# 说 明 书 附 图

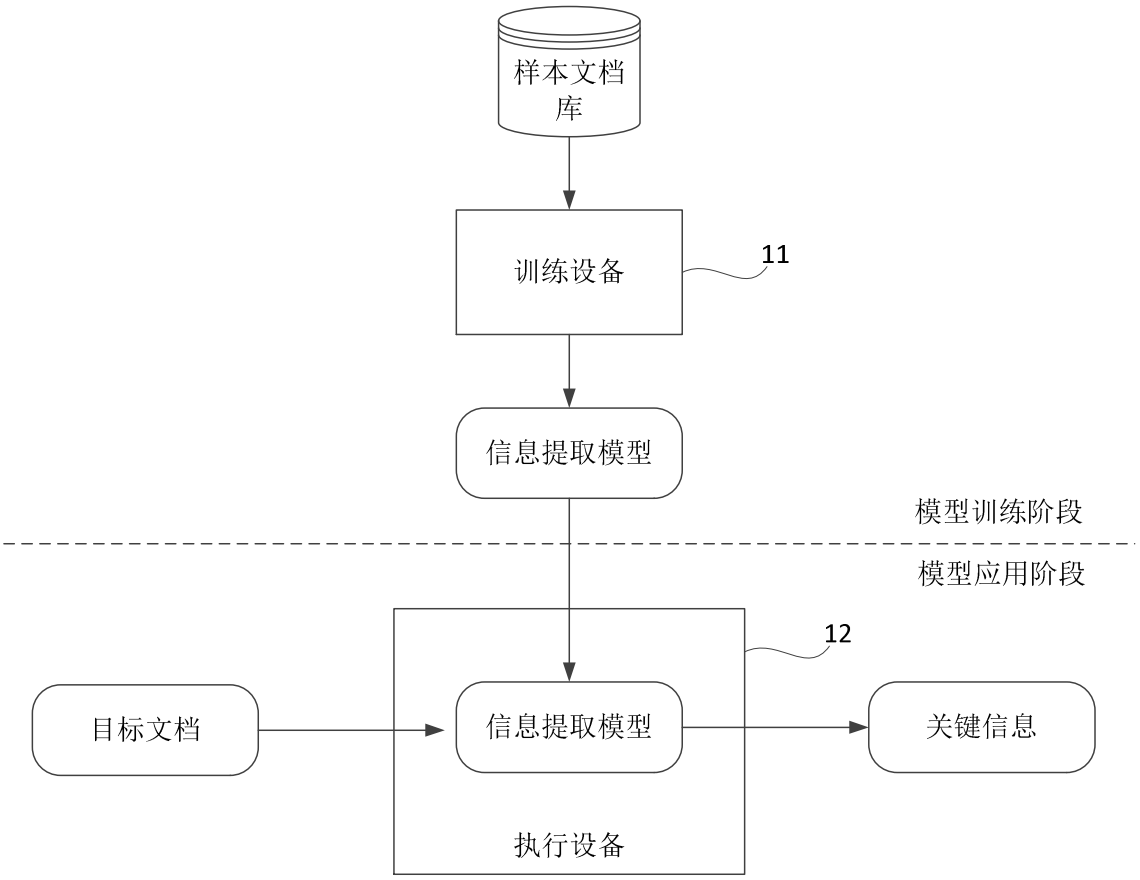


图 1

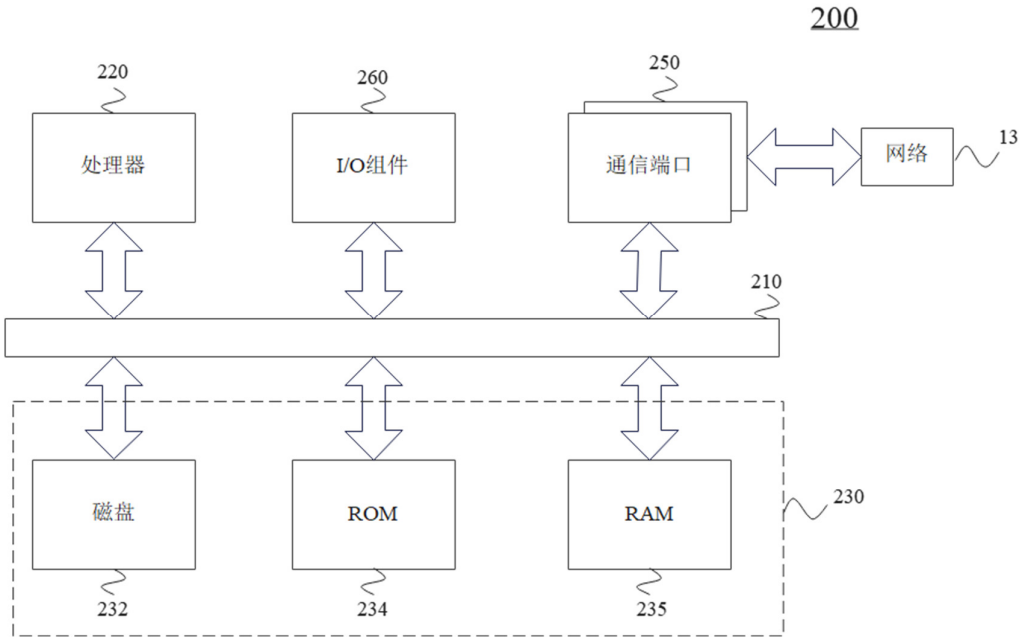
200
220
处理器
260
I/O组件
250
通信端口
网络
13
210
磁盘
ROM
RAM
230
232
234
235

图 2

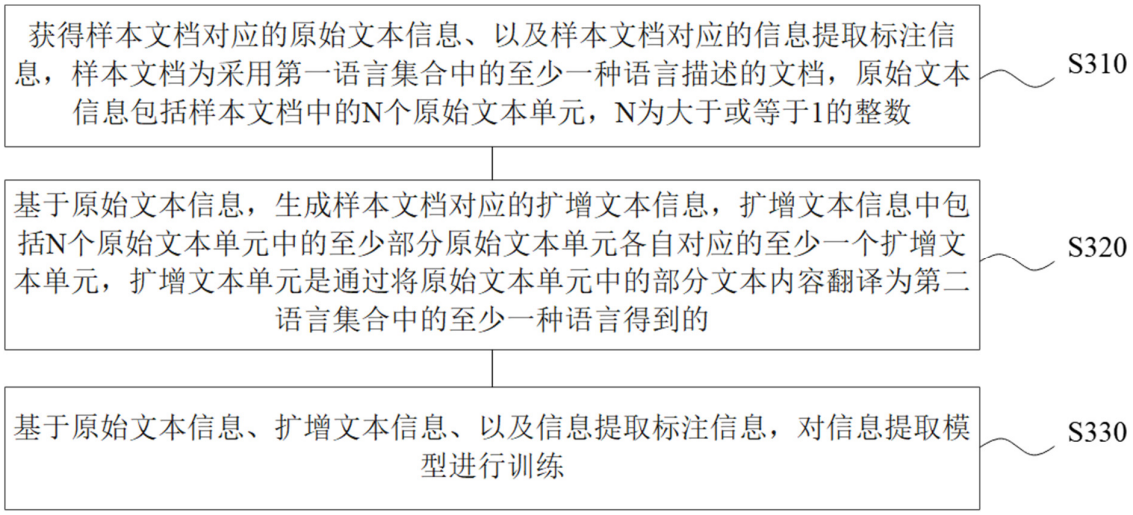
获得样本文档对应的原始文本信息、以及样本文档对应的信息提取标注信息，样本文档为采用第一语言集合中的至少一种语言描述的文档，原始文本信息包括样本文档中的N个原始文本单元，N为大于或等于1的整数
S310
基于原始文本信息，生成样本文档对应的扩增文本信息，扩增文本信息中包括N个原始文本单元中的至少部分原始文本单元各自对应的至少一个扩增文本单元，扩增文本单元是通过将原始文本单元中的部分文本内容翻译为第二语言集合中的至少一种语言得到的
S320
基于原始文本信息、扩增文本信息、以及信息提取标注信息，对信息提取模型进行训练
S330

图 3

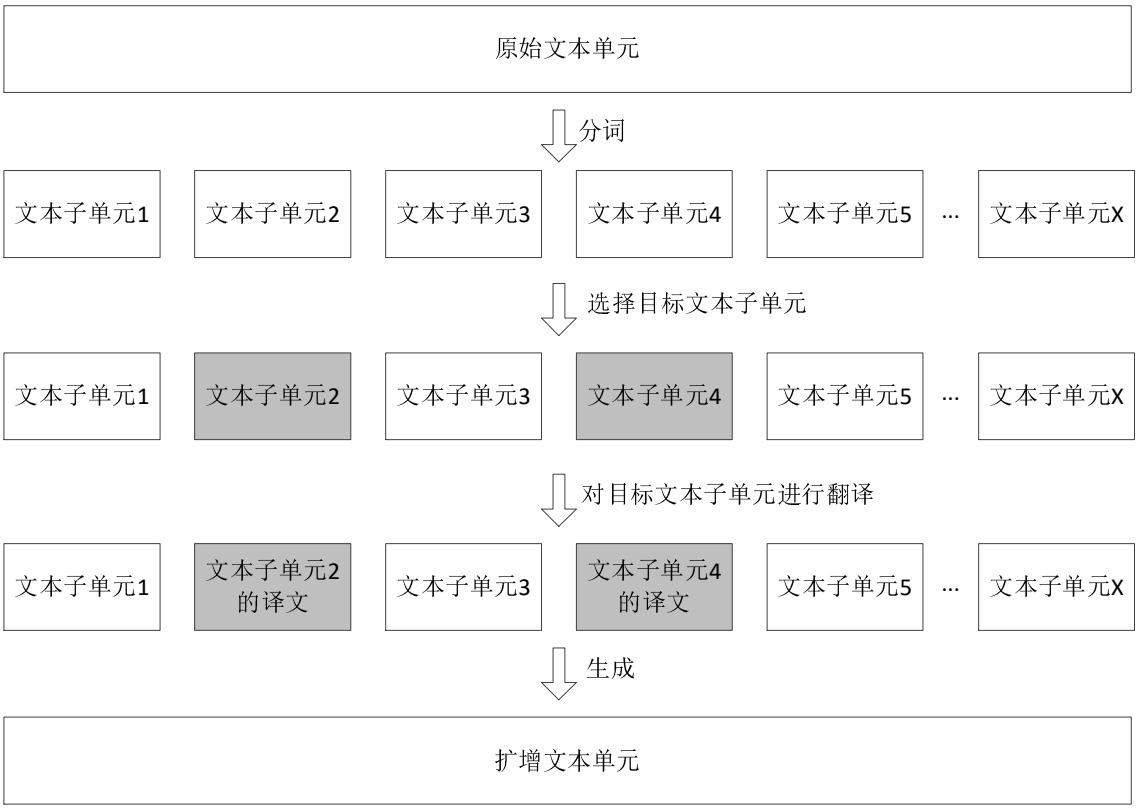
原始文本单元
分词
文本子单元1
文本子单元2
文本子单元3
文本子单元4
文本子单元5
…
文本子单元X
选择目标文本子单元
文本子单元1
文本子单元2
文本子单元3
文本子单元4
文本子单元5
…
文本子单元X
对目标文本子单元进行翻译
文本子单元1
文本子单元2的译文
文本子单元3
文本子单元4的译文
文本子单元5
…
文本子单元X
生成
扩增文本单元

图 4


原始文本单元
Account Name
:
HONGKONG ABC TECHNOLOGY
CO., LIMITED
翻译为日文
翻译为日文
アカウント名
:
HONGKONG ABC TECHNOLOGY
有限会社
扩增文本单元

图 5A


原始文本单元
Account Name
:
HONGKONG ABC TECHNOLOGY
CO., LIMITED
翻译为日文
翻译为韩文
アカウント名
:
HONGKONG ABC TECHNOLOGY
유한회사
扩增文本单元

图 5B


原始文本单元
Account Name
:
HONGKONG ABC TECHNOLOGY
CO., LIMITED
翻译为韩文
翻译为日文
계정 이름
:
HONGKONG ABC TECHNOLOGY
有限会社
扩增文本单元

图 5C

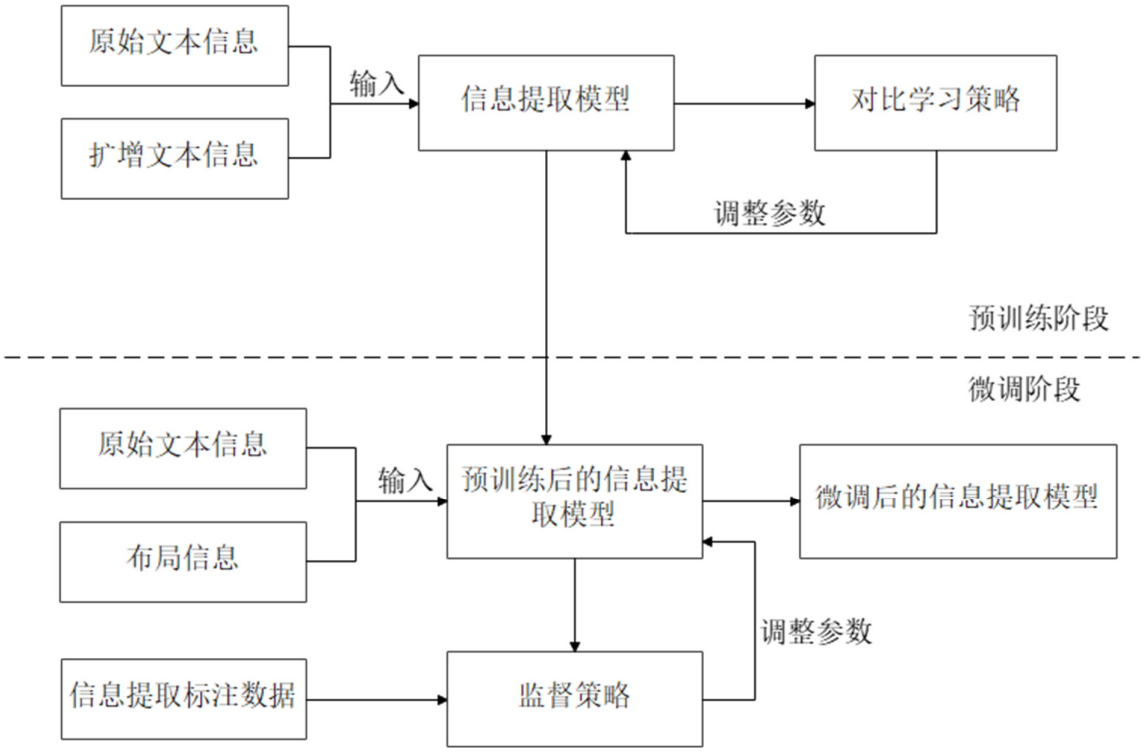
原始文本信息
扩增文本信息
输入
信息提取模型
对比学习策略
调整参数
预训练阶段
微调阶段
原始文本信息
布局信息
输入
预训练后的信息提取模型
微调后的信息提取模型
调整参数
信息提取标注数据
监督策略

图 6

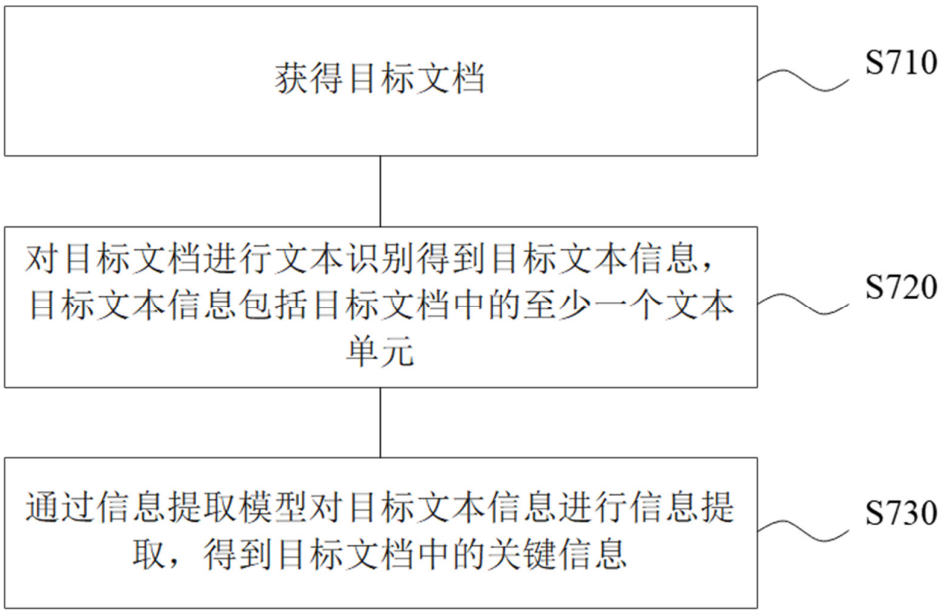
获得目标文档
S710
对目标文档进行文本识别得到目标文本信息，目标文本信息包括目标文档中的至少一个文本单元
S720
通过信息提取模型对目标文本信息进行信息提取，得到目标文档中的关键信息
S730

图 7