12 语料审核

对指定语料进行审核,能查看审核痕迹。

8.2 术语库管理

8.2.1 常用功能

主要包括术语库的新建、删除、修改、查找等功能。

8.2.2 术语库导入

批量导入术语到术语库。

8.2.3 术语库导出

将术语库中选定的术语导出到指定类型的数据文件。

8.2.4 术语库合并

合并多个术语库为一个术语库。

8.2.5 术语库收藏

标记术语或术语库,以便快速访问。

10 应用服务

10.1 信息检索

信息检索应包含但不限于以下功能：

- a) 统一检索：使用同一界面进行多语种检索，并展示检索结果；
- b) 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；
- c) 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。

10.2 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。

10.3 用户登录

用户能够通过密码、手机验证码等多种方式登录；不同用户账号之间能够相互切换；能够修改密码；密码失效后，应重新登录。

10.4 结果展示

以多种形式展示检索结果。

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 T/TAC 7.1—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内容：

- a) 审批管理新创建团队；
- b) 定义审批流程模版；
- c) 自定义事件的审批流程；
- d) 多节点审批。

示例：模版内容包括文件权限、团队组成、人员邀请等。

12.4 操作行为管控

对可能导致文档泄密的操作行为进行监督、管理、提醒及预警。

12.5 预览安全防护

防止文档预览过程中可能出现的泄密行为,如截屏、复制、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。支持数据分块存储、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,动态调整能实时生效。能自动清理、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时,宜进行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。

中国翻译协会
团体标准
中国特色话语翻译 高端语料库建设
第2部分:系统架构
T/TAC 7.2—2021

*

中国标准出版社出版发行
北京市朝阳区和平里西街甲2号(100029)
北京市西城区三里河北街16号(100045)
网址 www.spc.net.cn
总编室:(010)68533533 发行中心:(010)51780238
读者服务部:(010)68523946
中国标准出版社秦皇岛印刷厂印刷
各地新华书店经销

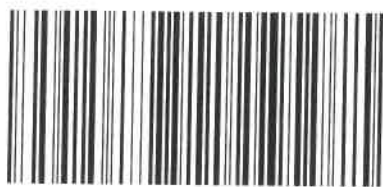
*

开本 880×1230 1/16 印张 1.25 字数 30 千字
2022年3月第一版 2022年3月第一次印刷

*

书号: 155066·5-4237 定价 31.00 元

如有印装差错 由本社发行中心调换
版权专有 侵权必究
举报电话:(010)68510107



T/TAC 7.2-2021



码上扫一扫 正版服务到