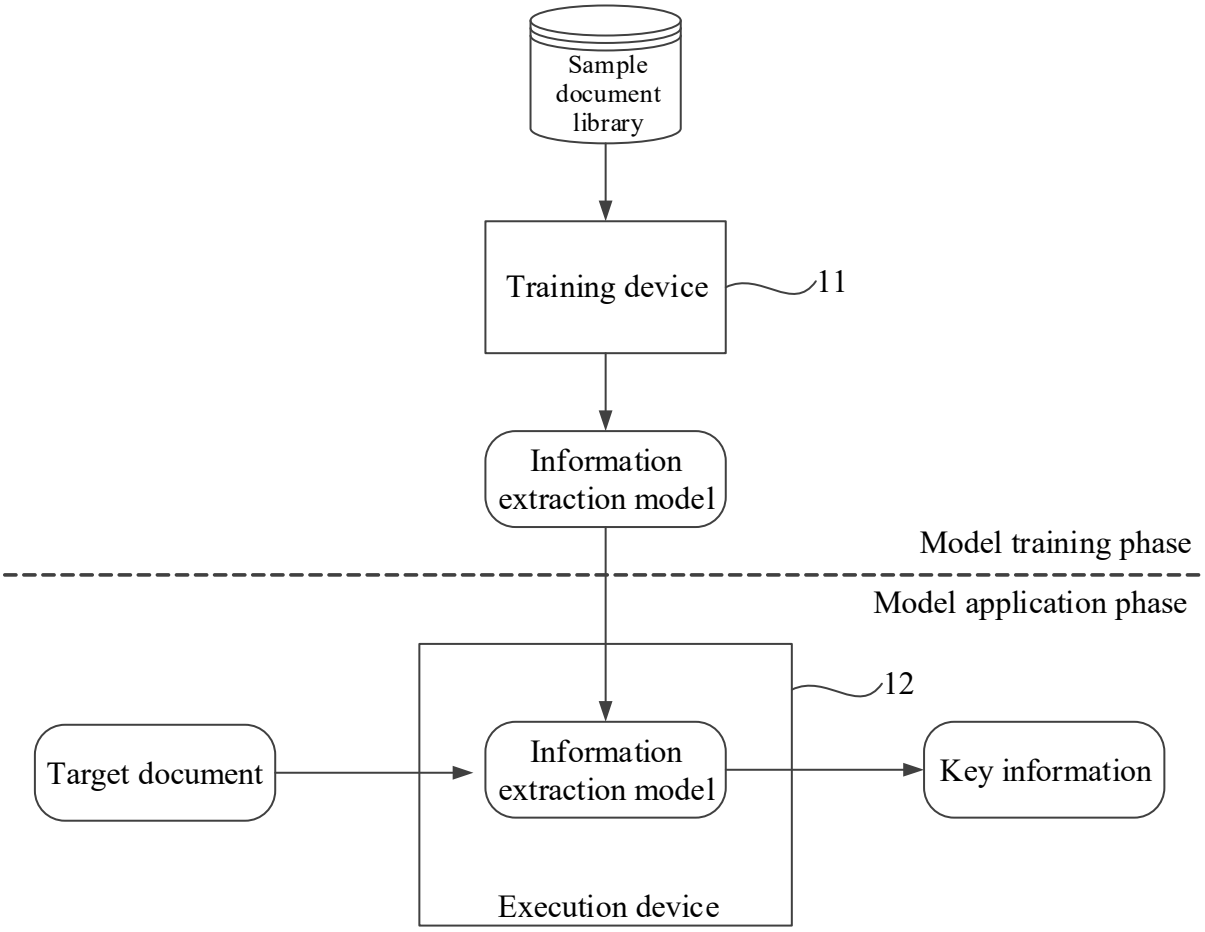

Sample document library
Training device
11
Information extraction model
Model training phase
Model application phase
12
Target document
Information extraction model
Key information
Execution device

FIG. 1

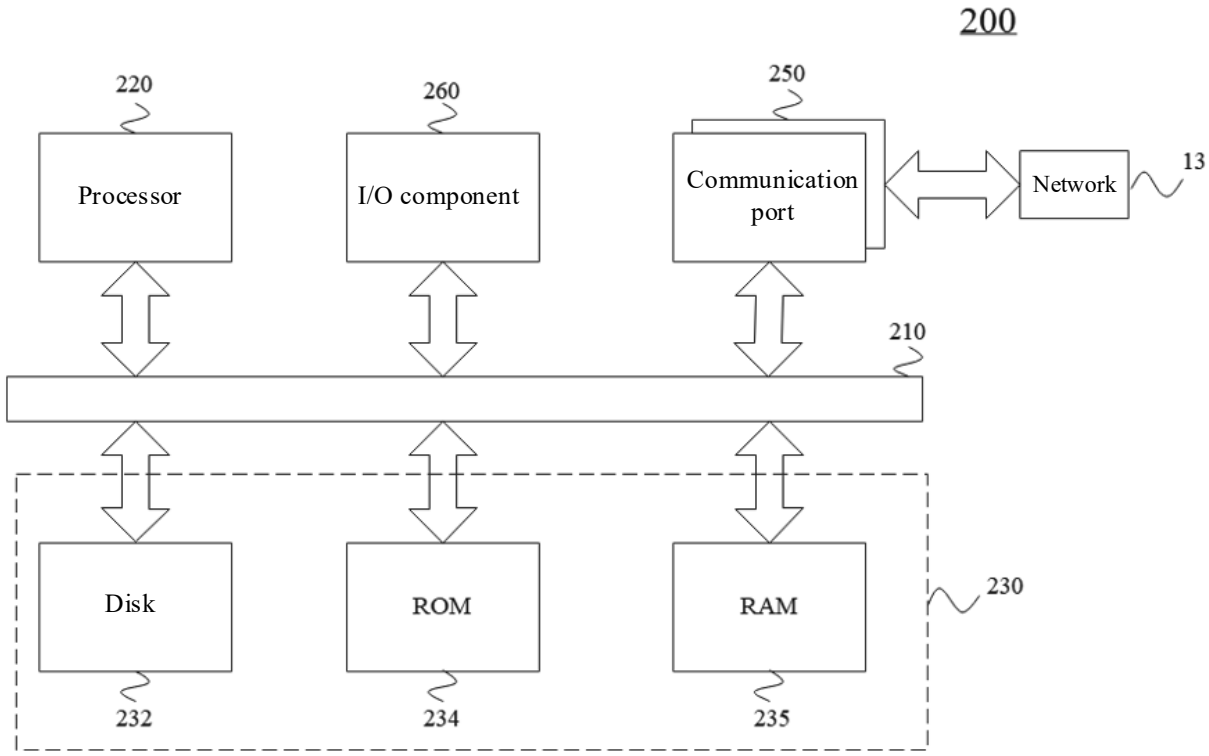

200
220
260
250
Processor
I/O component
Communication port
Network
13
210
Disk
ROM
RAM
230
232
234
235

FIG. 2

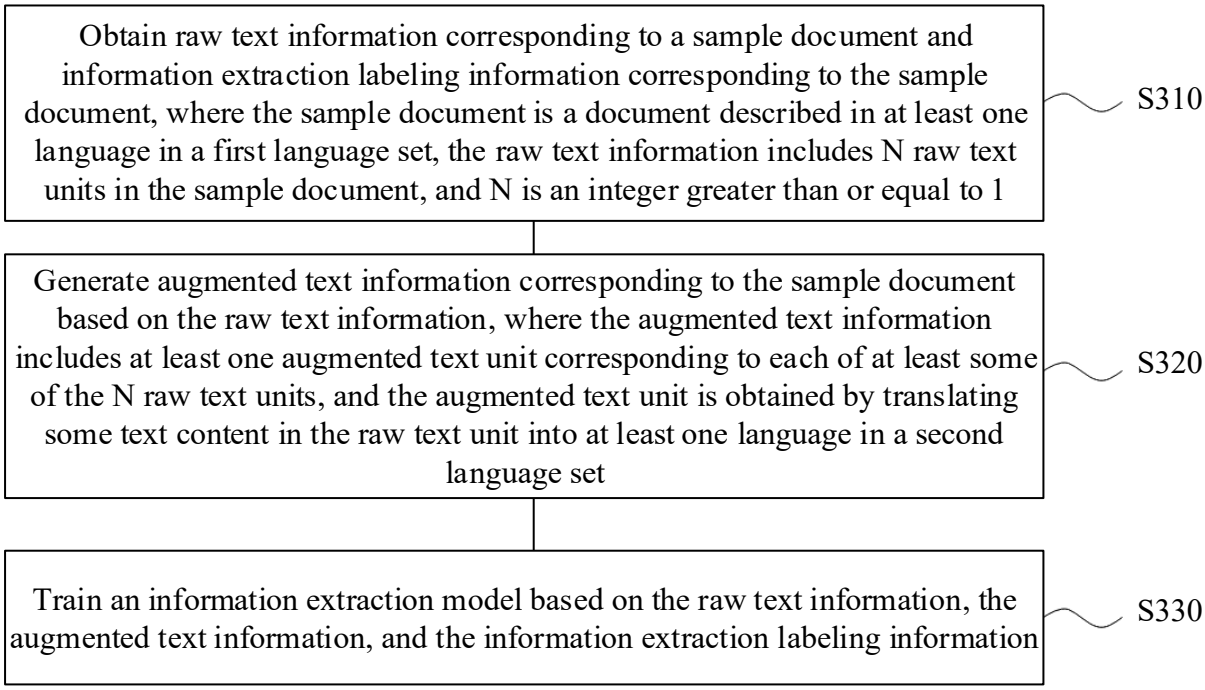

Obtain raw text information corresponding to a sample document and information extraction labeling information corresponding to the sample document, where the sample document is a document described in at least one language in a first language set, the raw text information includes N raw text units in the sample document, and N is an integer greater than or equal to 1
S310
Generate augmented text information corresponding to the sample document based on the raw text information, where the augmented text information includes at least one augmented text unit corresponding to each of at least some of the N raw text units, and the augmented text unit is obtained by translating some text content in the raw text unit into at least one language in a second language set
S320
Train an information extraction model based on the raw text information, the augmented text information, and the information extraction labeling information
S330

FIG. 3

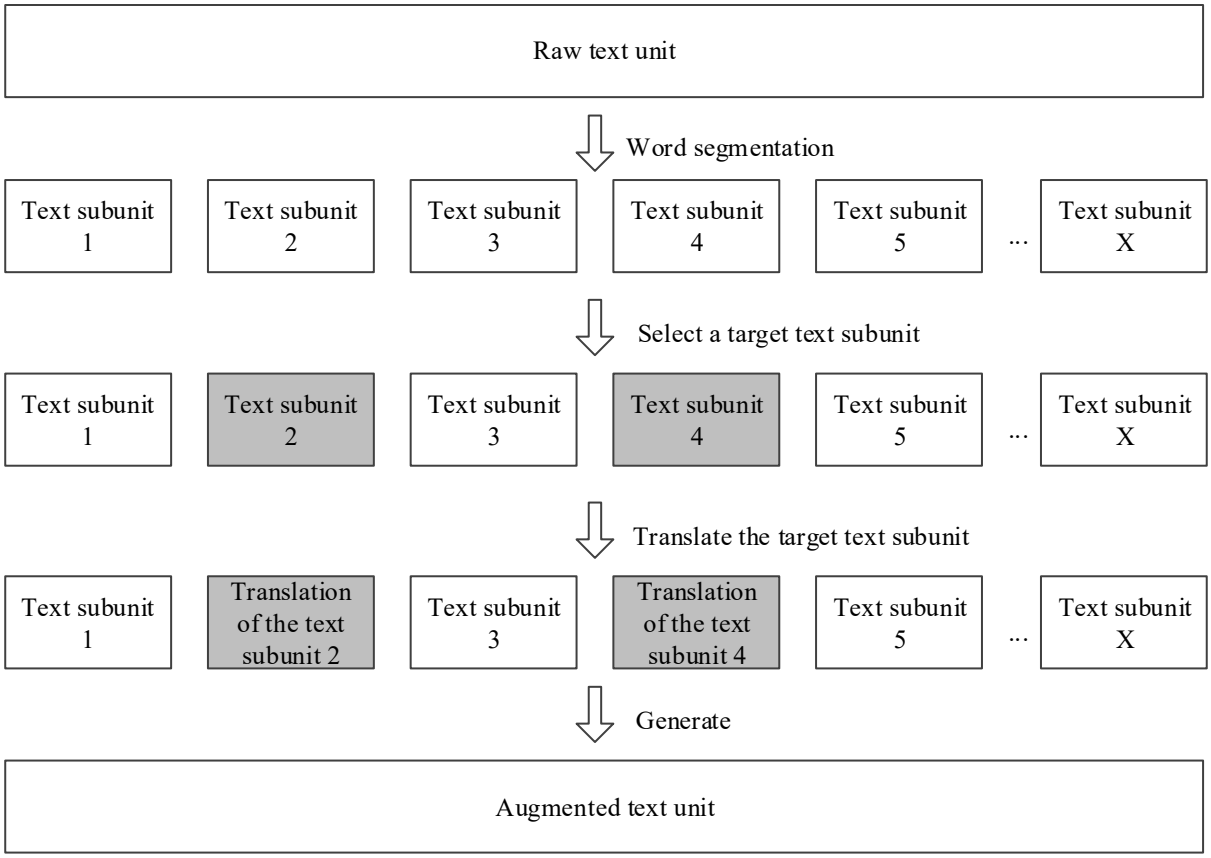

Raw text unit
Word segmentation
Text subunit 1
Text subunit 2
Text subunit 3
Text subunit 4
Text subunit 5
...
Text subunit X
Select a target text subunit
Text subunit 1
Text subunit 2
Text subunit 3
Text subunit 4
Text subunit 5
...
Text subunit X
Translate the target text subunit
Text subunit 1
Translation of the text subunit 2
Text subunit 3
Translation of the text subunit 4
Text subunit 5
...
Text subunit X
Generate
Augmented text unit

FIG. 4

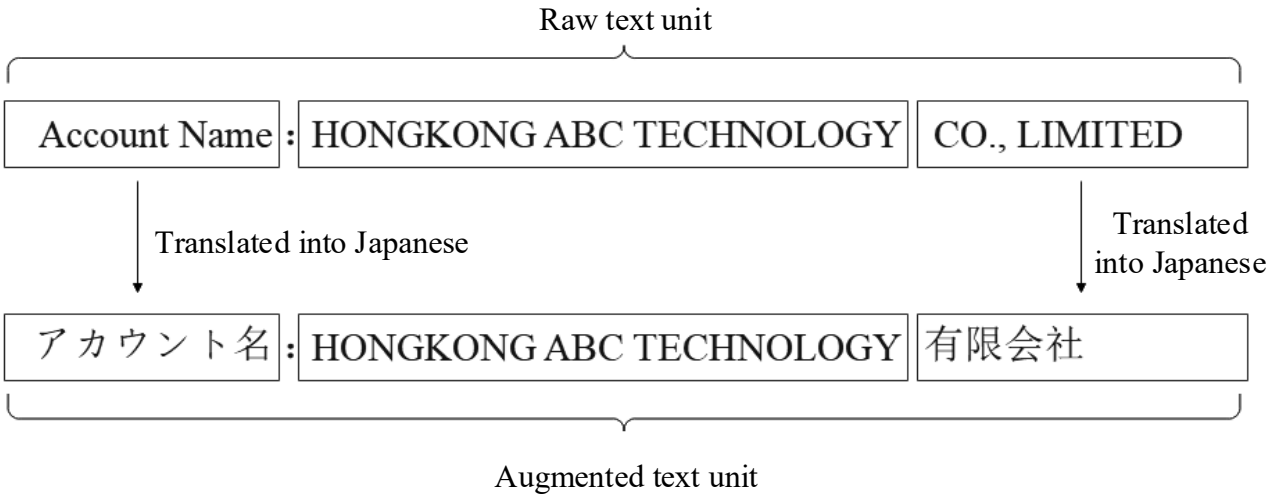

Raw text unit
Account Name
:
HONGKONG ABC TECHNOLOGY
CO., LIMITED
Translated into Japanese
Translated into Japanese
アカウント名
:
HONGKONG ABC TECHNOLOGY
有限会社
Augmented text unit

FIG. 5A

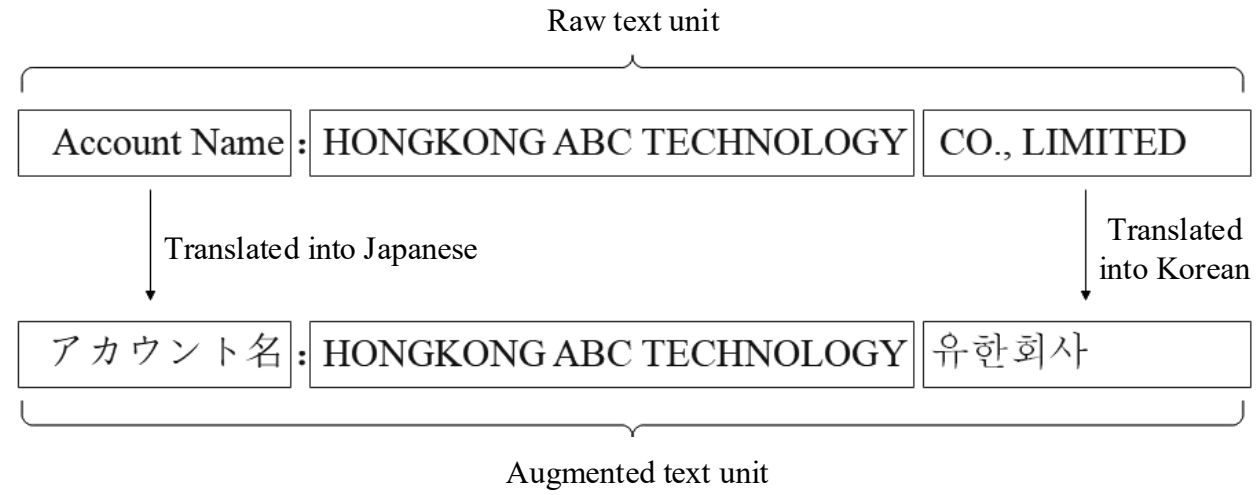

Raw text unit
Account Name
:
HONGKONG ABC TECHNOLOGY
CO., LIMITED
Translated into Japanese
Translated into Korean
アカウント名
:
HONGKONG ABC TECHNOLOGY
유한회사
Augmented text unit

FIG. 5B

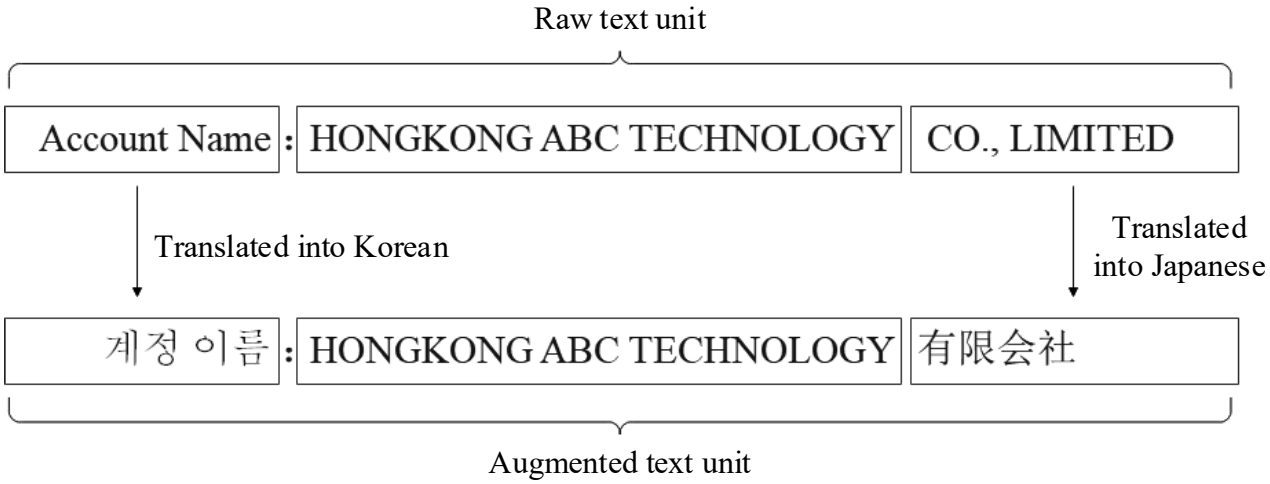

Raw text unit
Account Name
:
HONGKONG ABC TECHNOLOGY
CO., LIMITED
Translated into Korean
Translated into Japanese
계정 이름
:
HONGKONG ABC TECHNOLOGY
有限会社
Augmented text unit

FIG. 5C

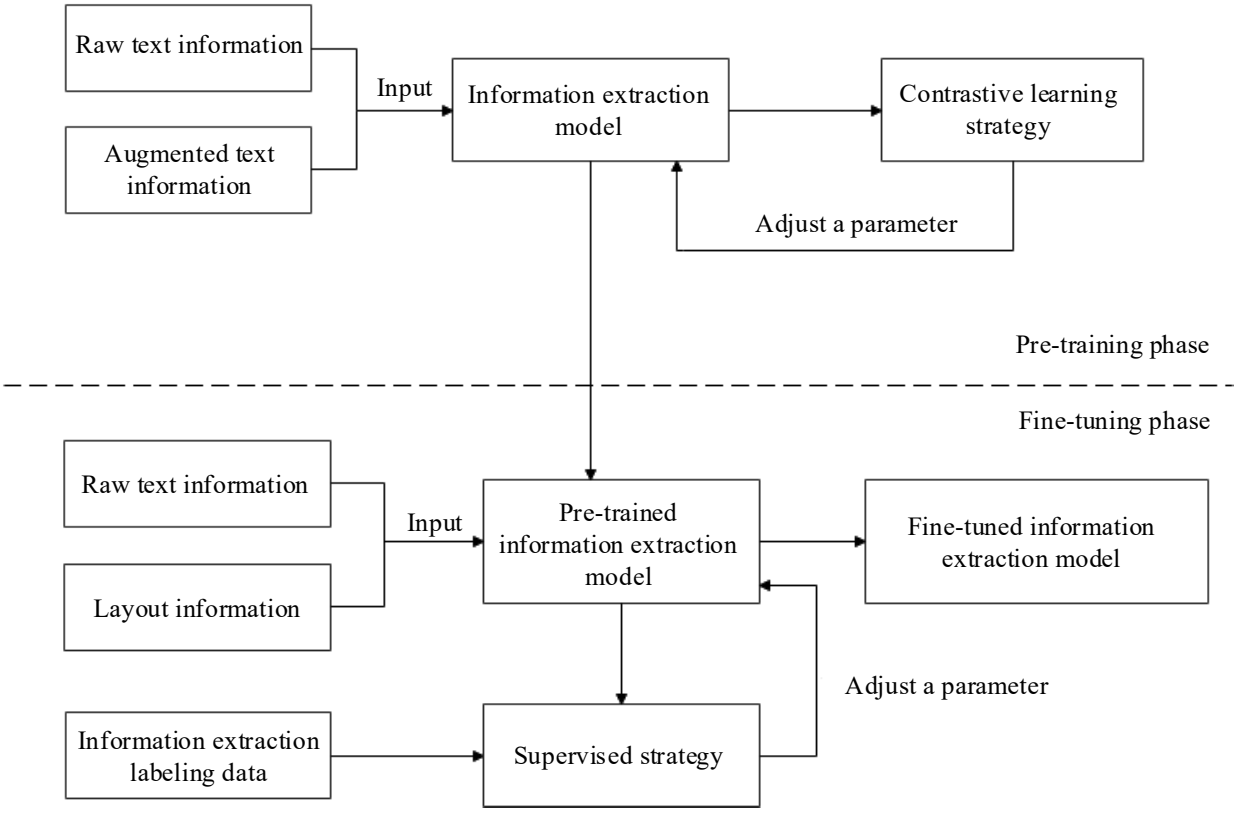

Raw text information
Augmented text information
Input
Information extraction model
Contrastive learning strategy
Adjust a parameter
Pre-training phase
Fine-tuning phase
Raw text information
Layout information
Input
Pre-trained information extraction model
Fine-tuned information extraction model
Adjust a parameter
Information extraction labeling data
Supervised strategy

FIG. 6

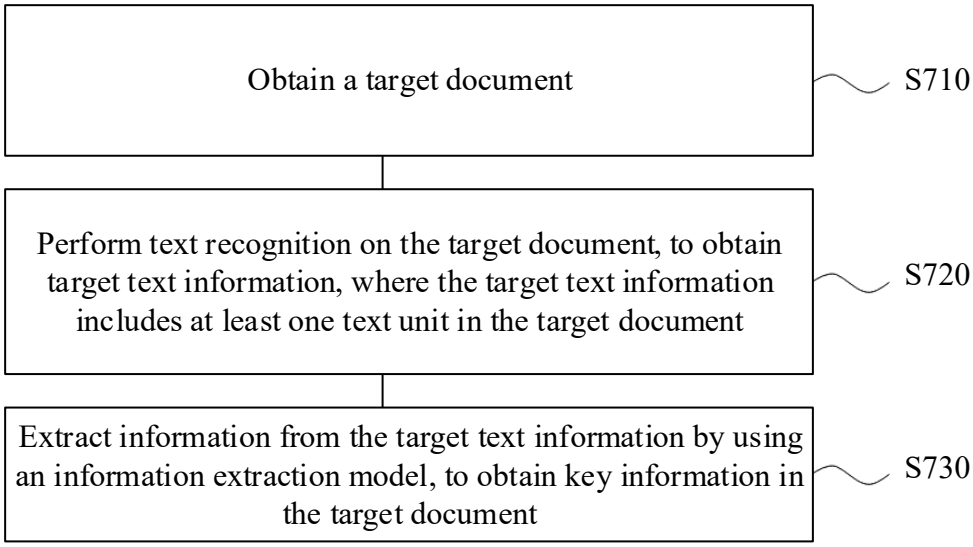

Obtain a target document
S710
Perform text recognition on the target document, to obtain target text information, where the target text information includes at least one text unit in the target document
S720
Extract information from the target text information by using an information extraction model, to obtain key information in the target document
S730

FIG. 7