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。