(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 109522538 B
(45)授权公告日 2021.10.29

(21)申请号 201811437473.2
(22)申请日 2018.11.28
(65)同一申请的已公布的文献号
申请公布号 CN 109522538 A
(43)申请公布日 2019.03.26
(73)专利权人 腾讯科技(深圳)有限公司
地址 518057 广东省深圳市南山区高新区科技中一路腾讯大厦35层
(72)发明人 张凝 钱昉 刘江伟 熊唯 刘德峰 于海通
(74)专利代理机构 北京三高永信知识产权代理有限责任公司 11138
代理人 张所明
(51)Int.Cl.
G06F 40/177(2020.01)

(56)对比文件
CN 107818075 A,2018.03.20
CN 104090850 A,2014.10.08
US 2010057704 A1,2010.03.04
王彦博.Excel分列技巧一则.《电脑知识与技术》.2013,43.

审查员 何祥鹏

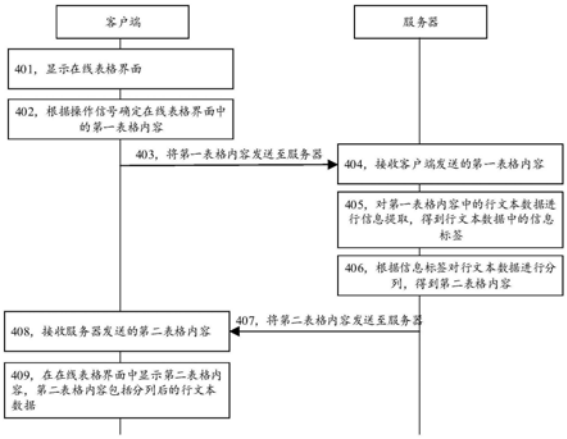
权利要求书5页 说明书26页 附图23页

(54)发明名称

表格内容的自动分列方法、装置、设备及存储介质

(57)摘要

本申请公开了一种表格内容的自动分列方法、装置、设备及存储介质，该方法包括：接收客户端发送的第一表格内容，所述第一表格内容包括待分列的行文本数据；对所述第一表格内容中的行文本数据进行信息提取，得到所述行文本数据中的信息标签；根据所述信息标签对所述行文本数据进行分列，得到第二表格内容，所述第二表格内容包括分列后的行文本数据；向所述客户端发送所述第二表格内容。本申请能够将属于不同信息标签的实体信息分列至不同的列中，不需要依赖于简单的分隔符作为分列条件，而是利用实体信息的语义或特征来进行分列，从而提高了分列功能的准确性和成功率。



CN 109522538 B

1.一种表格内容的自动分列方法,其特征在于,所述方法包括:

接收客户端发送的第一表格内容,所述第一表格内容包括待分列的行文本数据,所述第一表格内容是根据在线表格界面中的第二单元格区域中的单元格渲染值确定的,所述第二单元格区域中的单元格是根据区域选择信号选定的,所述区域选择信号是所述客户端接收到的;

对所述第一表格内容中的行文本数据进行信息提取,得到所述行文本数据中的信息标签,所述信息标签是用于标识实体信息的信息类别的信息;

根据所述信息标签对所述行文本数据进行分列,得到第二表格内容,所述第二表格内容包括分列后的行文本数据;

向所述客户端发送所述第二表格内容,所述客户端用于在所述在线表格界面中显示所述第二表格内容。

2.根据权利要求1所述的方法,其特征在于,所述对所述第一表格内容中的行文本数据进行信息提取,得到所述行文本数据中的信息标签,包括:

提取所述第一表格内容中的多个行文本数据;

对于所述多个行文本数据中的任一行文本数据,将所述行文本数据输入信息提取模型中,得到所述信息提取模型所提取的实体信息;

将所述行文本数据中的所述实体信息,标注与所述信息提取模型对应的信息标签;

其中,所述信息提取模型是多个信息提取模型中的一个。

3.根据权利要求2所述的方法,其特征在于,所述信息提取模型包括:采用机器学习特征进行信息提取的第一信息提取模型;

所述将所述行文本数据输入信息提取模型中,得到所述信息提取模型所提取到的实体信息,包括:

将所述行文本数据输入所述第一信息提取模型中,预测出文字特征符合所述机器学习特征的第一文字串;

当预测出文字特征符合所述机器学习特征的第一文字串时,将所述第一文字串确定为提取到的所述实体信息。

4.根据权利要求2所述的方法,其特征在于,所述信息提取模型包括:采用枚举词库进行信息提取的第二信息提取模型;

所述将所述行文本数据输入信息提取模型中,得到所述信息提取模型所提取到的实体信息,包括:

将所述行文本数据输入所述第二信息提取模型中,确定是否存在与所述枚举词库匹配的第二文字串;

当存在与所述枚举词库匹配的第二文字串时,将所述第二文字串确定为提取到的所述实体信息。

5.根据权利要求2所述的方法,其特征在于,所述信息提取模型包括:采用正则表达式进行信息提取的第三信息提取模型;

所述将所述行文本数据输入信息提取模型中,得到所述信息提取模型所提取到的实体信息,包括:

将所述行文本数据输入所述第三信息提取模型中,确定是否存在与所述正则表达式匹

配的第三文字串；

当存在与所述正则表达式匹配的第三文字串时，将所述第三文字串确定为提取到的所述实体信息。

6.根据权利要求2至5任一所述的方法，其特征在于，所述第一表格内容是以文本形式存储的文本数据；

所述提取所述第一表格内容中的多个行文本数据，包括：

识别所述文本数据的初始分列位置和末尾分列位置；

对位于所述初始分列位置和所述末尾分列位置之间的文本数据片段，按照优先级顺序依次采用至少一种分行规则进行分行处理，并在分行处理成功时得到所述多个行文本数据；

其中，所述分行规则包括：序号分行规则、空格分行规则、分割符分行规则中的至少一种。

7.根据权利要求6所述的方法，其特征在于，所述识别所述文本数据的初始分列位置和末尾分列位置，包括：

在所述文本数据中识别段首特征，所述段首特征包括语义关键字、序号关键字、分隔符中的至少一种；

将所述段首特征所在位置的前一个位置或后一个位置识别为所述初始分列位置；

将所述文本数据的最后一个位置识别为所述末尾分列位置。

8.根据权利要求1至5任一所述的方法，其特征在于，所述根据所述信息标签对所述行文本数据进行分列，得到第二表格内容，包括：

将各个所述行文本数据中具有相同信息标签的实体信息对齐至同一列，将具有不同信息标签的实体信息对齐至不同列，得到所述第二表格内容。

9.根据权利要求8所述的方法，其特征在于，所述将各个所述行文本数据中具有相同信息标签的实体信息对齐至同一列，将具有不同信息标签的实体信息对齐至不同列，得到所述第二表格内容，包括：

获取n个所述行文本数据中的实体信息对，所述实体信息对包括所述实体信息和与所述实体信息对应的信息标签，n为正整数；

将第i个所述行文本数据中的所述实体信息对添加到第i个栈中，所述行文本数据和所述栈一一对应，每个实体信息对是所述栈中的一个栈元素，i为不大于n的正整数；

统计各个栈中的栈首元素中出现次数最多的第一参考信息标签，以及次栈首元素中出现次数最多的第二参考信息标签；

当第j个栈中的栈首元素的信息标签与所述第一参考信息标签不同，且与所述第二参考信息标签相同时，将所述第j个栈中的栈元素向栈尾方向移动一位，并使用空白栈元素补齐所述第j个栈中的栈首元素，j为不大于n的正整数；

当所述第j个栈中的栈首元素的信息标签与所述第一参考信息标签不同，且与所述第二参考信息标签不同时，将所述第j个栈中的栈元素向栈首方向移动一位，并将移动后的第一个栈元素设置为新增栈元素，所述新增栈元素是位于所述栈首元素之上的元素；

当n个栈中的栈首元素的信息标签均为所述第一参考信息标签时，将所述n个栈中的栈首元素移出至目标表格中序号最小的同一个空白列中。

10.根据权利要求1至5任一所述的方法，其特征在于，所述根据所述信息标签对所述行文本数据进行分列，得到第二表格内容之后，还包括：

根据各个所述行文本数据中的所述信息标签的信息标签个数，统计出所述信息标签的正常值；

当存在所述信息标签个数大于所述正常值的第一行文本数据，且所述第一行文本数据的相邻行的所述信息标签个数等于所述正常值时，将所述第一行文本数据进行重新分行，并对重新分行后的行文本数据进行重新分列；

当存在所述信息标签个数小于所述正常值的第二行文本数据，且与所述第二行文本数据相邻的第三行文本数据的所述信息标签个数大于所述正常值时，将所述第二行文本数据和所述第三行文本数据进行重新分行，并对重新分行后的行文本数据进行重新分列。

11.根据权利要求1至5任一所述的方法，其特征在于，所述根据所述信息标签对所述行文本数据进行分列，得到第二表格内容之后，还包括：

根据各个所述行文本数据中的所述信息标签的信息标签个数，统计出所述信息标签个数的正常值；

当存在所述信息标签个数大于所述正常值的最后一行文本数据，且所述最后一行文本数据的上一行的所述信息标签个数等于所述正常值时，在所述最后一行文本数据中确定干扰文字串；

去除所述最后一行文本数据中的所述干扰文字串。

12.一种表格内容的自动分列方法，其特征在于，所述方法包括：

显示在线表格界面；

接收区域选择信号；

根据所述区域选择信号，选定所述在线表格界面中的第二单元格区域中的单元格；

将所述第二单元格区域中的单元格渲染值，确定为第一表格内容，将所述第一表格内容发送至服务器；

接收所述服务器发送的第二表格内容，所述第二表格内容是所述服务器根据信息标签对所述第一表格内容中的行文本数据进行分列后得到的，所述信息标签由所述服务器对所述第一表格内容中的行文本数据进行信息提取得到，所述信息标签是用于标识实体信息的信息类别的信息；

在所述在线表格界面中显示所述第二表格内容，所述第二表格内容包括分列后的行文本数据。

13.根据权利要求12所述的方法，其特征在于，所述根据操作信号确定所述在线表格界面中的第一表格内容，包括：

接收粘贴信号；

根据所述粘贴信号，向第一单元格区域中的单元格粘贴表格内容；

将粘贴的所述表格内容确定为所述第一表格内容。

14.根据权利要求12所述的方法，其特征在于，所述将所述第二单元格区域中的单元格渲染值，确定为所述第一表格内容，包括：

按照数据分布概率模型所指示的优先级顺序，搜索所述第二单元格区域的单元格中的连续非空白行列区域；所述数据分布概率模型是根据历史数据统计出的表格内容在表格区

域中不同分布位置中出现的概率模型；

将搜索到的所述连续非空白行列区域中的单元格渲染值，确定为所述第一表格内容。

15.根据权利要求12至14任一所述的方法，其特征在于，在所述在线表格界面中显示所述第二表格内容，包括：

确定所述第二表格内容中的分列后的行文本数据所需占据的第三表格区域；

当所述第三表格区域占据除所述第二单元格区域之外的非空单元格时，显示询问控件，所述询问控件用于询问是否允许覆盖所述非空单元格；

当所述询问控件上接收到允许操作信号时，将所述分列后的行文本数据分列显示在所述第三表格区域中。

16.一种表格内容的自动分列装置，其特征在于，所述装置包括：

接收模块，用于接收客户端发送的第一表格内容，所述第一表格内容包括待分列的行文本数据，所述第一表格内容是根据在线表格界面中的第二单元格区域中的单元格渲染值确定的，所述第二单元格区域中的单元格是根据区域选择信号选定的，所述区域选择信号是所述客户端接收到的；

提取模块，用于对所述第一表格内容中的行文本数据进行信息提取，得到所述行文本数据中的信息标签，所述信息标签是用于标识实体信息的信息类别的信息；

分列模块，用于根据所述信息标签对所述行文本数据进行分列，得到第二表格内容，所述第二表格内容包括分列后的行文本数据；

发送模块，用于向所述客户端发送所述第二表格内容，所述客户端用于在所述在线表格界面中显示所述第二表格内容。

17.一种表格内容的自动分列装置，其特征在于，所述装置包括：

显示模块，用于显示在线表格界面；

确定模块，用于接收区域选择信号；根据所述区域选择信号，选定所述在线表格界面中的第二单元格区域中的单元格；将所述第二单元格区域中的单元格渲染值，确定为第一表格内容，将所述第一表格内容发送至服务器；

接收模块，用于接收所述服务器发送的第二表格内容，所述第二表格内容是所述服务器根据信息标签对所述第一表格内容中的行文本数据进行分列后得到的，所述信息标签是所述服务器对所述第一表格内容中的行文本数据进行信息提取得到的，所述信息标签是用于标识实体信息的信息类别的信息；

所述显示模块，用于在所述在线表格界面中显示所述第二表格内容，所述第二表格内容包括分列后的行文本数据。

18.一种服务器，其特征在于，所述服务器包括：处理器和存储器，所述存储器中存储有至少一条指令，至少一段程序、代码集或指令集，所述至少一条指令，至少一段程序、代码集或指令集由所述处理器加载并执行以实现如权利要求1至11任一所述的表格内容的自动分列方法。

19.一种终端，其特征在于，所述终端包括：处理器和存储器，所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如权利要求12至15任一所述的表格内容的自动分列方法。

20.一种计算机可读存储介质，其特征在于，所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由处理器加载并执行以实现如权利要求1至11任一所述的表格内容的自动分列方法，或者，如权利要求12至15任一所述的表格内容的自动分列方法。

21.一种计算机系统，其特征在于，所述系统包括如权利要求18所述的服务器，以及如权利要求19所述的终端。

表格内容的自动分列方法、装置、设备及存储介质

技术领域

[0001] 本申请实施例涉及办公软件领域，特别涉及一种表格内容的自动分列方法、装置、设备及存储介质。

背景技术

[0002] 表格处理软件是办公软件中使用最为广泛的软件之一。表格处理软件提供有分列功能。

[0003] 分列功能是将表格中选定的单列文本数据进行指定规则的列拆分的功能。分列功能的一个示例性应用场景为：多个用户在同一个聊天群里提供各自的一行数据，由一个用户将多行数据汇总至一个表格的同一列中，然后使用分列功能将该列数据分列为多列数据。比如，第一列的第一个单元格填写有“张三男21岁”，第一列的第二个单元格填写有“李四女22岁”，第一列的第三个单元格填写有“王五男30岁”。当用户设置采用空格作为分列规则时，表格处理软件将上述一列数据按照空格分列为三列，如下表一所示：

[0004] 表一

[0005]

张三	男	21岁
李四	女	22岁
王五	男	30岁

[0006] 上述分列功能需要每行数据均使用相同的分列规则，但由于多个用户在聊天群中提供数据时，存在一些用户所使用的分列符是不同的，导致用户在汇总后还需要手动调整多次，才能得到较为精确的分列结果。

发明内容

[0007] 本申请实施例提供了一种表格内容的自动分列方法、装置、设备及存储介质，可以解决相关技术中存在一些用户所使用的分列符是不同的，导致用户在汇总后还需要手动调整多次，才能得到较为精确的分列结果的问题。所述技术方案如下：

[0008] 根据本申请的一个方面，提供了一种表格内容的自动分列方法，所述方法包括：

[0009] 接收客户端发送的第一表格内容，所述第一表格内容包括待分列的行文本数据；

[0010] 对所述第一表格内容中的行文本数据进行信息提取，得到所述行文本数据中的信息标签；

[0011] 根据所述信息标签对所述行文本数据进行分列，得到第二表格内容，所述第二表格内容包括分列后的行文本数据；

[0012] 向所述客户端发送所述第二表格内容。

[0013] 根据本申请的另一方面，提供了一种表格内容的自动分列方法，所述方法包括：

[0014] 显示在线表格界面；

[0015] 根据操作信号确定所述在线表格界面中的第一表格内容，将所述第一表格内容发送至服务器；

[0016] 接收所述服务器发送的第二表格内容，所述第二表格内容是所述服务器根据信息标签对所述第一表格内容中的行文本数据进行分列后得到的，所述信息标签是所述服务器对所述第一表格内容中的行文本数据进行信息提取得到的；

[0017] 在所述在线表格界面中显示所述第二表格内容，所述第二表格内容包括分列后的行文本数据。

[0018] 根据本申请的另一方面，提供了一种表格内容的自动分列装置，所述装置包括：

[0019] 接收模块，用于接收客户端发送的第一表格内容，所述第一表格内容包括待分列的行文本数据；

[0020] 提取模块，用于对所述第一表格内容中的行文本数据进行信息提取，得到所述行文本数据中的信息标签；

[0021] 分列模块，用于根据所述信息标签对所述行文本数据进行分列，得到第二表格内容，所述第二表格内容包括分列后的行文本数据；

[0022] 发送模块，用于向所述客户端发送所述第二表格内容。

[0023] 根据本申请的另一方面，提供了一种表格内容的自动分列装置，所述装置包括：

[0024] 显示模块，用于显示在线表格界面；

[0025] 确定模块，用于根据操作信号确定所述在线表格界面中的第一表格内容，将所述第一表格内容发送至服务器；

[0026] 接收模块，用于接收所述服务器发送的第二表格内容，所述第二表格内容是所述服务器根据信息标签对所述第一表格内容中的行文本数据进行分列后得到的，所述信息标签是所述服务器对所述第一表格内容中的行文本数据进行信息提取得到的；

[0027] 所述显示模块，用于在所述在线表格界面中显示所述第二表格内容，所述第二表格内容包括分列后的行文本数据。

[0028] 根据本申请的另一方面，提供了一种服务器，所述服务器包括：处理器和存储器，所述存储器中存储有至少一条指令，至少一段程序、代码集或指令集，所述至少一条指令，至少一段程序、代码集或指令集由所述处理器加载并执行以实现如上方面所述的表格内容的自动分列方法。

[0029] 根据本申请的另一方面，提供了一种终端，所述终端包括：处理器和存储器，所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如上方面所述的表格内容的自动分列方法。

[0030] 另一方面，提供了一种计算机可读存储介质，所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如上方面所述的表格内容的自动分列方法。

[0031] 另一方面，提供了一种计算机程序产品，当所述计算机程序产品在计算机设备上运行时，使得计算机设备执行时实现如上方面所述的表格内容的自动分列方法。

[0032] 本申请实施例通过服务器对第一表格内容中的行文本数据进行信息提取，得到行文本数据中的信息标签；根据信息标签对行文本数据进行分列，得到第二表格内容；能够将属于不同信息标签的实体信息分列至不同的列中，不需要依赖于简单的分隔符作为分列条件，而是利用实体信息的语义或特征来进行分列，从而提高了分列功能的准确性和成功率。

附图说明

[0033] 为了更清楚地说明本申请实施例中的技术方案，下面将对实施例描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0034] 图1是相关技术中的分列功能的实施示意图；

[0035] 图2是相关技术中的分列功能的实施示意图；

[0036] 图3是本申请示例性实施例的计算机系统的结构框图；

[0037] 图4是本申请一个示例性实施例的表格内容的自动分列方法的流程图；

[0038] 图5是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0039] 图6是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0040] 图7是本申请另一个示例性实施例的表格内容的自动分列方法的实施示意图；

[0041] 图8是本申请另一个示例性实施例的表格内容的自动分列方法的实施示意图；

[0042] 图9是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0043] 图10是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0044] 图11是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0045] 图12是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0046] 图13是本申请另一个示例性实施例的对齐过程的示意图；

[0047] 图14是本申请另一个示例性实施例的对齐过程的示意图；

[0048] 图15是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0049] 图16是本申请另一个示例性实施例的对齐过程的示意图；

[0050] 图17是本申请另一个示例性实施例的对齐过程的示意图；

[0051] 图18是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0052] 图19是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0053] 图20是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0054] 图21是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0055] 图22是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0056] 图23是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0057] 图24是本申请另一个示例性实施例的表格内容的自动分列方法的流程图；

[0058] 图25是本申请另一个示例性实施例的表格内容的自动分列装置的框图；

[0059] 图26是本申请另一个示例性实施例的表格内容的自动分列装置的框图；

[0060] 图27是本申请一个示例性实施例提供的服务器的框图；

[0061] 图28是本申请一个示例性实施例提供的终端的框图。

具体实施方式

[0062] 为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。

[0063] 将一段不规整的文本数据粘贴或导入Excel表格中时，如“姓名”、“电话号码”等有效信息，手动分类填写到对应的单个单元格中会花费用户大量的时间。传统的表格处理程

序中提供了分列功能。该分列功能是将表格中选定的单列文本数据进行指定规则的列拆分的功能。相关技术中在不同信息类型之间设置分隔符、空格、逗号或分割线等方式对文本内容进行分列。以下以空格以及分割线为例来说明。

[0064] 1)以空格作为分列条件；

[0065] 由于原始的文本数据会存在大量参差不齐的分隔符，使用单一的规则来实现分列无法使数据按理想的条件来进行数据拆分。有些文本应该被拆分到B列，但因为使用的条件为空格，某些文本中的分割符未使用空格，导致内容仍然留在了A列。如图1所示，用户将“制表符”和“空格”设置为分隔符，第三行中的“谢燕冰，女，2011.09.11，13800138000 2位”中使用逗号作为分隔符，导致该行文本数据无法被正常分列，“谢燕冰，女，2011.09.11，13800138000”均被分列至第一列中。

[0066] 2)将分割线作为分列条件；

[0067] 用户也可以设置分割线作为分列条件，然后在表格中手动添加分割线来进行分列。如图2所示，用户在每行文本数据的相应位置中增加分割线，然后由表格处理程序对行文本数据进行分列。但由于手工添加分割线的次数较多，与用户直接手动分列相比，并未提高分列效率。

[0068] 本实施例提供了一种表格内容的自动分列方案。该技术方案利用人工智能(Artificial Intelligence，AI)技术对表格中的文本数据进行学习识别后，利用实体抽取技术实现了一键抽取行文本数据中的实体信息，相应地将每个不同类型的实体信息归类整理到相应的列中。比如，将每行数据中的信息标签为“姓名”的实体信息，归类到姓名列中；将每行数据中的信息标签为“性别”的实体信息，归类到性别列中。

[0069] 图3示出了本申请一个示例性实施例提供的计算机系统100的框图。该计算机系统100可以是一个即时通讯系统、在线办公系统、团队协作系统中的任意一种，本申请实施例对此不加以限定。该计算机系统100包括：终端120和服务器140。

[0070] 终端120可以是手机、平板电脑、电子书阅读器、MP3(Moving Picture Experts Group Audio Layer III，动态影像专家压缩标准音频层面3)播放器、MP4(Moving Picture Experts Group Audio Layer IV，动态影像专家压缩标准音频层面4)播放器、膝上型便携计算机和台式计算机等等。终端120中安装有支持在线表格处理功能的客户端。该客户端是用于以程序形式进行表格处理的客户端、用于以Web网页形式进行表格处理的客户端、用于以web网页形式进行表格处理的小程序中的至少一种。小程序是依赖于母应用程序进行运行的程序，同一个母应用程序上可运行多个不同的小程序。

[0071] 终端120通过无线网络或有线网络与服务器140相连。

[0072] 服务器140可以包括一台服务器、多台服务器、云计算平台和虚拟化中心中的至少一种。服务器140用于为支持声音消息的应用程序提供后台服务。可选地，服务器140承担主要计算工作，终端120承担次要计算工作；或者，服务器140承担次要计算工作，终端120承担主要计算工作；或者，终端120和服务器140之间采用分布式计算架构进行协同计算。该服务器140中运行有计算机程序，该计算机程序用于实现如下方法实施例中的分列功能。

[0073] 图4示出了本申请一个示例性实施例提供的表格内容的自动分列方法的流程图。本实施例以该方法应用于图3所述的计算机系统中来举例说明。该计算机系统的终端中运行有支持在线表格处理的客户端。该方法包括：

[0074] 步骤401,客户端显示在线表格界面;

[0075] 客户端支持在线表格处理。可选地,客户端是操作系统中安装和运行的本地程序,或者该客户端是以web网页形式提供服务的网页程序(简称前端程序)。本实施例对客户端的具体形式不加以限定。

[0076] 客户端上显示有在线表格界面。该在线表格界面是用于对在线表格进行编辑的用户界面。该在线表格界面是程序界面类型,或网页界面类型。

[0077] 该在线表格界面上显示有按照行列分布的多个单元格、每个单元格的行号、每个单元格的列号、以及与表格编辑有关的多个控件。

[0078] 所述在线表格为支持多人协作的在线文档。客户端无需下载安装,打开网址即可立即开始编辑文档、新建文档或导入本地文档。不同的用户可以通过各自的账号,如即时通信账号一键登录进入在线表格界面,实现多人在线协作编辑。

[0079] 步骤402,客户端根据操作信号确定在线表格界面中的第一表格内容;

[0080] 第一表格内容包括待分列的行文本数据。待分列的行文本数据为一行或多行。可选地,第一表格内容是属于同一列的多行文本数据,同一行文本数据中包括待分列的至少两项实体信息。

[0081] 客户端可通过如下两种操作方式中的任意一种方式,确定在线表格界面中的第一表格内容。

[0082] 1、复制粘贴方式:

[0083] 用户将待分列的行文本数据粘贴至在线表格界面中的单元格中,客户端根据用户的粘贴复制操作,将至少一行文本数据存储至在线表格界面中。并且,客户端在确定行文本数据中存在待分列的至少两项实体信息时,将粘贴的表格内容确定为第一表格内容。

[0084] 2、选定区域方式:

[0085] 当在线表格界面中已存储有表格内容时,用户可以使用鼠标、键盘或触摸屏方式对在线表格界面中的目标单元格进行选定,客户端根据用户的单元格选定操作,将目标单元格中的表格内容确定为第一表格内容。

[0086] 步骤403,客户端将第一表格内容发送至服务器;

[0087] 当客户端识别出第一表格内容中存在待分列的行文本数据时,客户端将第一表格内容发送至服务器。

[0088] 可选地,在线表格界面上显示有分列功能控件,当该分列功能控件接收到触发信号时,客户端将第一表格内容发送至服务器。比如,当客户端识别出第一表格内容中存在待分列的行文本数据时,弹出提示信息框“您可能希望对数据进行自动整理,确定或取消”,当接收到用户对“确定”按钮的点击信号时,客户端将第一表格内容发送至服务器。

[0089] 可选地,客户端将帐号和第一表格内容发送至服务器中。

[0090] 步骤404,服务器接收客户端发送的第一表格内容;

[0091] 步骤405,服务器对第一表格内容中的行文本数据进行信息提取,得到行文本数据中的信息标签;

[0092] 待分列的同一个行文本数据中会包括至少两项实体信息,服务器对第一表格内容中的行文本数据进行信息提取,得到行文本数据中的信息标签。

[0093] 信息标签是用于标识实体信息的信息类别的信息。示意性地,信息标签包括:姓

名、性别、年龄、时间、手机号、银行卡号、邮政编码、数字串、地址中的至少一种，本实施例对此不加以限定。

[0094] 可选地，服务器通过AI技术对第一表格内容中的行文本数据进行信息提取，得到行文本数据中的各个实体信息的信息标签。

[0095] 步骤406，服务器根据信息标签对行文本数据进行分列，得到第二表格内容；

[0096] 当待分列的行文本数据为一行时，服务器将行文本数据中属于不同信息标签的实体信息分列至不同的单元格中，得到第二表格内容。

[0097] 当待分列的行文本数据大于一行时，服务器将同一行文本数据中属于不同信息标签的实体信息分列至不同的单元格中，将不同行文本数据中属于相同信息标签的实体信息分列至相同的单元格中，得到第二表格内容。

[0098] 步骤407，服务器向客户端发送第二表格内容；

[0099] 服务器向客户端发送第二表格内容，第二表格内容包括分列后的行文本数据。可选地，服务器根据帐号向客户端发送第二表格内容。

[0100] 步骤408，客户端接收服务器发送的第二表格内容；

[0101] 步骤409，客户端在在线表格界面中显示第二表格内容，第二表格内容包括分列后的行文本数据。

[0102] 综上所述，本实施例提供的方法，通过服务器对第一表格内容中的行文本数据进行信息提取，得到行文本数据中的信息标签；根据信息标签对行文本数据进行分列，得到第二表格内容；能够将属于不同信息标签的实体信息分列至不同的列中，不需要依赖于简单的分隔符作为分列条件，而是利用实体信息的语义或特征来进行分列，从而提高了分列功能的准确性和成功率。

[0103] 参考图5所示，服务器将第一表格内容利用AI技术进行分列的过程可分为五个阶段：

[0104] 第一阶段：文本定位。由于第一表格内容的头部和尾部可能存在冗余信息，首先从第一表格内容中定位出待分列的文本数据。

[0105] 可选地，服务器识别出待分列的文本数据的初始分列位置和末尾分列位置。将位于初始分列位置和末尾分列位置之间的文本数据片段，确定为待分列的文本数据。

[0106] 服务器可基于多维度的信息对初始分列位置和末尾分列位置进行综合判断，比如一些关键字信息(例如，接龙、报名等)，有规律的序号，明显的分行符或分隔符等等。

[0107] 第二阶段：分行处理。将待分列的文本数据，拆分为一个个行文本数据。通常，每个行文本数据包括至少两个待分列的实体信息。

[0108] 可选地，服务器基于多种分行规则进行分行，比如基于序号进行分行、基于空格进行分行、基于换行符进行分行、基于其它分割符进行分行等等。

[0109] 受限于待分列的文本数据的复杂性，在本阶段中可能会引入少量的分行错误，也即并不能保证分行结果的100％准确性。

[0110] 第三阶段：解析分列。本阶段可以包括两个子阶段：信息提取+实体信息对齐。

[0111] 在信息提取子阶段中，对于每个行文本数据，对该行文本数据进行信息提取，得到至少两个实体信息以及每个实体信息所对应的信息标签。

[0112] 在实体信息对齐子阶段中，服务器利用每个实体信息的信息标签，对不同的行文

本数据中具有相同信息标签的实体信息进行对齐，并将具有同一个信息标签的实体信息对齐至同一个列中，具有不同信息标签的实体信息对齐至不同的列中，从而实现智能分列。

[0113] 第四阶段：后迭代处理。由于第三阶段是建立在相信第二阶段的分行结果的基础上的，但第二阶段的分行结果存在一定的错误可能性。本阶段利用第三阶段中的分列结果，以尽量降低分行错误以及各个信息提取模块的错误，从而提高容错性。

[0114] 本阶段依赖的方法包括：行分割、行合并、枚举型信息推理中的至少一种。

[0115] 行分割：将已经分好的行文本数据(异常行)进行再次分割，并重新进行分列处理，以便获得更好的分列效果。

[0116] 行合并：将已经分好的多个行文本数据(异常行)进行合并后再次分割，并重新进行分列处理，以便获得更好的分列效果。

[0117] 枚举型信息推理：利用枚举型信息列的实体信息，对相邻列的实体信息进行调整。枚举型信息列是指该列中的实体信息的取值是可枚举的有限数量的取值，比如性别列只包括“男”和“女”这两种枚举型信息，当性别列存在多余的汉字时，该多余的汉字通常为相邻行的实体信息。

[0118] 第五阶段：输出表格。

[0119] 当不存在用户自定义的表头时，由服务器根据每列实体信息所共同具有的信息标签，对每列单元格的表头进行命名。比如，当信息标签为姓名时，将该列实体信息命名为姓名；当信息标签为性别时，将该列实体信息命名为性别。

[0120] 下面采用一个示意性的实施例来对上述过程进行举例说明。图6示出了本申请一个示例性实施例提供的表格内容的自动分列方法的流程图。本实施例以该方法应用于图3所示的服务器中来举例说明。该方法包括：

[0121] 步骤601，接收客户端发送的第一表格内容；

[0122] 第一表格内容包括待分列的行文本数据。可选地，该第一表格内容是以文本形式表示的文本数据。

[0123] 步骤602，提取第一表格内容中的多个行文本数据；

[0124] 服务器识别以文本形式表示的文本数据的初始分列位置和末尾分列位置。将位于初始分列位置和末尾分列位置之间的文本数据片段进行分行处理，得到多个行文本数据。

[0125] 示意性的参考图7，服务器根据关键字“接龙：”识别出初始分列位置71，并将第一表格内容的尾部识别为末尾分列位置72，将位于初始分列位置71和末尾分列位置72的文本数据片段，确定为待分列的行文本数据。然后将待分列的行文本数据，按照序号拆分为多个行文本数据，如图8所示。

[0126] 需要说明的是，图7和图8中的姓名和书名均为示意性的名称，并不指代真实世界中的用户和书籍。

[0127] 步骤603，对于多个行文本数据中的任一行文本数据，将行文本数据输入信息提取模型中，得到信息提取模型所提取到的实体信息；

[0128] 服务器中设置有多个信息提取模型，每个信息提取模型用于提取一种类型的实体信息，该类型可以采用信息标签进行表示。

[0129] 根据标签类型的不同，信息提取模型包括但不限于：用于提取姓名的模型一、用于提取性别的模型二、用于提取年龄的模型三、用于提取序号的模型四、用于提取亲属关系的

模型五、用于提取时间的模型六、用于提取手机号码的模型七、用于提取身份证号码的模型八中的至少一个。

[0130] 根据提取原理的不同，信息提取模型包括但不限于：采用机器学习特征进行信息提取的第一信息提取模型、采用枚举词库进行信息提取的第二信息提取模型、采用正则表达式进行信息提取的第三信息提取模型中的至少一个。

[0131] 对于多个行文本数据中的任一行文本数据，服务器将该行文本数据分别输入各个信息提取模型中，得到信息提取模型所提取到的实体信息。

[0132] 比如，对于行文本数据“刘子瑞多了一本英语2课一试单元测试卷，少了经典诵读”，分别输入模型一至八中，对实体信息“刘子瑞”、“多了一本英语2课一试单元测试卷”、“少了经典诵读”进行提取。

[0133] 步骤604，将行文本数据中的实体信息，标注与信息提取模型对应的信息标签；

[0134] 由于行文本数据中的各个实体信息，均为采用信息提取模型所提取出的信息，而每个信息提取模型均对应各自的信息标签。因此，对于行文本数据中的任一实体信息，采用提取出该实体信息的实体信息模型所对应的信息标签进行标注。

[0135] 步骤605，将各个行文本数据中具有相同信息标签的实体信息对齐至同一列，将具有不同信息标签的实体信息对齐至不同列，得到第二表格内容；

[0136] 对于每个行文本数据，服务器均会识别出该行文本数据中的各个实体信息，以及每个实体信息对应的信息标签。

[0137] 根据每个行文本数据中的各个实体信息的信息标签，服务器将具有相同信息标签的实体信息对齐至同一列中，将具有不同信息标签的实体信息对齐至不同列，从而得到第二表格内容。

[0138] 步骤606，根据每个列对应的信息标签，生成第二表格内容的表头名称；

[0139] 当第二表格内容不存在表头名称时，由于同一列中的实体信息是具有相同信息标签的，服务器根据每个列对应的信息标签，生成第二表格内容中每个列的表头名称。

[0140] 当第二表格内容存在用户自定义的表头名称时，可沿用用户自定义的表头名称，忽略执行本步骤。

[0141] 步骤607，向客户端发送第二表格内容。

[0142] 服务器将第二表格内容发送给客户端，以便客户端在在线表格界面上显示第二表格内容，第二表格内容包括分列后的行文本数据。

[0143] 针对第一阶段/第二阶段：文本定位阶段/分行处理阶段，可参考如下实施例：

[0144] 在一个可选的实施例中，第一表格内容是以文本形式存储的文本数据，上述步骤602可包括如下子步骤602a至602d，如图9所示：

[0145] 步骤602a，在文本数据中识别段首特征，段首特征包括语义关键字、序号关键字、分隔符中的至少一种；

[0146] 服务器按照优先级顺序在第一表格内容对应的文本数据中识别段首特征。该段首特征包括语义关键字、序号关键字、分隔符中的至少一种。在一个可选的实施例中，语义关键字的优先级＞序号关键字的优先级＞分隔符的优先级，符号＞表示大于。

[0147] 服务器先在第一表格内容对应的文本数据中识别语义关键字，当识别出语义关键字时，将语义关键字作为段首特征。语义关键字包括但不限于：接龙、报名、团购、统计中的

至少一种。本实施例对语义关键字的具体形式不加以限定。

[0148] 当未识别出语义关键字时，服务器在第一表格内容对应的文本数据中识别序号关键字，当识别出序号关键字时，将序号关键字作为段首特征。序号关键字包括但不限于：一、1、①、⑴、I中的至少一种。本实施例对序号关键字的具体形式不加以限定。

[0149] 当未识别出序号关键字时，服务器在第一表格内容对应的文本数据中识别首次出现的分隔符，将首次出现的分隔符确定为段首特征。分隔符可以是空格、竖线分割符、横线分割符中的至少一种。本实施例对分隔符的具体形式不加以限定。

[0150] 当未识别出分隔符时，将第一表格内容对应的文本数据的段首位置，直接确定为初始分列位置。

[0151] 步骤602b，将段首特征所在位置的前一个位置或后一个位置识别为初始分列位置；

[0152] 当段首特征是语义关键字或分隔符时，将段首特征所在位置的后一个位置识别为初始分列位置。比如，当段首特征是语义关键字“接龙：”时，将“接龙：”的后一个位置识别为初始分列位置。

[0153] 当段首特征是序号关键字时，将段首特征所在位置的前一个位置识别为初始分列位置。比如，当段首特征是序号关键字“1”时，将“1”的前一个位置识别为初始分列位置。

[0154] 步骤602c，将文本数据的最后一个位置识别为末尾分列位置；

[0155] 结合图7可知，服务器将初始分列位置71和末尾分列位置72之间的文本数据片段，识别为待分行的文本数据。

[0156] 步骤602d，对位于初始分列位置和末尾分列位置之间的文本数据片段，按照优先级顺序依次采用至少一种分行规则进行分行处理，并在分行处理成功时得到多个行文本数据；

[0157] 服务器将位于初始分列位置和末尾分列位置之间的文本数据片段，按照优先级顺序依次尝试采用至少一种分行规则进行分行处理，并在分行处理成功时得到多个行文本数据。其中，分行规则包括：序号分行规则、空格分行规则、分割符分行规则中的至少一种。

[0158] 序号分行规则包括：当待分行的文本数据中存在多个数值连续的序号时，根据每个序号的序号位置对文本数据进行分行，得到多个行文本数据的规则。

[0159] 空格分行规则包括：当待分行的文本数据中存在多个空格时，根据每个空格(连续出现的空格可视为一个空格)的出现位置对文本数据进行分行，得到多个行文本数据的规则。

[0160] 分割符分行规则包括：当待分行的文本数据中存在多个分割符时，根据每个分割符(连续出现的分割符可视为一个分割符)的出现位置对文本数据进行分行，得到多个行文本数据的规则。

[0161] 服务器先使用序号分行规则对文本数据进行分行处理，当分行处理成功时得到多个行文本数据；当分行失败时，服务器使用空格分行规则对文本数据进行分行处理，当分行处理成功时得到多个行文本数据；当分行失败时，服务器使用分割符分行规则对文本数据进行分行处理。

[0162] 综上所述，本实施例提供的方法，通过利用优先级顺序采用不同的识别方式识别段首特征，能够提高识别不同文本内容的初始分列位置时的兼容性和准确率。即便在不同

的使用场景下，用户所采用的段首特征不同时，服务器也能较为准确地识别出初始分列位置。

[0163] 本实施例提供的方法，还通过利用优先级顺序采用不同的分行规则对文本数据进行分行处理，能够提高不同分行场景下进行分行处理时的兼容性和准确率。即便在不同的使用场景下，用户所采用的分行符号不同时，服务器也能较为准确地拆分出不同的行文本数据。

[0164] 针对第三阶段：解析分列。该解析分列阶段可包括：实体信息提取子阶段和实体信息对齐子阶段。

[0165] 针对实体信息提取子阶段，可参考如下实施例：

[0166] 在一个可选的实施例中，信息提取模型包括如下模型中的至少一种：采用机器学习特征进行信息提取的第一信息提取模型、采用枚举词库进行信息提取的第二信息提取模型、采用正则表达式进行信息提取的第三信息提取模型。本实施例以同时包括这三种模型为例，步骤603可包括如下子步骤603a至603d，如图10所示：

[0167] 步骤603a，将行文本数据输入第一信息提取模型中，预测出文字特征符合机器学习特征的第一文字串；

[0168] 可选地，第一信息提取模型是机器学习模型，该机器学习模型包括不限于：命名实体识别模型、条件随机场(Conditional Random Fields)模型、隐马尔科夫模型、深度学习模型中的至少一种。

[0169] 可选地，第一信息提取模型是预先通过训练样本进行训练得到模型，训练样本包括：人工标注的样本实体信息以及样本信息标签。

[0170] 步骤603b，当预测出文字特征符合机器学习特征的第一文字串时，将第一文字串确定为提取到的实体信息；

[0171] 以时间识别为例，由于时间的格式非常多样，比如：明天下午三点、预定28号的房间、预定今天到30号的房间。本申请实施例可采用命名实体识别模型对时间信息进行识别，得到具有时间信息标签的第一实体信息。

[0172] 以姓名识别为例，可采用序列标注模型对姓名进行标注。图11示出了该序列标注模型的原理示意图，对于给定的观察序列“我喜欢刘德华的歌曲”，该序列标注模型会利用概率矩阵预测出隐藏序列“000BIE000”。也即，“刘德华”为姓名。

[0173] 其中，B代表姓名首字、I代表姓名中字、E代表姓名尾字、0代表其它字。概率矩阵用于表示从观察序列中的一个状态跳转到另一个状态的概率，比如：B→I的概率大于B→0的概率。

[0174] 步骤603c，将行文本数据输入第二信息提取模型中，确定是否存在与枚举词库匹配的第二文字串；

[0175] 某些实体信息的信息类型是可以枚举出的有限数量的信息。比如性别只包括男和女两个取值；又比如亲属关系包括父亲、母亲、奶奶、爷爷、外婆、外公等可枚举完毕的信息范围。

[0176] 对于这类实体信息，可预先构建与该实体信息对应的枚举词库，该枚举词库包括枚举出的文字串。服务器将行文本数据输入第二信息提取模型中后，确定该行文本数据中是否与枚举词库中的任一文字串所匹配。

[0177] 步骤603d，当存在与枚举词库匹配的第二文字串时，将第二文字串确定为提取到的实体信息；

[0178] 以亲属关系为例，当存在与枚举词库匹配的第二文字串“爷爷”时，将行文本数据中的第二文字串“爷爷”确定为提取到的实体信息。

[0179] 步骤603e，将行文本数据输入第三信息提取模型中，确定是否存在与正则表达式匹配的第三文字串；

[0180] 某些实体信息的排列规律存在较强的规律性，比如手机号、邮政编码、银行卡号，这些实体信息可以利用正则表达式来表达这些实体信息的排列规则。服务器中存储有基于正则表达式构建的第三信息提取模型，在将行文本数据输入第三信息提取模型中时，确定是否存在与正则表达式匹配的第三文本串。

[0181] 以手机号码识别为例，可利用正则表达式(?:(?<!\\d\\w)(1[0-9]{5}[0-9xX]{5})(?!\\d\\w))来识别手机号码。

[0182] 步骤603f，当存在与正则表达式匹配的第三文字串时，将第三文字串确定为提取到的实体信息。

[0183] 当行文本数据中存在与正则表达式匹配的第三文字串时，将第三文字串确定为提取到的实体信息。

[0184] 需要说明的是，同一信息标签的实体信息可采用至少两种不同的实体信息模型进行提取，比如姓名类型的实体信息可以采用第一信息提取模型和第二信息提取模型进行结合提取，本实施例对此不加以限定。

[0185] 综上所述，本实施例提供的方法，通过利用不同的信息提取模型对行文本数据中的实体信息进行提取，可以综合利用实体信息的机器特征、枚举特性、排列规律中的至少一种特征进行信息提取，从而提高实体信息的提取成功率，以及不同实体信息的提取兼容性。

[0186] 针对实体信息对齐子阶段，可参考如下实施例：

[0187] 在一个可选的实施例中，针对步骤605，服务器可采用如下算法将各个行文本数据中具有相同的信息标签的实体对齐至同一列，将具有不同的信息标签的实体信息标签对齐至不同列，得到第二表格内容。步骤605可包括如下子步骤6051至6059，如图12所示：

[0188] 步骤6051，获取n个行文本数据中的实体信息对，实体信息对包括实体信息和与实体信息对应的信息标签，n为正整数；

[0189] 在服务器对每个行文本数据进行实体信息提取后，得到每个行文本数据中的实体信息对。同一个行文本数据中的实体信息对，可按照实体信息的出现顺序进行排序。

[0190] 可选地，每个行文本信息的提取结果，可表示为：$r_i=[(t_1,c_1),\cdots,(t_m,c_m)]$，$r_i$代表同一个行文本数据，$t_i$代表该行文本数据中的第i个实体信息，$c_i$代表第i个实体信息的信息标签；i为不大于m的正整数。每个括号代表一个实体信息对。不同实体信息对的排列顺序，按照实体信息在行文本数据中的出现顺序进行排序。

[0191] 步骤6052，将第i个行文本数据中的实体信息对添加到第i个栈中，行文本数据和栈一一对应，i为不大于n的正整数；

[0192] 服务器初始化n个栈，n个行文本数据与n个栈之间存在一一对应关系。

[0193] 服务器将第i个行文本数据中的每个实体信息对作为一个栈元素，添加到第i个栈中。可选地，服务器按照实体信息对的出现顺序，将最后出现的实体信息对添加至栈尾，将

最早出现的实体信息对添加至栈首。

[0194] 同一个栈中包括多个栈元素，位于栈首的栈元素称为栈首元素，位于栈尾的栈元素称为栈尾元素，位于栈首元素的下一个栈元素称为次栈首元素。

[0195] 基于相同的处理方式，将n个行文本数据中的各个实体信息对，添加至各自对应的栈中，得到n个栈。

[0196] 可选地，服务器通过补齐栈尾元素(空白栈元素)的方式，保证n个栈中的实体信息对的个数相同。也即，当n个行文本数据中的实体信息对的个数不同时，确定各个栈中的实体信息对的最多个数，将不足最多个数的栈的栈尾中补齐空白栈尾元素，使得每个栈中的栈元素均为最多个数。

[0197] 步骤6053，统计各个栈中的栈首元素中出现次数最多的第一参考信息标签，以及次栈首元素中出现次数最多的第二参考信息标签；

[0198] 对于每个栈的栈首元素和次栈首元素，统计栈首元素中出现次数最多的第一参考信息标签，比如第一参考信息标签为“姓名”；统计次栈首元素中出现次数最多的第二参考信息标签，比如第二参考信息标签为“性别”。

[0199] 次栈首元素是位于栈首元素之后的下一个栈元素。当栈首为上且栈尾为下时，次栈首元素是位于栈首元素下方的一个栈元素；当栈首为左且栈尾为右时，次栈首元素是位于栈首元素右侧的一个栈元素。

[0200] 结合参考图13，各个栈中的栈元素为四个，栈首元素中出现次数最多的第一参考信息标签为“姓名”，次栈首元素中出现次数最多的第二参考信息标签为“年龄”。

[0201] 步骤6054，检测当前栈的栈首元素的信息标签是否与第一参考信息标签相同；

[0202] 当前栈是n个栈中的任意一个栈。本实施例以当前栈是第j个栈来举例说明。

[0203] 当与第一参考信息标签相同时，进入步骤6055；当与第一参考信息标签不同时，进入步骤6056。

[0204] 步骤6055，当与第一参考信息标签相同时，将下一个栈确定为当前栈；

[0205] 若当前栈不是最后一个栈，则对下一个栈中的栈首元素重复该检测；若当前栈是最后一个栈，则进入步骤6056；

[0206] 步骤6056，当与第一参考信息标签不同时，检测是否与第二参考信息标签相同。

[0207] 当存在第j个栈中的栈首元素的信息标签与第一参考信息标签不同，但与第二参考信息标签不同时，进入步骤6057；

[0208] 当存在第j个栈中的栈首元素的信息标签与第一参考信息标签不同，但与第二参考信息标签相同时，进入步骤6058。

[0209] 步骤6057，当第j个栈中的栈首元素的信息标签与第一参考信息标签不同，但与第二参考信息标签相同时，将第j个栈中的栈元素向栈尾方向移动一位，并使用空白栈元素补齐第j个栈中的栈首元素；

[0210] 此时，第j个栈中的栈首元素将会被移动为次栈首元素。

[0211] 当栈首为上且栈尾为下时，服务器将第j个栈中的所有栈元素均下移一位，此时栈首元素变为次栈首元素。

[0212] 当栈首为左且栈尾为右时，服务器将第j个栈中的所有栈元素均右移一位，此时栈首元素变为次栈首元素。

[0213] 可选地，服务器还在原栈首元素位置，采用空白栈元素进行补齐。j为不大于n的正整数。

[0214] 结合参考图13，第3个栈中的栈首元素为(19，年龄)，栈首元素的信息标签不同于第一参考信息标签“姓名”，但相同于第二参考信息标签“年龄”，将第3个栈中所有栈元素进行右移，并在第3个栈的栈首位置补齐空白栈元素。

[0215] 步骤6058，当第j个栈中的栈首元素的信息标签与第一参考信息标签不同，但与第二参考信息标签不同时，将第j个栈中的栈元素向栈首方向移动一位，并将移动后的第一个栈元素设置为新增栈元素，新增栈元素是位于栈首元素之上的元素；

[0216] 当栈首为上且栈尾为下时，服务器将第j个栈的所有栈元素上移一位，此时原栈首元素变为位于当前栈首元素之上的新增栈元素，然后使用空白栈元素补齐空出的栈元素位置。

[0217] 当栈首为左且栈尾为右时，服务器将第j个栈的所有栈元素左移一位，此时原栈首元素变为位于当前栈首元素左侧的新增栈元素，然后使用空白栈元素补齐空出的栈元素位置。

[0218] 结合参考图14，第3个栈中的栈首元素为(班长，职务)，栈首元素的信息标签不同于第一参考信息标签“姓名”，且不同于第二参考信息标签“年龄”，将第3个栈中所有栈元素进行左移，并将原栈首元素(班长，职务)设置为新增栈元素，以及在第3个栈的栈尾位置补齐空白栈元素。

[0219] 步骤6059，当n个栈中的栈首元素的信息标签均为第一参考信息标签时，将n个栈中的栈首元素移出至目标表格中序号最小的同一个空白列中；

[0220] 可选地，目标表格是位于缓存中的空白表格，或者，目标表格是用户选取的表格区域。

[0221] 当每次栈首元素的信息标签均为第一参考信息标签时，均将n个栈中的栈首元素移出至目标表格中序号最小的同一个空白列中。若每个栈中的栈元素为n个，则该移出过程可能会执行n次。可选地，当n个栈中的栈首元素的信息标签均为第一参考信息标签，且存在新增栈元素时，在输出位置中将新增栈元素插入至新增列中，该新增列是序号最小的空白列和序号最大的一个非空白列之间的表格列。

[0222] 结合参考图14，将信息标签为“姓名”的栈首元素均移出至第1列中，然后将新增栈元素“职务”插入至第0列和第1列之间的新增列中。

[0223] 在一个示意性的例子中，上述对齐过程可采用如下对齐算法实现：

[0224] 输入：每行抽取结果$r_i=[(t_1,c_1),\cdots,(t_m,c_m)]$，$r_i$代表同一个行文本数据，$t_i$代表该行文本数据中的第i个实体信息，$c_i$代表第i个实体信息的信息标签；i为不大于m的正整数。

[0225] 输出：对齐表格T，表格列名H；

[0226] 1、初始化栈S和H，将每行抽取结果依次添加到栈S_i中，并且进行补齐，使得每一行在栈中元素个数相等；

[0227] 2、如果栈$S=\emptyset$，则算法结束，否则转入步骤3；

[0228] 3、从所有行栈首元素找出最常见的信息标签m_t，并且m_t≠'PAD'，如果m_t＝'NAME'，则再次对栈首所有t≠'NAME'的元素判断是否是姓名，如果是，则修改信息标签；

[0229] 4、对每一行栈首t≠m_t的元素进行左移或者右移操作，然后对S进行补齐；如果信息标签t和次栈首元素最常见的信息标签一致则右移；如果信息标签t和次栈首元素最常见的信息标签不一致则左移。

[0230] 本实施例中，以栈首为左且栈尾为右为例。

[0231] 右移是指将栈首元素移动为次栈首元素。可选地，将该栈内的所有栈元素均进行右移，并对各个栈进行空白栈补齐，使得每个栈中的元素个数相同。

[0232] 左移是指将栈首元素移动为新增栈元素，该新增栈元素位于栈首元素的左侧。比如，栈首元素的编号为0，次栈首元素的编号为1，则新增栈元素的编号为-1。

[0233] 5、如果每一行栈首元素的信息标签一致，则将所有行的栈首元素移出，添加到表格T，同时将信息标签t添加到H中，转入步骤2。

[0234] 可选地，将信息标签t作为表头添加至表格列名H中。

[0235] 综上所述，本实施例提供的方法，通过多个栈来实现对齐算法，使得具有相同信息标签的实体信息被分列至同一列，具有不同信息标签的实体信息被分列至不同列。即便存在一些信息标签是异常标签的情况下，也能够通过左移或右移操作来实现对齐，使得该对齐算法具有极高的容错性。

[0236] 针对第四阶段：后迭代处理阶段，可参考如下实施例：

[0237] 在一个可选的实施例中，上述步骤605之后，还包括如下步骤608至611，如图15所示：

[0238] 步骤608，根据各个行文本数据中的信息标签的信息标签个数，统计出信息标签个数的正常值；

[0239] 行文本数据可以是多行，服务器统计每个行文本数据中的信息标签的信息标签个数，统计出行文本数据中的信息标签个数的正常值。

[0240] 比如，第一个行文本数据的信息标签个数为3、第二个行文本数据的信息标签个数为3、第三个行文本数据的信息标签个数为6，第四个行文本数据的信息标签个数为3，则服务器可统计出信息标签个数的正常值为3。

[0241] 步骤609，当存在信息标签个数大于正常值的第一行文本数据，且第一行文本数据的相邻行的信息标签个数等于正常值时，将第一行文本数据进行重新分行，并对重新分行后的行文本数据进行重新分列；

[0242] 当存在第一行文本数据的信息标签个数大于正常值(或为正常值的倍数)，且第一行文本数据的相邻行的信息标签个数等于正常值时，表明该第一行文本数据是存在分行错误的异常行，而且存在较大可能性是将多行内容拆分至同一行文本数据中。

[0243] 服务器根据第一行文本数据中的信息标签，对第一行文本数据进行重新分行，并对再次分行后的行文本数据进行重新分列。比如，当第一行文本数据中的信息标签个数是正常值的n倍时，根据信息标签将第一行文本数据重新拆分为n个行文本数据。

[0244] 结合参考图16，第2个栈中的行文本数据对应的信息标签包括“姓名、性别、年龄、姓名、性别、年龄”时，服务器将第2个栈中的行文本数据拆分为按照“姓名、性别、年龄”和“姓名、性别、年龄”划分的两个行文本数据，然后重新分列。

[0245] 步骤610，当存在信息标签个数小于正常值的第二行文本数据，且与第二行文本数据相邻的第三行文本数据的信息标签个数大于正常值时，将第二行文本数据和第三行文本

数据进行重新分行，并对重新分行后的行文本数据进行重新分列。

[0246] 当存在第二行文本数据的信息标签个数小于正常值，且相邻的第三行文本数据的信息标签个数大于正常值时，表明第二行文本数据和第三行文本数据是存在分行错误的异常行，而且存在较大可能性是第二行文本数据和第三行文本数据的分行位置存在错误。

[0247] 服务器将第二行文本数据和第三行文本数据合并为一个合并文本数据后，根据信息标签对该合并文本数据进行重新分行，并对重新分行后的行文本数据进行重新分列。

[0248] 结合参考图17，第3个栈的行文本数据对应的信息标签包括“年龄、性别”、第2个栈的行文本数据对应的信息标签包括“姓名、年龄、性别、姓名”时，服务器将第2个栈和第3个栈的行文本数据合并后，重新拆分为按照“姓名、年龄、性别”和“姓名、年龄、性别”划分的两个行文本数据。

[0249] 步骤611，当存在信息标签个数大于正常值的最后一行文本数据，且最后一行文本数据的上一行的信息标签个数等于正常值时，在最后一行文本数据中确定干扰文字串；去除最后一行文本数据中的干扰文字串。

[0250] 当最后一行文本数据的信息标签个数大于正常值，且最后一行文本数据的上一行的信息标签个数等于正常值时，表明最后一行文本数据中存在冗余信息。

[0251] 服务器根据最后一行文本数据中的信息标签，在最后一行文本数据中确定干扰文字串，在最后一行文本数据中去除干扰文字串。

[0252] 上述过程可迭代实现多次，以获得校正后的第二表格内容。

[0253] 综上所述，本实施例提供的方法，通过统计每个行文本数据的信息标签个数的正常值(也称最常见值)，利用信息标签个数的正常值来进行行分割或行合并，使得即便在第二阶段存在分列错误时，也能够利用第三阶段的标签进行自动纠错，从而迭代出更准确的分列结果，减少因分行错误导致的分列不准确现象。

[0254] 在人机交互方面，用户存在至少两种不同的分列功能启动方式：

[0255] 第一，复制粘贴后，触发分列功能；

[0256] 第二，选定区域后，触发分列功能。

[0257] 针对第一种分列功能触发方式，参考如下实施例：

[0258] 在一个可选的实施例中，客户端根据用户的复制粘贴操作，确定出第一表格内容，步骤402可替代实现成为子步骤402a至步骤402c，如图18所示：

[0259] 步骤402a，接收粘贴信号；

[0260] 在显示在线表格界面后，用户可以从其它数据源向在线表格界面中复制文本信息。然后，用户可以在客户端内粘贴该文本信息。

[0261] 客户端接收用户的粘贴信号。该粘贴信号可以是鼠标右键菜单中的粘贴选项被点击的信号，也可以是粘贴快捷键信号，如Ctrl+V信号。

[0262] 步骤402b，根据粘贴信号，向第一单元格区域中的单元格粘贴表格内容；

[0263] 客户端根据用户的粘贴信号，将剪贴板的复制内容粘贴至第一单元格区域中的单元格中。第一单元格区域包括至少一个单元格。可选地，第一单元格区域包括位于同一列中的多个单元格。

[0264] 步骤402c，将粘贴的表格内容确定为第一表格内容。

[0265] 客户端将从剪切板粘贴的表格内容，确定为第一表格内容。

[0266] 在一个如图19所示出的示意性的例子中，步骤402c包括如下步骤：

[0267] S71，判断粘贴的表格内容是否满足智能分列条件；

[0268] 当满足智能分列条件时，进入步骤S72；当不满足智能分列条件时，进入步骤S75。

[0269] 可选地，智能分列条件包括但不限于如下条件中的至少一个：粘贴的表格内容的列数为一列、粘贴的表格内容为文本信息、粘贴的表格内容不包括图片。

[0270] S72，出现智能分列的tips(技巧)浮窗；

[0271] 当粘贴的表格内容满足智能分列条件时，客户端显示询问是否进行智能分列的询问信息。该询问信息可采用tips浮窗来显示。

[0272] 比如，客户端显示一个tips浮窗，该tips浮窗上显示有：是否对粘贴内容进行智能分列，确认或取消。其中，“确认”和“取消”为可点击的按钮控件。

[0273] S73，接收对tips浮窗上的确认按钮的触发信号；

[0274] 当客户端接收到对确认按钮的点击操作时，接收到对tips浮窗上的确认按钮的触发信号，进入步骤704；

[0275] 当客户端接收到对取消按钮的点击操作时，接收到对tips浮窗上的取消按钮的触发信号，取消tips浮窗的显示。

[0276] S74，对粘贴的表格内容确定为第一表格内容。

[0277] S75，不出现智能分列的提示信息，仅响应粘贴操作。

[0278] 综上所述，本实施例提供的方法，可以在用户对在线表格界面中进行复制粘贴时，触发智能分列功能的启动，较为适合用户从其它文本信息来源汇总数据的使用场景，能够提高在汇总数据时的分列效率。

[0279] 在一个示例性的例子中，以客户端为基于web的前端程序为例，通过复制粘贴操作来触发自动分列功能的过程。可示意性的如图20所示，该表格内容的自动分列方法包括如下步骤：

[0280] S701，前端程序接收粘贴操作；

[0281] S702，前端程序从剪切板中获取粘贴的表格内容；

[0282] S703，前端程序判断是否在空表格内粘贴；

[0283] 如果是在空表格内进行粘贴，则进入步骤704；如果是在非空表格内进行粘贴，则进入步骤715；

[0284] S704，前端程序判断粘贴的表格内容，是否从文本复制而来的？

[0285] 剪切板中存储有复制内容的数据来源，当数据来源为网页、聊天记录或邮件等来源时，认为粘贴的表格内容是从文本复制而来的。

[0286] 如果是从文本复制而来的，则进入步骤705；如果不是从文本复制而来的(比如从其它excel表格复制而来的)，则结束流程。

[0287] S705，前端程序判断粘贴的表格内容，是否符合智能分列条件？

[0288] 可选地，智能分列条件包括但不限于如下条件中的至少一个：粘贴的表格内容的列数为一列、粘贴的表格内容为文本信息、粘贴的表格内容不包括图片。

[0289] 如果符合智能分列条件，则进入步骤706；如果不符合智能分列条件，则结束流程。

[0290] S706，前端程序去除连续空白；

[0291] 当粘贴的表格内容中存在连续的空白单元格时，前端程序去除连续的空白单元

格。

[0292] 当粘贴的表格内容中存在连续的空格时，前端程序将连续的空格替换为单个空格。

[0293] S707，前端程序将粘贴的表格内容，发送到后台AI识别；

[0294] 后台AI可以是服务器，该服务器中集成有基于AI的信息提取模型。

[0295] S708，服务器判断是否能分列；

[0296] 当粘贴的表格内容能分列时，生成分列结果发送给客户端，进入步骤S709；

[0297] 当粘贴的表格内容不能分列时，结束流程。

[0298] S709，前端程序获得分列结果；

[0299] 前端程序从服务器获取分列结果。

[0300] S710，前端程序在在线表格界面上，对分列结果进行横幅提醒。

[0301] 该横幅提醒用于向用户提醒，存在粘贴的表格内容的智能分列结果。

[0302] S711，前端程序判断用户是否触发分列选项；

[0303] 当用户触发分列选项时，进入步骤712；当用户不触发分列选项时，结束流程。

[0304] S712，前端程序清扫原粘贴区域；

[0305] 前端程序将用户的初始粘贴区域清空。

[0306] S713，前端程序将分列结果按行列组织成表格数据(第二表格内容)；

[0307] S714，前端程序为第二表格内容附加样式格式；

[0308] S715，前端程序进行正常粘贴；

[0309] 若正常粘贴的内容为第二表格内容，则在线表格界面上显示有第二表格内容。

[0310] 作为步骤703的另一个分支，若用户是在非空表格内粘贴时，前端程序会先将粘贴的表格内容进行正常粘贴，并在符合智能分类条件时发送给AI后台进行智能分列。

[0311] S716，前端程序判断粘贴的表格内容，是否从文本复制而来的?

[0312] 剪切板中存储有复制内容的数据来源，当数据来源为网页、聊天记录或邮件等来源时，认为粘贴的表格内容是从文本复制而来的。

[0313] 如果是从文本复制而来的，则进入步骤717；如果不是从文本复制而来的(比如从其它excel表格复制而来的)，则结束流程。

[0314] S717，前端程序显示粘贴选项的选择性面板；

[0315] 可选地，该选择性面板上包括：仅粘贴内容、保留格式粘贴等与粘贴相关的功能选项。

[0316] 如果前端程序获得服务器发送的分列结果，则在选择性面板上增加显示智能分列选项。

[0317] S718，前端程序判断用户是否触发分列选项；

[0318] 当用户触发分列选项时，进入步骤712；当用户不触发分列选项时，结束流程。

[0319] 针对第二种分列功能触发方式，参考如下实施例：

[0320] 在另一个可选的实施例中，客户端根据用户的区域选定操作，确定出第一表格内容，步骤402可替代实现成为子步骤4021至步骤4023，如图21所示：

[0321] 步骤4021，接收区域选择信号；

[0322] 步骤4022，根据区域选择信号，选定第二单元格区域中的单元格；

[0323] 步骤4023,将第二单元格区域中的单元格渲染值,确定为第一表格内容;

[0324] 客户端将第二单元格区域中连续非空白行列区域的单元格渲染值,确定为第一表格内容。

[0325] 在一个如图23所示出的示意性的例子中,步骤4023包括如下步骤:

[0326] S81,判断当前选区是否满足智能分列条件;

[0327] 当满足智能分列条件时,进入步骤82;当不满足智能分列条件时,进入步骤83。

[0328] S82,智能分列选项高亮,或者,出现智能分列选项;

[0329] 当客户端是PC端时,将智能分列选项高亮;

[0330] 当客户端是web端时,出现智能分列选项。

[0331] S83,接收智能分列选项的触发信号;

[0332] 该触发信号可以是点击智能分列选项的信号。

[0333] S84,按照数据分布概率模型所指示的优先级顺序,搜索第二单元格区域的单元格中的连续非空白行列区域;

[0334] 由于用户选定的第二单元格区域中的单元格可能较多,比如选择了表格中的整个第一列,存在很多单元格还是空单元格。客户端可以按照数据分布概率模型所指示的优先级顺序,搜索第二单元格区域的单元格中的连续非空白行列区域;

[0335] 可选地,数据分布概率模型是根据历史数据统计出的表格内容在表格区域中不同分布位置中出现的概率模型。参考图22所示的一个示意性例子,在线表格界面中的第1-25行、第1-8列中出现连续非空白行列区域的数据概率为75%;在线表格界面中的第1-25行、第9-16行中出现连续非空白行列区域的数据概率为11%;在线表格界面中的第26-50行、第1-16列中出现连续非空白行列区域的数据概率为7%;在线表格界面中的第1-50行、第17-正无穷列中的数据概率为5%;在线表格界面中的余下行列中出现连续非空白行列区域的数据概率为2%。

[0336] S85,将搜索到的连续非空白行列区域中的单元格渲染值,确定为第一表格内容。

[0337] S86,智能分列选项置灰,或者,不出现智能分列选项;

[0338] 当客户端是PC端时,将智能分列选项置灰;

[0339] 当客户端是web端时,不出现智能分列选项。

[0340] 在一个示例性的例子中,以客户端为基于web的前端程序为例,通过选定区域操作来触发自动分列功能的过程。可示意性的如图24所示,该表格内容的自动分列方法包括如下步骤:

[0341] S801,用户选择区域和手动触发智能分列;

[0342] S802,前端程序按连续数据划分选择区域;

[0343] 前端程序根据数据分布概率模型选择区域中的连续非空白区域。

[0344] S803,前端程序获得连续非空白区域的单元格渲染值;

[0345] S804,前端程序将各个区域内的数据以行列组合成纯文本;

[0346] 示意性的,前端程序将不同行之间采用TAB符隔开后,组合成纯文本。

[0347] S805,前端程序将纯文本(的第一表格内容)发送给后台AI进行识别。

[0348] S806,后台AI判断是否能够分列;

[0349] 如果能分列,则将分列结果发送给前端程序,进入步骤807;如果不能分列,进入步

骤813，提示用户不能分列。

[0350] S807，前端程序根据分列结果将行列组装成表格数据(第二表格内容)；

[0351] S808，前端程序判断将要粘贴区域是否和未选择区域交叉；

[0352] 当将要粘贴区域和未选择区域存在相同的单元格时，认为粘贴区域和未选择区域存在交叉。

[0353] 如果交叉，则进入步骤809；如果不交叉，则进入步骤810；

[0354] S809，会覆盖到未选择区域时，前端程序判断用户是否继续操作?

[0355] 可选地，前端程序弹出询问窗口，询问用户是否继续粘贴第二表格内容。

[0356] 若用户选择继续，则进入步骤810；如果用户选择不继续，则结束流程。

[0357] S810，前端程序清除原来选择区域的内容；

[0358] S811，前端程序设置粘贴数据的样式格式；

[0359] S812，前端程序粘贴数据。

[0360] 综上所述，本实施例提供的方法，可以在用户对在线表格界面中的选择区域操作触发智能分列功能的启动，较为适合用户在线编辑表格时的使用场景，能够提高在汇总数据时的分列效率。

[0361] 在一个可选的实施例中，由于分列后的第二表格内容会占据较多的表格区域，步骤409可替代实现成为子步骤409a至步骤409c，如图18或图21所示：

[0362] 步骤409a，确定第二表格内容中的分列后的行文本数据所需占据的第三表格区域；

[0363] 步骤409b，当第三表格区域占据除第二单元格区域之外的非空单元格时，显示询问控件，询问控件用于询问是否允许覆盖非空单元格；

[0364] 步骤409c，当询问控件上接收到允许操作信号时，将分列后的行文本数据分列显示在第三表格区域中。

[0365] 综上所述，本实施例提供的方法，可以减少直接粘贴覆盖第三表格区域后，将用户的有用数据覆盖掉，从而丢失用户的有用数据的问题。能够在用户确认本次覆盖无误时，才将第二表格内容复制到第三表格区域中，减少不必要的表格数据丢失情况。

[0366] 以下为本申请的装置实施例，对于装置实施例中未详细描述的细节，可参考上述方法实施例中的详细细节。

[0367] 图25示出了本申请一个示例性实施例提供的表格内容的自动分列装置的框图，该装置可以通过软件、硬件或两者的结合实现成为服务器的全部或一部分，该装置包括：接收模块2120、提取模块2140、分列模块2160和发送模块2180。

[0368] 接收模块2120，用于接收客户端发送的第一表格内容，所述第一表格内容包括待分列的行文本数据；

[0369] 提取模块2140，用于对所述第一表格内容中的行文本数据进行信息提取，得到所述行文本数据中的信息标签；

[0370] 分类模块2160，用于根据所述信息标签对所述行文本数据进行分列，得到第二表格内容，所述第二表格内容包括分列后的行文本数据；

[0371] 发送模块2180，用于向所述客户端发送所述第二表格内容。

[0372] 在一个可选的实施例中，提取模块2140，用于提取所述第一表格内容中的多个行

文本数据;对于所述多个行文本数据中的任一行文本数据,将所述行文本数据输入信息提取模型中,得到所述信息提取模型所提取的所述实体信息;将所述行文本数据中的所述实体信息,标注与所述信息提取模型对应的信息标签;

[0373] 其中,所述信息提取模型是多个信息提取模型中的一个。

[0374] 在一个可选的实施例中,所述信息提取模型包括:采用机器学习特征进行信息提取的第一信息提取模型;

[0375] 所述提取模块2140,用于将所述行文本数据输入所述第一信息提取模型中,预测出文字特征符合所述机器学习特征的第一文字串;当预测出文字特征符合所述机器学习特征的第一文字串时,将所述第一文字串确定为提取到的实体信息。

[0376] 在一个可选的实施例中,所述信息提取模型包括:采用枚举词库进行信息提取的第二信息提取模型;

[0377] 所述提取模块2140,用于将所述行文本数据输入所述第二信息提取模型中,确定是否存在与所述枚举词库匹配的第二文字串;当存在与所述枚举词库匹配的第二文字串时,将所述第二文字串确定为提取到的所述实体信息。

[0378] 在一个可选的实施例中,所述信息提取模型包括:采用正则表达式进行信息提取的第三信息提取模型;

[0379] 所述提取模块2140,用于将所述行文本数据输入所述第三信息提取模型中,确定是否存在与所述正则表达式匹配的第三文字串;当存在与所述正则表达式匹配的第三文字串时,将所述第三文字串确定为提取到的所述实体信息。

[0380] 在一个可选的实施例中,所述第一表格内容是以文本形式存储的文本数据;

[0381] 所述提取模块2140,用于识别所述文本数据的初始分列位置和末尾分列位置;对位于所述初始分列位置和所述末尾分列位置之间的文本数据片段,按照优先级顺序依次采用至少一种分行规则进行分行处理,并在分行处理成功时得到所述多个行文本数据;其中,所述分行规则包括:序号分行规则、空格分行规则、分割符分行规则中的至少一种。

[0382] 在一个可选的实施例中,所述提取模块2140,在所述文本数据中识别段首特征,所述段首特征包括语义关键字、序号关键字、分隔符中的至少一种;

[0383] 将所述段首特征所在位置的前一个位置或后一个位置识别为所述初始分列位置;

[0384] 将所述文本数据的最后一个位置识别为所述末尾分列位置。

[0385] 在一个可选的实施例中,所述分列模块2160,用于将各个所述行文本数据中具有相同信息标签的实体信息对齐至同一列,将具有不同信息标签的实体信息对齐至不同列,得到所述第二表格内容。

[0386] 在一个可选的实施例中,所述分列模块2160,用于获取n个所述行文本数据中的实体信息对,所述实体信息对包括所述实体信息和与所述实体信息对应的信息标签,n为正整数;

[0387] 将第i个所述行文本数据中的所述实体信息对作为栈元素,添加到第i个栈中,所述行文本数据和所述栈一一对应,i为不大于n的正整数;

[0388] 统计各个栈中的栈首元素中出现次数最多的第一参考信息标签,以及次栈首元素中出现次数最多的第二参考信息标签;

[0389] 当第j个栈中的栈首元素的信息标签与所述第一参考信息标签不同,且与所述第

二参考信息标签相同时，将所述第j个栈中的栈元素向栈尾方向移动一位，并使用空白栈元素补齐所述第j个栈中的栈首元素，j为不大于n的正整数；

[0390] 当所述第j个栈中的栈首元素的信息标签与所述第一参考信息标签不同，且与所述第二参考信息标签不同时，将所述第j个栈中的栈元素向栈首方向移动一位，并将移动后的第一个栈元素设置为新增栈元素，所述新增栈元素是位于所述栈首元素之上的元素；

[0391] 当所述n个栈中的栈首元素的信息标签均为所述第一参考信息标签时，将所述n个栈中的栈首元素移出至目标表格中序号最小的同一个空白列中。

[0392] 在一个可选的实施例中，所述装置还包括：迭代模块2190，用于根据各个所述行文本数据中的所述信息标签的信息标签个数，统计出所述信息标签的正常值；

[0393] 当存在所述信息标签个数大于所述正常值的第一行文本数据，且所述第一行文本数据的相邻行的所述信息标签个数等于所述正常值时，将所述第一行文本数据进行重新分行，并对重新分行后的行文本数据进行重新分列；

[0394] 当存在所述信息标签个数小于所述正常值的第二行文本数据，且与所述第二行文本数据相邻的第三行文本数据的所述信息标签个数大于所述正常值时，将所述第二行文本数据和所述第三行文本数据进行重新分行，并对重新分行后的行文本数据进行重新分列。

[0395] 在一个可选的实施例中，所述装置还包括：迭代模块2190，用于根据各个所述行文本数据中的所述信息标签的信息标签个数，统计出所述信息标签个数的正常值；

[0396] 当存在所述信息标签个数大于所述正常值的最后一行文本数据，且所述最后一行文本数据的上一行的所述信息标签个数等于所述正常值时，在所述最后一行文本数据中确定干扰文字串；去除所述最后一行文本数据中的所述干扰文字串。

[0397] 图26示出了本申请一个示例性实施例提供的表格内容的自动分列装置的框图，该装置可以通过软件、硬件或两者的结合实现成为终端的全部或一部分，该装置包括：显示模块2220、确定模块2240、接收模块2260。

[0398] 显示模块2220，用于显示在线表格界面；

[0399] 确定模块2240，用于根据操作信号确定所述在线表格界面中的第一表格内容，将所述第一表格内容发送至服务器；

[0400] 接收模块2260，用于接收所述服务器发送的第二表格内容，所述第二表格内容是所述服务器根据信息标签对所述第一表格内容中的行文本数据进行分列后得到的，所述信息标签是所述服务器对所述第一表格内容中的行文本数据进行信息提取得到的；

[0401] 所述显示模块2220，用于在所述在线表格界面中显示所述第二表格内容，所述第二表格内容包括分列后的行文本数据。

[0402] 在一个可选的实施例中，确定模块2240，用于接收粘贴信号；根据所述粘贴信号，向第一单元格区域中的单元格粘贴表格内容；将粘贴的所述表格内容确定为所述第一表格内容。

[0403] 在一个可选的实施例中，确定模块2240，用于接收区域选择信号；根据所述区域选择信号，选定第二单元格区域中的单元格；将所述第二单元格区域中的单元格渲染值，确定为所述第一表格内容。

[0404] 在一个可选的实施例中，确定模块2240，用于按照数据分布概率模型所指示的优先级顺序，搜索所述第二单元格区域的单元格中的连续非空白行列区域；所述数据分布概

率模型是根据历史数据统计出的表格内容在表格区域中不同分布位置中出现的概率模型；将搜索到的所述连续非空白行列区域中的单元格渲染值，确定为所述第一表格内容。

[0405] 在一个可选的实施例中，显示模块2220，用于确定所述第二表格内容中的分列后的行文本数据所需占据的第三表格区域；

[0406] 当所述第三表格区域占据除所述第二单元格区域之外的非空单元格时，显示询问控件，所述询问控件用于询问是否允许覆盖所述非空单元格；

[0407] 当所述询问控件上接收到允许操作信号时，将所述分列后的行文本数据分列显示在所述第三表格区域中。

[0408] 图27示出了本申请一个实施例提供的服务器的结构示意图。该服务器用于实施上述实施例中提供的表格内容的自动分列方法。具体来讲：

[0409] 所述服务器2700包括中央处理单元(CPU)2701、包括随机存取存储器(RAM)2702和只读存储器(ROM)2703的系统存储器2704，以及连接系统存储器2704和中央处理单元2701的系统总线2705。所述服务器2700还包括帮助计算机内的各个器件之间传输信息的基本输入/输出系统(I/O系统)2706，和用于存储操作系统2713、应用程序2714和其他程序模块2715的大容量存储设备2707。

[0410] 所述基本输入/输出系统2706包括有用于显示信息的显示器2708和用于用户输入信息的诸如鼠标、键盘之类的输入设备2709。其中所述显示器2708和输入设备2709都通过连接到系统总线2705的输入输出控制器2710连接到中央处理单元2701。所述基本输入/输出系统2706还可以包括输入输出控制器2710以用于接收和处理来自键盘、鼠标、或电子触控笔等多个其他设备的输入。类似地，输入输出控制器2710还提供输出到显示屏、打印机或其他类型的输出设备。

[0411] 所述大容量存储设备2707通过连接到系统总线2705的大容量存储控制器(未示出)连接到中央处理单元2701。所述大容量存储设备2707及其相关联的计算机可读介质为服务器2700提供非易失性存储。也就是说，所述大容量存储设备2707可以包括诸如硬盘或者CD-ROM驱动器之类的计算机可读介质(未示出)。

[0412] 不失一般性，所述计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或其他数据等信息的任何方法或技术实现的易失性和非易失性、可移动和不可移动介质。计算机存储介质包括RAM、ROM、EPROM、EEPROM、闪存或其他固态存储其技术，CD-ROM、DVD或其他光学存储、磁带盒、磁带、磁盘存储或其他磁性存储设备。当然，本领域技术人员可知所述计算机存储介质不局限于上述几种。上述的系统存储器2704和大容量存储设备2707可以统称为存储器。

[0413] 根据本申请的各种实施例，所述服务器2700还可以通过诸如因特网等网络连接到网络上的远程计算机运行。也即服务器2700可以通过连接在所述系统总线2705上的网络接口单元2711连接到网络2712，或者说，也可以使用网络接口单元2711来连接到其他类型的网络或远程计算机系统(未示出)。

[0414] 所述存储器还包括一个或者一个以上的程序，所述一个或者一个以上程序存储于存储器中，且经配置以由一个或者一个以上处理器执行。上述一个或者一个以上程序包含用于实现上述表格内容的自动分列方法的计算机程序。

[0415] 图28示出了本发明一个示例性实施例提供的终端2800的结构框图。该终端2800可

以是:智能手机、平板电脑、MP3播放器(Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)、MP4(Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面4)播放器、笔记本电脑或台式电脑。终端2800还可能被称为用户设备、便携式终端、膝上型终端、台式终端等其他名称。

[0416] 通常,终端2800包括有:处理器2801和存储器2802。

[0417] 处理器2801可以包括一个或多个处理核心,比如4核心处理器、8核心处理器等。处理器2801可以采用DSP(Digital Signal Processing,数字信号处理)、FPGA(Field-Programmable Gate Array,现场可编程门阵列)、PLA(Programmable Logic Array,可编程逻辑阵列)中的至少一种硬件形式来实现。处理器2801也可以包括主处理器和协处理器,主处理器是用于对在唤醒状态下的数据进行处理的处理器,也称CPU(Central Processing Unit,中央处理器);协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。在一些实施例中,处理器2801可以在集成有GPU(Graphics Processing Unit,图像处理器),GPU用于负责显示屏所需要显示的内容的渲染和绘制。一些实施例中,处理器2801还可以包括AI(Artificial Intelligence,人工智能)处理器,该AI处理器用于处理有关机器学习的计算操作。

[0418] 存储器2802可以包括一个或多个计算机可读存储介质,该计算机可读存储介质可以是非暂态的。存储器2802还可包括高速随机存取存储器,以及非易失性存储器,比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中,存储器2802中的非暂态的计算机可读存储介质用于存储至少一个指令,该至少一个指令用于被处理器2801所执行以实现本申请中方法实施例提供的表格内容的自动分列方法。

[0419] 在一些实施例中,终端2800还可选包括有:外围设备接口2803和至少一个外围设备。处理器2801、存储器2802和外围设备接口2803之间可以通过总线或信号线相连。各个外围设备可以通过总线、信号线或电路板与外围设备接口2803相连。具体地,外围设备包括:射频电路2804、触摸显示屏2805、摄像头2806、音频电路2807、定位组件2808和电源2809中的至少一种。

[0420] 外围设备接口2803可被用于将I/O(Input/Output,输入/输出)相关的至少一个外围设备连接到处理器2801和存储器2802。在一些实施例中,处理器2801、存储器2802和外围设备接口2803被集成在同一芯片或电路板上;在一些其他实施例中,处理器2801、存储器2802和外围设备接口2803中的任意一个或两个可以在单独的芯片或电路板上实现,本实施例对此不加以限定。

[0421] 射频电路2804用于接收和发射RF(Radio Frequency,射频)信号,也称电磁信号。射频电路2804通过电磁信号与通信网络以及其他通信设备进行通信。射频电路2804将电信号转换为电磁信号进行发送,或者,将接收到的电磁信号转换为电信号。可选地,射频电路2804包括:天线系统、RF收发器、一个或多个放大器、调谐器、振荡器、数字信号处理器、编解码芯片组、用户身份模块卡等等。射频电路2804可以通过至少一种无线通信协议来与其它终端进行通信。该无线通信协议包括但不限于:万维网、城域网、内联网、各代移动通信网络(2G、3G、4G及5G)、无线局域网和/或WiFi(Wireless Fidelity,无线保真)网络。在一些实施例中,射频电路2804还可以包括NFC(Near Field Communication,近距离无线通信)有关的电路,本申请对此不加以限定。

[0422] 显示屏2805用于显示UI(User Interface,用户界面)。该UI可以包括图形、文本、图标、视频及其它们的任意组合。当显示屏2805是触摸显示屏时,显示屏2805还具有采集在显示屏2805的表面或表面上方的触摸信号的能力。该触摸信号可以作为控制信号输入至处理器2801进行处理。此时,显示屏2805还可以用于提供虚拟按钮和/或虚拟键盘,也称软按钮和/或软键盘。在一些实施例中,显示屏2805可以为一个,设置终端2800的前面板;在另一些实施例中,显示屏2805可以为至少两个,分别设置在终端2800的不同表面或呈折叠设计;在再一些实施例中,显示屏2805可以是柔性显示屏,设置在终端2800的弯曲表面上或折叠面上。甚至,显示屏2805还可以设置成非矩形的不规则图形,也即异形屏。显示屏2805可以采用LCD(Liquid Crystal Display,液晶显示屏)、OLED(Organic Light-Emitting Diode,有机发光二极管)等材质制备。

[0423] 摄像头组件2806用于采集图像或视频。可选地,摄像头组件2806包括前置摄像头和后置摄像头。通常,前置摄像头设置在终端的前面板,后置摄像头设置在终端的背面。在一些实施例中,后置摄像头为至少两个,分别为主摄像头、景深摄像头、广角摄像头、长焦摄像头中的任意一种,以实现主摄像头和景深摄像头融合实现背景虚化功能、主摄像头和广角摄像头融合实现全景拍摄以及VR(Virtual Reality,虚拟现实)拍摄功能或者其它融合拍摄功能。在一些实施例中,摄像头组件2806还可以包括闪光灯。闪光灯可以是单色温闪光灯,也可以是双色温闪光灯。双色温闪光灯是指暖光闪光灯和冷光闪光灯的组合,可以用于不同色温下的光线补偿。

[0424] 音频电路2807可以包括麦克风和扬声器。麦克风用于采集用户及环境的声波,并将声波转换为电信号输入至处理器2801进行处理,或者输入至射频电路2804以实现语音通信。出于立体声采集或降噪的目的,麦克风可以为多个,分别设置在终端2800的不同部位。麦克风还可以是阵列麦克风或全向采集型麦克风。扬声器则用于将来自处理器2801或射频电路2804的电信号转换为声波。扬声器可以是传统的薄膜扬声器,也可以是压电陶瓷扬声器。当扬声器是压电陶瓷扬声器时,不仅可以将电信号转换为人类可听见的声波,也可以将电信号转换为人类听不见的声波以进行测距等用途。在一些实施例中,音频电路2807还可以包括耳机插孔。

[0425] 定位组件2808用于定位终端2800的当前地理位置,以实现导航或LBS(Location Based Service,基于位置的服务)。定位组件2808可以是基于美国的GPS(Global Positioning System,全球定位系统)、中国的北斗系统或俄罗斯的伽利略系统的定位组件。

[0426] 电源2809用于为终端2800中的各个组件进行供电。电源2809可以是交流电、直流电、一次性电池或可充电电池。当电源2809包括可充电电池时,该可充电电池可以是有线充电电池或无线充电电池。有线充电电池是通过有线线路充电的电池,无线充电电池是通过无线线圈充电的电池。该可充电电池还可以用于支持快充技术。

[0427] 在一些实施例中,终端2800还包括有一个或多个传感器2810。该一个或多个传感器2810包括但不限于:加速度传感器2811、陀螺仪传感器2812、压力传感器2813、指纹传感器2814、光学传感器2815以及接近传感器2816。

[0428] 加速度传感器2811可以检测以终端2800建立的坐标系的三个坐标轴上的加速度大小。比如,加速度传感器2811可以用于检测重力加速度在三个坐标轴上的分量。处理器

2801可以根据加速度传感器2811采集的重力加速度信号，控制触摸显示屏2805以横向视图或纵向视图进行用户界面的显示。加速度传感器2811还可以用于游戏或者用户的运动数据的采集。

[0429] 陀螺仪传感器2812可以检测终端2800的机体方向及转动角度，陀螺仪传感器2812可以与加速度传感器2811协同采集用户对终端2800的3D动作。处理器2801根据陀螺仪传感器2812采集的数据，可以实现如下功能：动作感应(比如根据用户的倾斜操作来改变UI)、拍摄时的图像稳定、游戏控制以及惯性导航。

[0430] 压力传感器2813可以设置在终端2800的侧边框和/或触摸显示屏2805的下层。当压力传感器2813设置在终端2800的侧边框时，可以检测用户对终端2800的握持信号，由处理器2801根据压力传感器2813采集的握持信号进行左右手识别或快捷操作。当压力传感器2813设置在触摸显示屏2805的下层时，由处理器2801根据用户对触摸显示屏2805的压力操作，实现对UI界面上的可操作性控件进行控制。可操作性控件包括按钮控件、滚动条控件、图标控件、菜单控件中的至少一种。

[0431] 指纹传感器2814用于采集用户的指纹，由处理器2801根据指纹传感器2814采集到的指纹识别用户的身份，或者，由指纹传感器2814根据采集到的指纹识别用户的身份。在识别出用户的身份为可信身份时，由处理器2801授权该用户执行相关的敏感操作，该敏感操作包括解锁屏幕、查看加密信息、下载软件、支付及更改设置等。指纹传感器2814可以被设置终端2800的正面、背面或侧面。当终端2800上设置有物理按键或厂商Logo时，指纹传感器2814可以与物理按键或厂商Logo集成在一起。

[0432] 光学传感器2815用于采集环境光强度。在一个实施例中，处理器2801可以根据光学传感器2815采集的环境光强度，控制触摸显示屏2805的显示亮度。具体地，当环境光强度较高时，调高触摸显示屏2805的显示亮度；当环境光强度较低时，调低触摸显示屏2805的显示亮度。在另一个实施例中，处理器2801还可以根据光学传感器2815采集的环境光强度，动态调整摄像头组件2806的拍摄参数。

[0433] 接近传感器2816，也称距离传感器，通常设置在终端2800的前面板。接近传感器2816用于采集用户与终端2800的正面之间的距离。在一个实施例中，当接近传感器2816检测到用户与终端2800的正面之间的距离逐渐变小时，由处理器2801控制触摸显示屏2805从亮屏状态切换为息屏状态；当接近传感器2816检测到用户与终端2800的正面之间的距离逐渐变大时，由处理器2801控制触摸显示屏2805从息屏状态切换为亮屏状态。

[0434] 本领域技术人员可以理解，图28中示出的结构并不构成对终端2800的限定，可以包括比图示更多或更少的组件，或者组合某些组件，或者采用不同的组件布置。

[0435] 根据本申请实施例的另一方面，还提供了一种计算机可读存储介质，所述存储介质中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如上述方面所述的表格内容的自动分列方法。

[0436] 根据本申请实施例的另一方面，还提供了一种计算机程序产品，当所述计算机程序产品在计算机设备上运行时，使得计算机设备执行时实现如上述方面所述的表格内容的自动分列方法。

[0437] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件

来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

[0438] 以上所述仅为本申请的较佳实施例，并不用以限制本申请，凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。


学生姓名 性别 出生年月日 家长联系电话 参团人数
文本分列向导 - 第 2 步，共 3 步
您可以在此屏幕中设置您的数据包含的分隔符。
分隔符
制表符
分号
逗号
空格
其他:
连续分隔符号视为单个处理
文本限定符:
预览选定数据:
取消
< 上一步
下一步 >
完成

图1

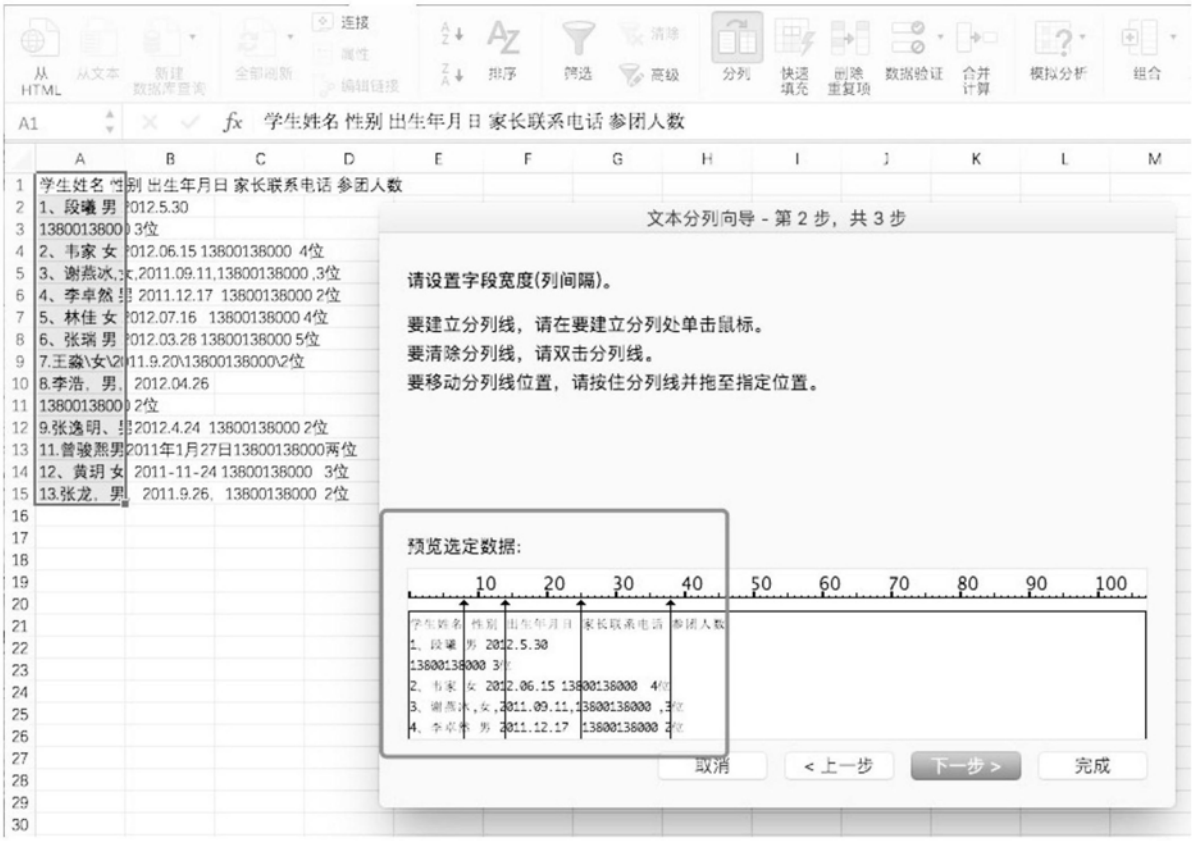

学生姓名 性别 出生年月日 家长联系电话 参团人数
文本分列向导 - 第 2 步，共 3 步
请设置字段宽度(列间隔)。
要建立分列线，请在要建立分列处单击鼠标。
要清除分列线，请双击分列线。
要移动分列线位置，请按住分列线并拖至指定位置。
预览选定数据:
取消
< 上一步
下一步 >
完成

图2

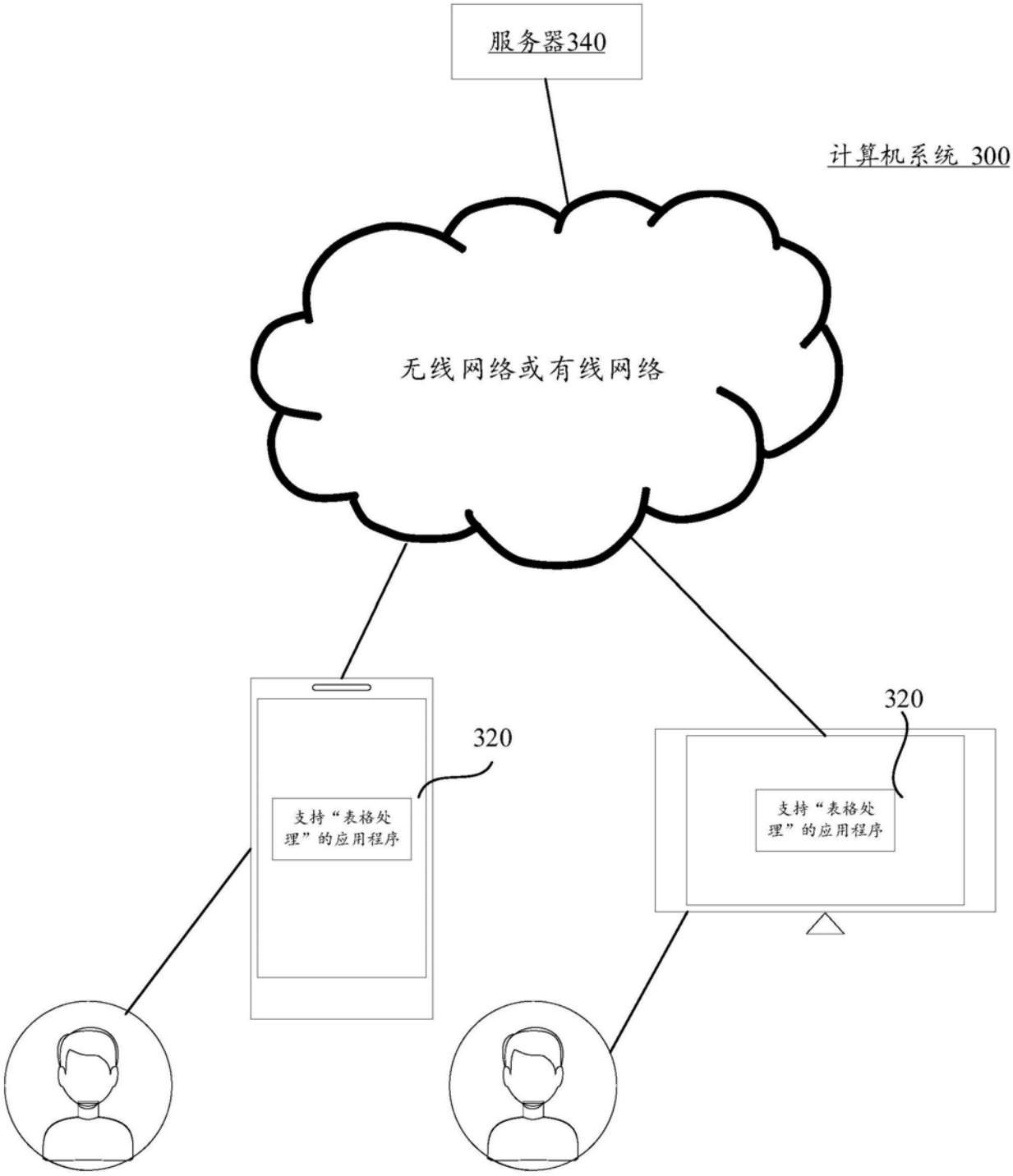

服务器340
计算机系统 300
无线网络或有线网络
320
支持“表格处理”的应用程序
320
支持“表格处理”的应用程序

图3

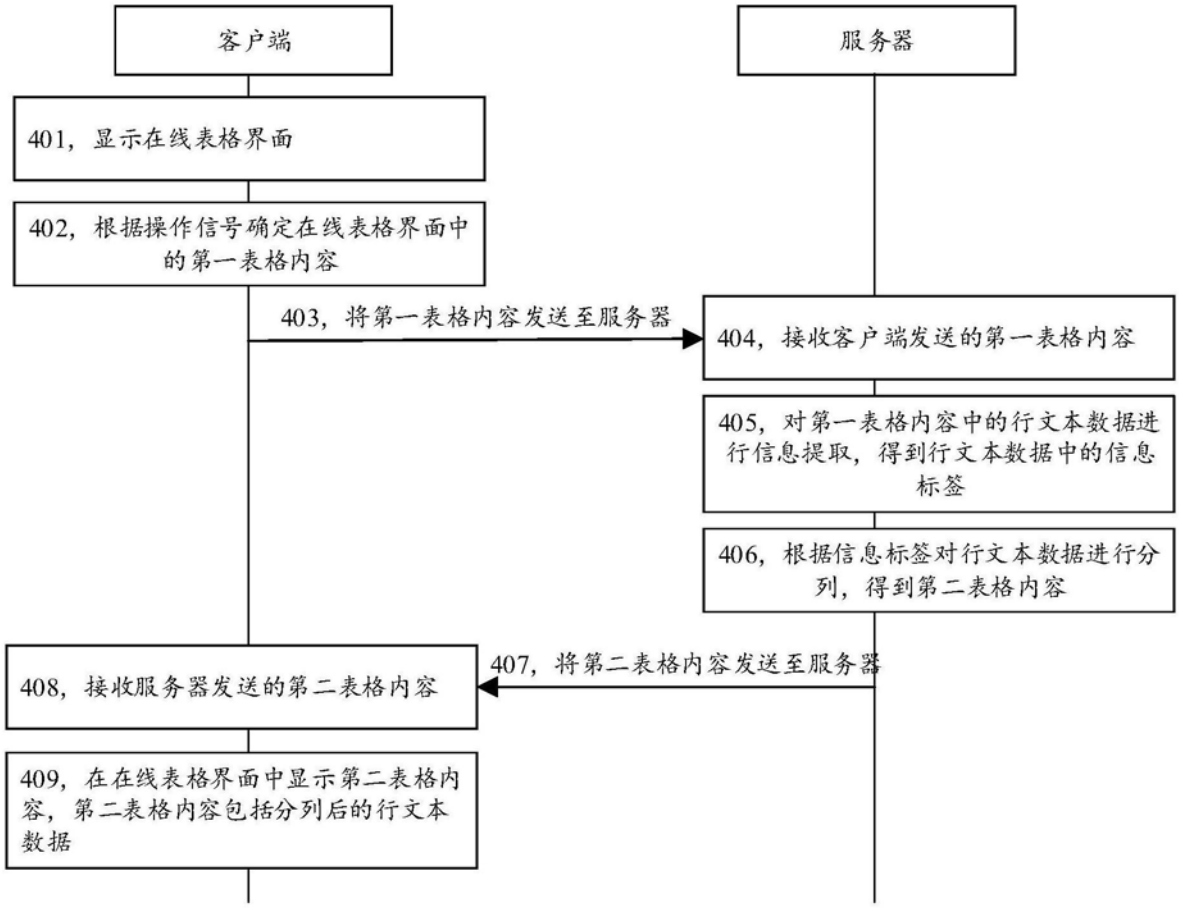

客户端
服务器
401，显示在线表格界面
402，根据操作信号确定在线表格界面中的第一表格内容
403，将第一表格内容发送至服务器
404，接收客户端发送的第一表格内容
405，对第一表格内容中的行文本数据进行信息提取，得到行文本数据中的信息标签
406，根据信息标签对行文本数据进行分列，得到第二表格内容
407，将第二表格内容发送至服务器
408，接收服务器发送的第二表格内容
409，在在线表格界面中显示第二表格内容，第二表格内容包括分列后的行文本数据

图4

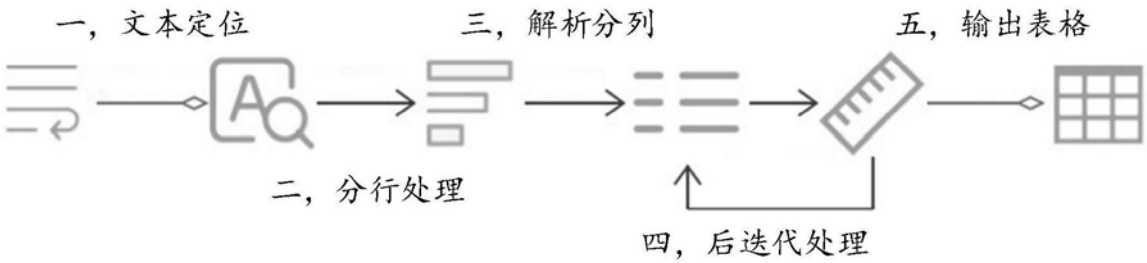

一，文本定位
二，分行处理
三，解析分列
四，后迭代处理
五，输出表格

图5

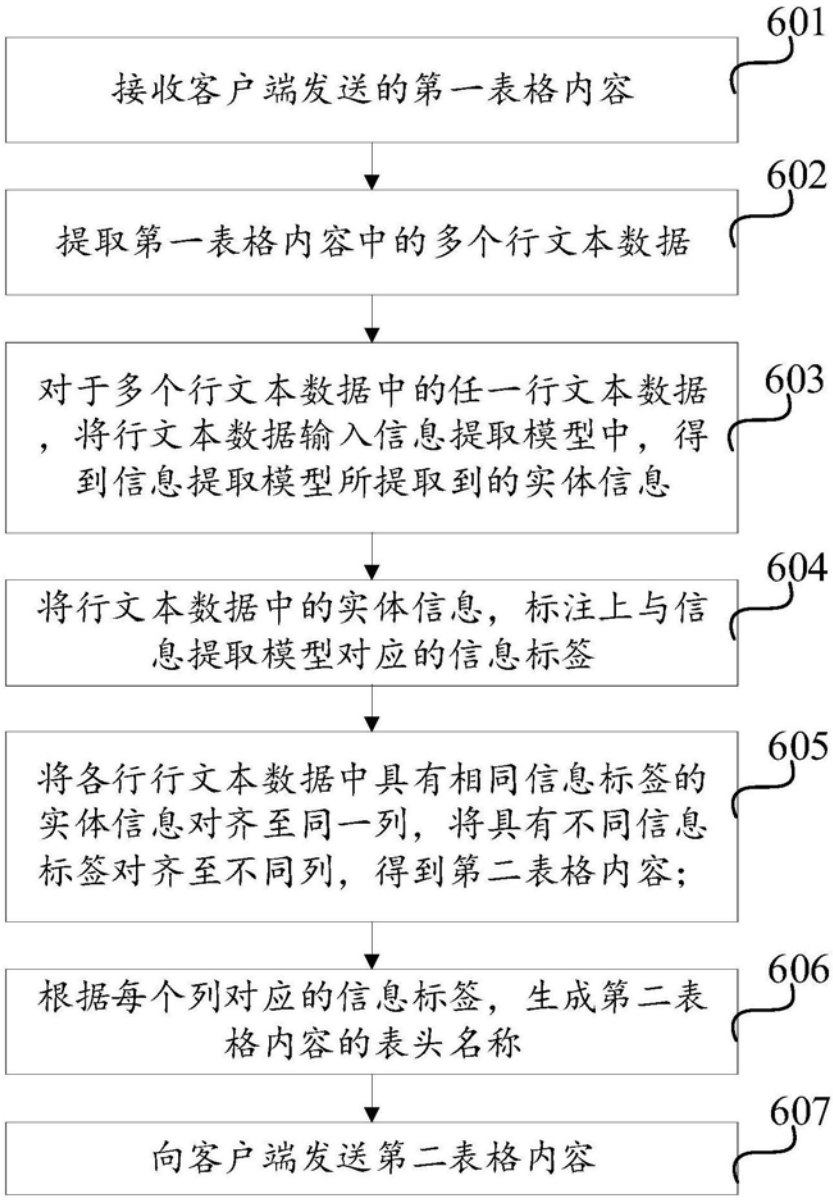

601
接收客户端发送的第一表格内容
602
提取第一表格内容中的多个行文本数据
603
对于多个行文本数据中的任一行文本数据，将行文本数据输入信息提取模型中，得到信息提取模型所提取到的实体信息
604
将行文本数据中的实体信息，标注上与信息提取模型对应的信息标签
605
将各行行文本数据中具有相同信息标签的实体信息对齐至同一列，将具有不同信息标签对齐至不同列，得到第二表格内容；
606
根据每个列对应的信息标签，生成第二表格内容的表头名称
607
向客户端发送第二表格内容

图6

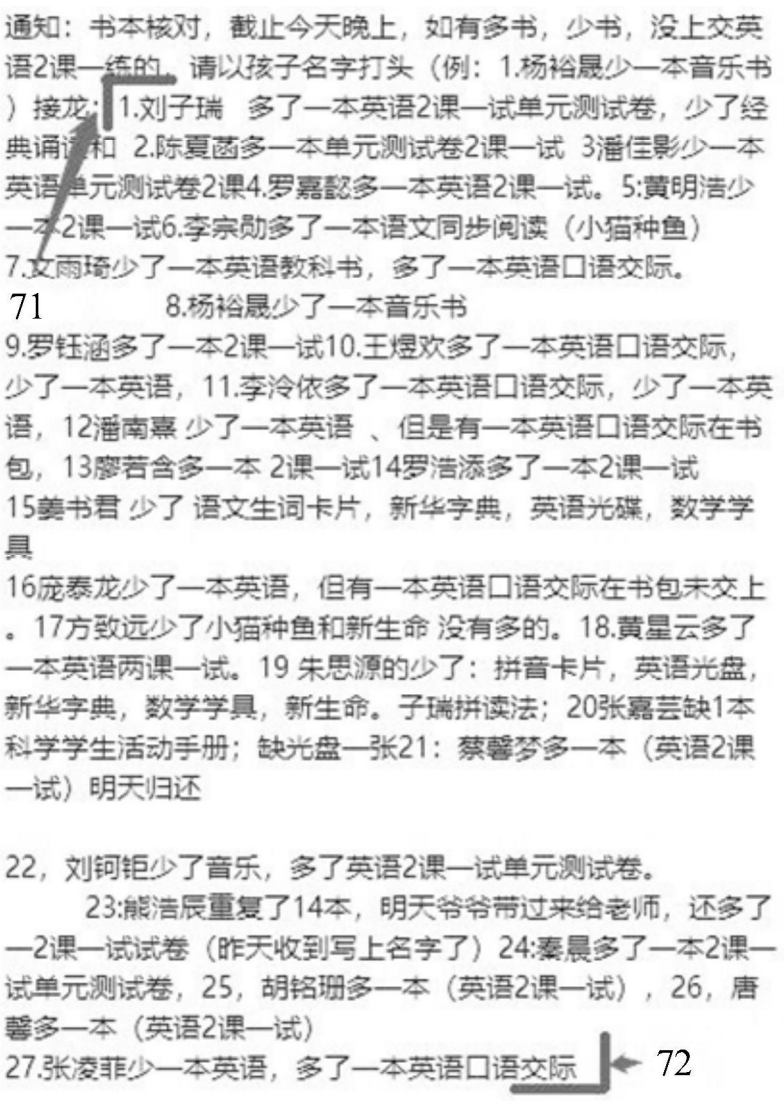

通知：书本核对，截止今天晚上，如有多书，少书，没上交英语2课一练的。请以孩子名字打头（例：1.杨裕晟少一本音乐书）接龙：1.刘子瑞 多了一本英语2课一试单元测试卷，少了经典诵读和 2.陈夏函多一本单元测试卷2课一试 3潘佳影少一本英语单元测试卷2课4.罗嘉懿多一本英语2课一试。5:黄明浩少一本2课一试6.李宗勋多了一本语文同步阅读（小猫种鱼）
7.文雨琦少了一本英语教科书，多了一本英语口语交际。
71
8.杨裕晟少了一本音乐书
9.罗钰涵多了一本2课一试10.王煜欢多了一本英语口语交际，少了一本英语，11.李泠依多了一本英语口语交际，少了一本英语，12潘南熹 少了一本英语 、但是有一本英语口语交际在书包，13廖若含多一本 2课一试14罗浩添多了一本2课一试
15姜书君 少了 语文生词卡片，新华字典，英语光碟，数学学具
16庞泰龙少了一本英语，但有一本英语口语交际在书包未交上。17方致远少了小猫种鱼和新生命 没有多的。18.黄量云多了一本英语两课一试。19 朱思源的少了：拼音卡片，英语光盘，新华字典，数学学具，新生命。子瑞拼读法；20张嘉芸缺1本科学学生活动手册；缺光盘一张21：蔡馨梦多一本（英语2课一试）明天归还
22，刘轲钜少了音乐，多了英语2课一试单元测试卷。
23:熊浩辰重复了14本，明天爷爷带过来给老师，还多了一2课一试试卷（昨天收到写上名字了）24:秦晨多了一本2课一试单元测试卷，25，胡铭珊多一本（英语2课一试），26，唐馨多一本（英语2课一试）
27.张凌菲少一本英语，多了一本英语口语交际
72

图7

1. 刘子瑞多了一本英语2课一试单元测试卷，少了经典诵读和
2. 刘夏菡多一本单元测试卷2课一试
3. 潘佳影少一本英语单元测试卷2课
4. 罗嘉懿多一本英语2课一试
5. 黄明浩少一本2课一试
6. 李宗勋多了一本语文同步阅读（小猫种鱼）
7. 文雨琦少了一本英语教科书，多了一本英语口语交际。
8. 杨裕晟少了一本音乐书
9. 罗钰涵多了一本2课一试
10. 王煜欢多了一本英语口语交际，少了一本英语，
11. 李泠依多了一本英语口语交际，少了一本英语，
12. 潘南熹少了一本英语、但是有一本英语口语交际在书包，
13. 廖若含多一本2课一试
14. 罗浩添多了一本2课一试
15. 姜书君少了语文生词卡片，新华字典，英语光碟，数学学具
16. 庞泰龙少了一本英语，但又一本英语口语交际在书包未交上。
17. 方致远少了小猫种鱼和新生命，没有多的。
18. 黄星云多了一本英语两课一试。、
19. 朱思源的少了：拼音卡片，英语光盘，新华字典，数学学具，新生命。子瑞拼读法；
20. 张嘉芸缺1本科学学生活动手册；缺光盘一张
21. 蔡馨梦多一本（英语2课一试）明天归还
22. 刘轲钜少了音乐，多了英语2课一试单元测试卷。
23. 熊浩辰重复了14本，明天爷爷带过来给老师，还多了一2课一试试卷（昨天收到写上名字了）
24. 秦晨多了一本2课一试单元测试卷。
25. 胡铭珊多一本（英语2课一试），
26. 唐馨多一本（英语2课一试）
27. 张凌菲少一本英语，多了一本英语口语交际

图8

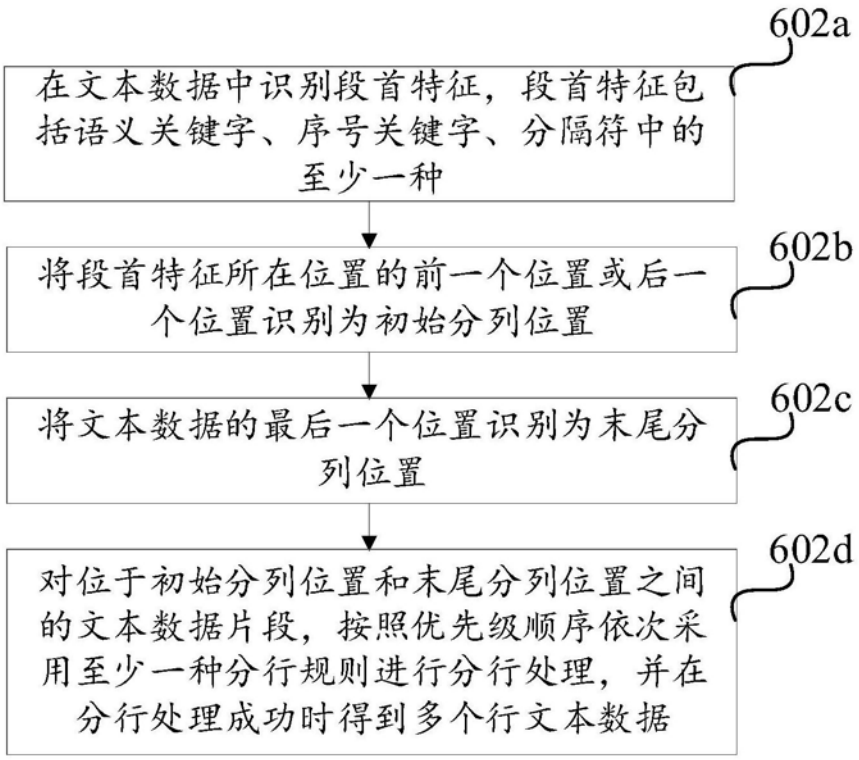

602a
在文本数据中识别段首特征，段首特征包括语义关键字、序号关键字、分隔符中的至少一种
602b
将段首特征所在位置的前一个位置或后一个位置识别为初始分列位置
602c
将文本数据的最后一个位置识别为末尾分列位置
602d
对位于初始分列位置和末尾分列位置之间的文本数据片段，按照优先级顺序依次采用至少一种分行规则进行分行处理，并在分行处理成功时得到多个行文本数据

图9

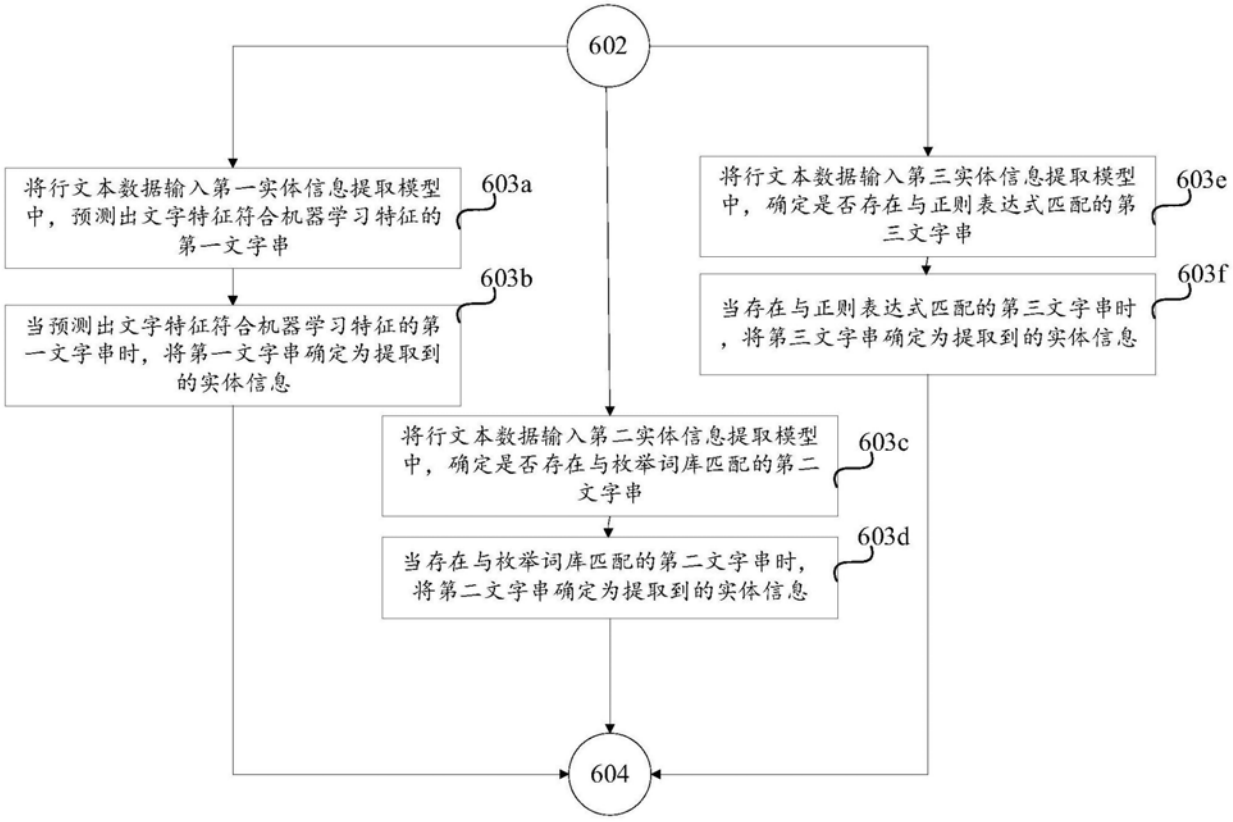

602
将行文本数据输入第一实体信息提取模型中，预测出文字特征符合机器学习特征的第一文字串
603a
当预测出文字特征符合机器学习特征的第一文字串时，将第一文字串确定为提取到的实体信息
603b
将行文本数据输入第二实体信息提取模型中，确定是否存在与枚举词库匹配的第二文字串
603c
当存在与枚举词库匹配的第二文字串时，将第二文字串确定为提取到的实体信息
603d
将行文本数据输入第三实体信息提取模型中，确定是否存在与正则表达式匹配的第三文字串
603e
当存在与正则表达式匹配的第三文字串时，将第三文字串确定为提取到的实体信息
603f
604

图10

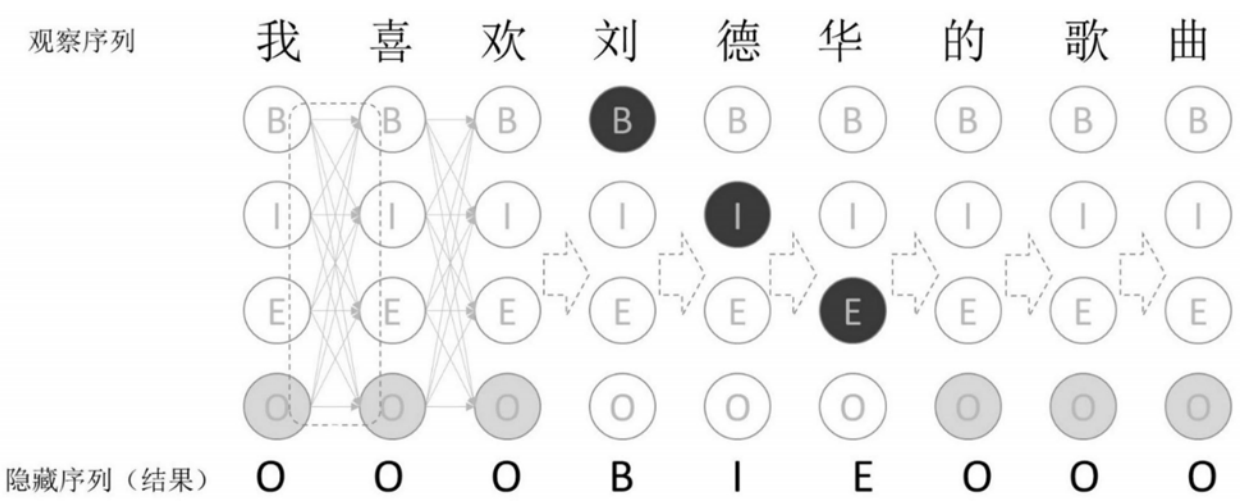

观察序列
我 喜 欢 刘 德 华 的 歌 曲
B B B B B B B B B
I I I I I I I I I
E E E E E E E E E
O O O O O O O O O
隐藏序列（结果）
O O O B I E O O O

图11

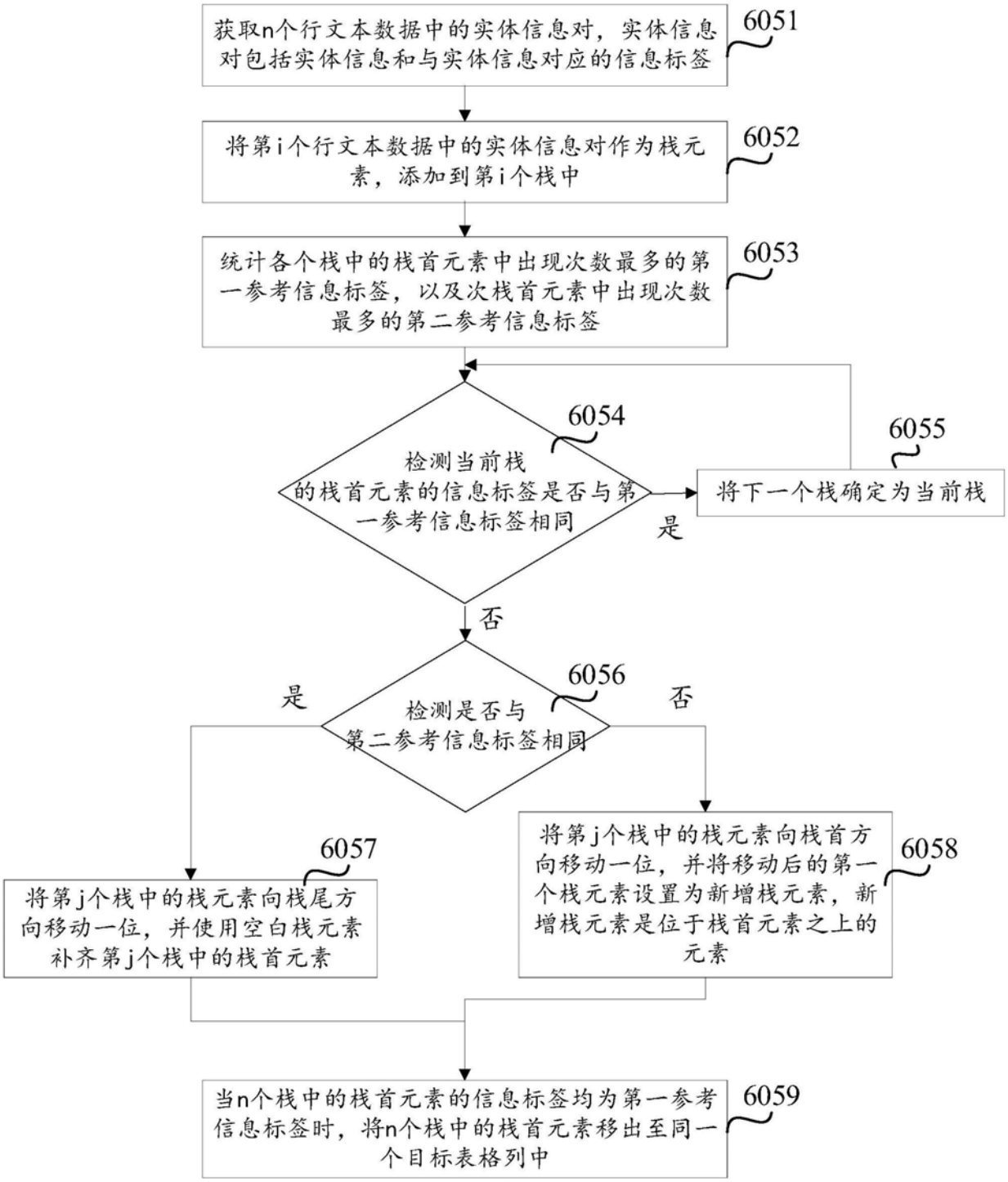
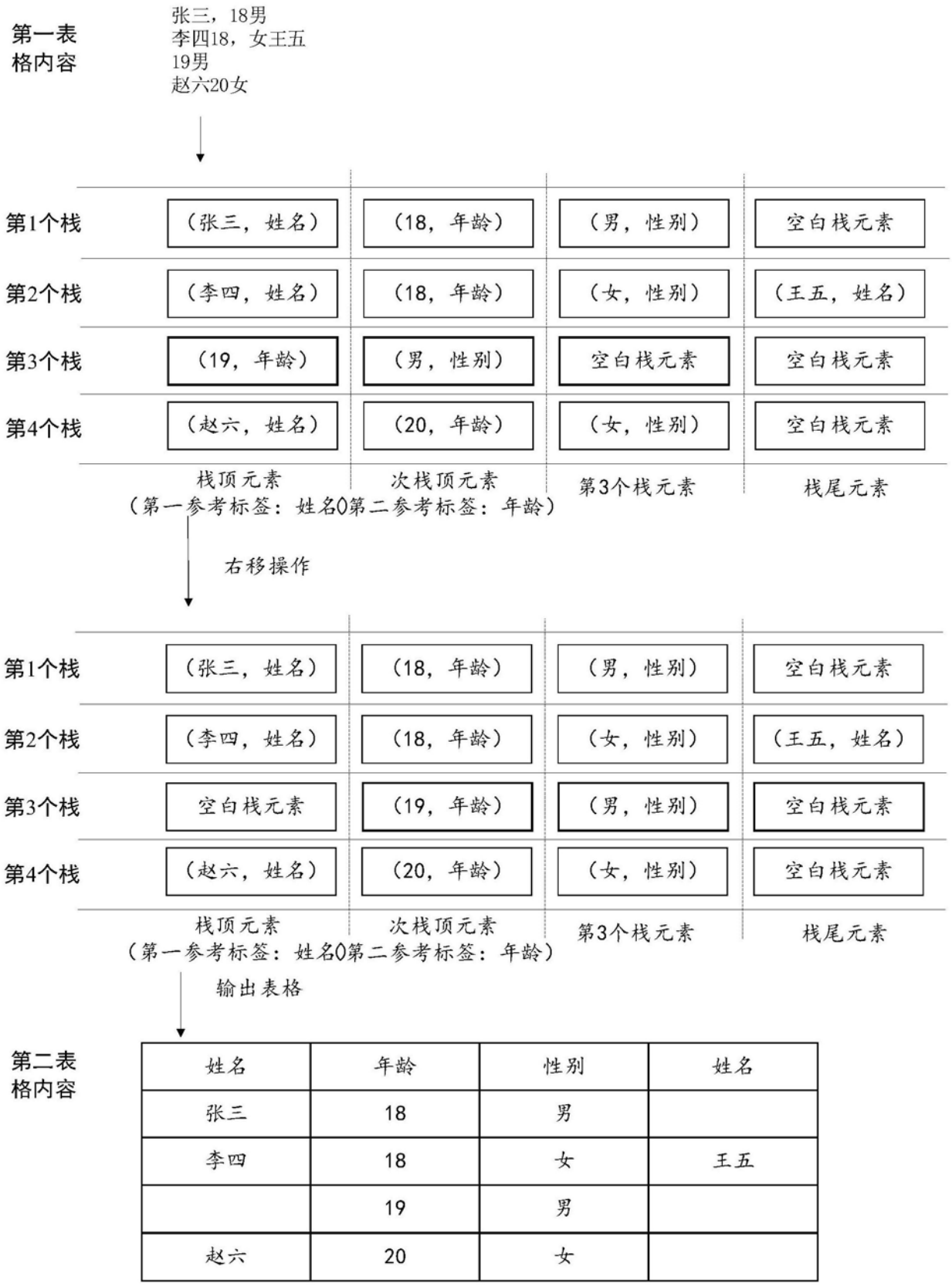

获取n个行文本数据中的实体信息对，实体信息对包括实体信息和与实体信息对应的信息标签
6051
将第i个行文本数据中的实体信息对作为栈元素，添加到第i个栈中
6052
统计各个栈中的栈首元素中出现次数最多的第一参考信息标签，以及次栈首元素中出现次数最多的第二参考信息标签
6053
检测当前栈的栈首元素的信息标签是否与第一参考信息标签相同
6054
是
将下一个栈确定为当前栈
6055
否
检测是否与第二参考信息标签相同
6056
是
否
将第j个栈中的栈元素向栈尾方向移动一位，并使用空白栈元素补齐第j个栈中的栈首元素
6057
将第j个栈中的栈元素向栈首方向移动一位，并将移动后的第一个栈元素设置为新增栈元素，新增栈元素是位于栈首元素之上的元素
6058
当n个栈中的栈首元素的信息标签均为第一参考信息标签时，将n个栈中的栈首元素移出至同一个目标表格列中
6059

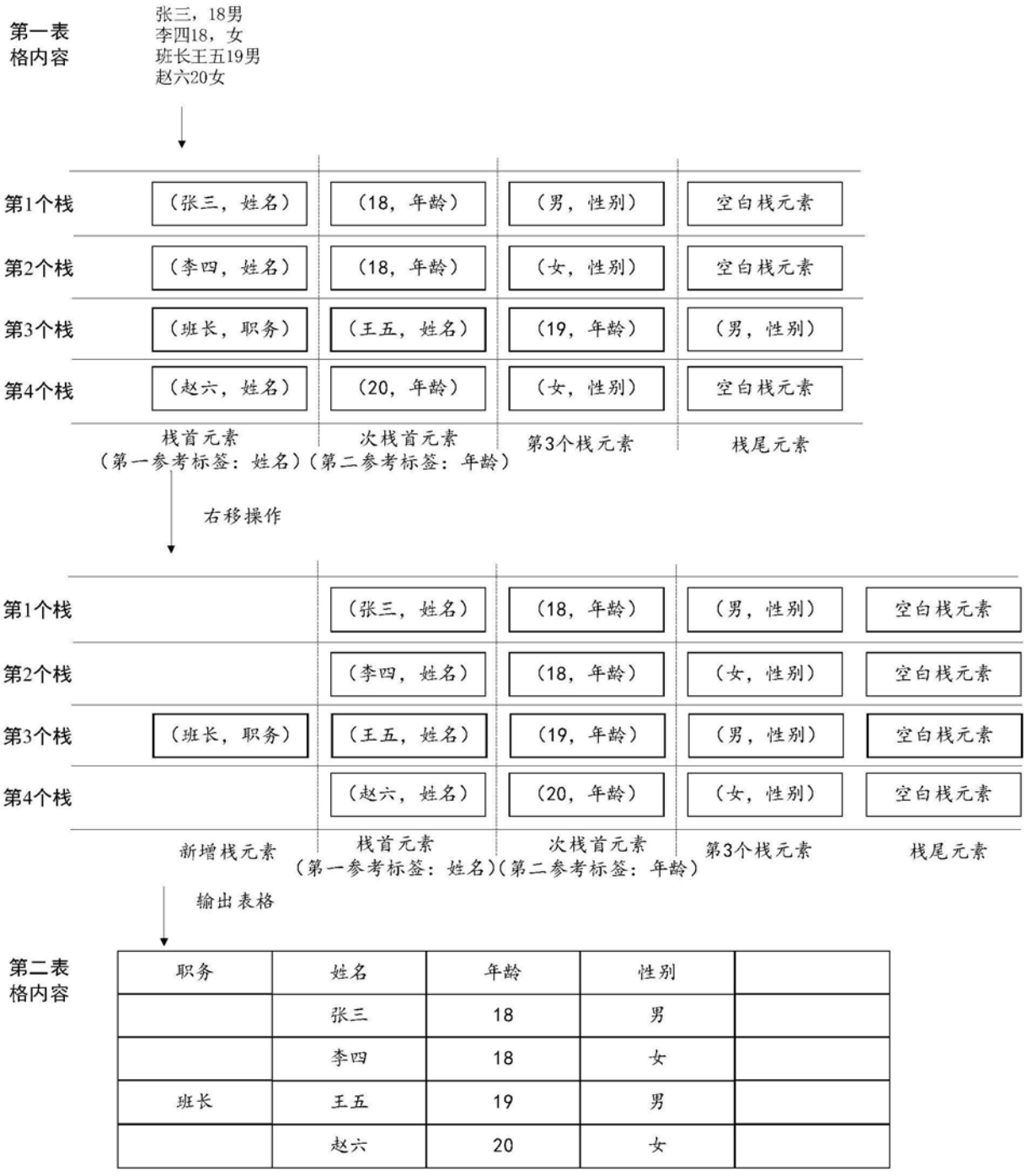
图12



第二表格内容

姓名	年龄	性别	姓名
张三	18	男	
李四	18	女	王五
	19	男	
赵六	20	女	

图13



职务	姓名	年龄	性别	
	张三	18	男	
	李四	18	女	
班长	王五	19	男	
	赵六	20	女	

图14

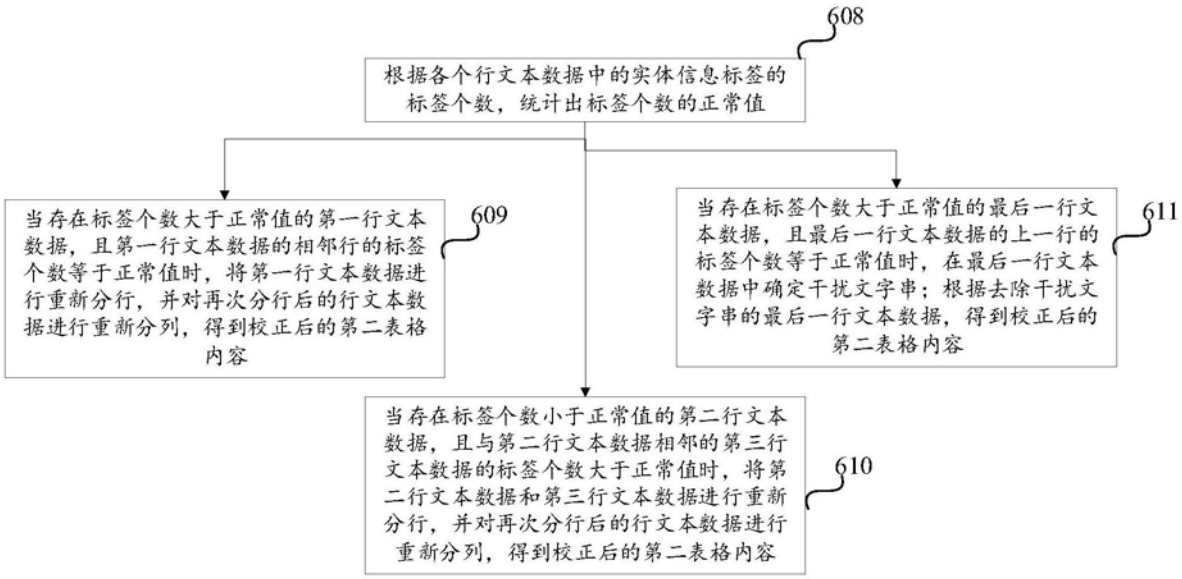
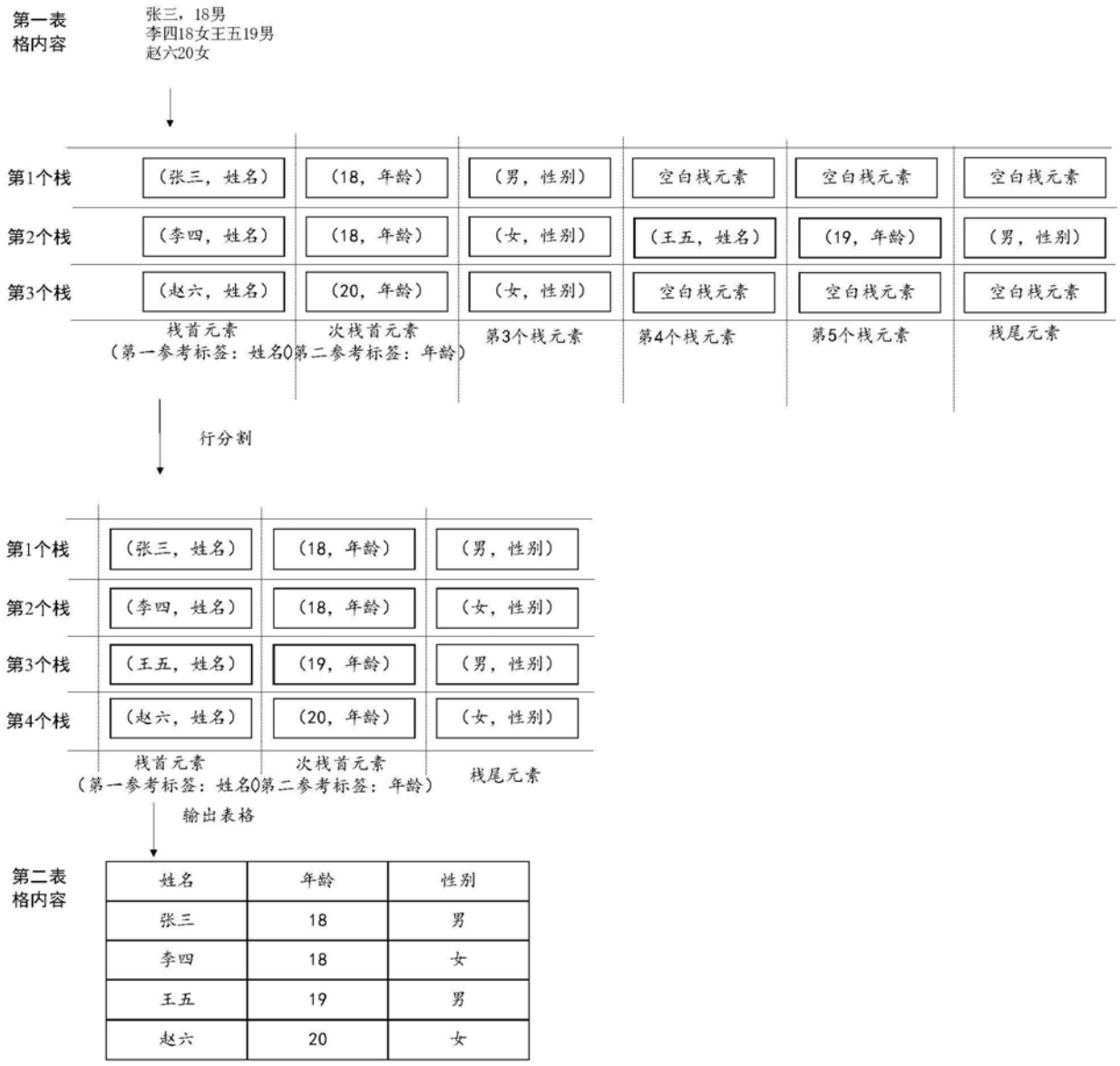
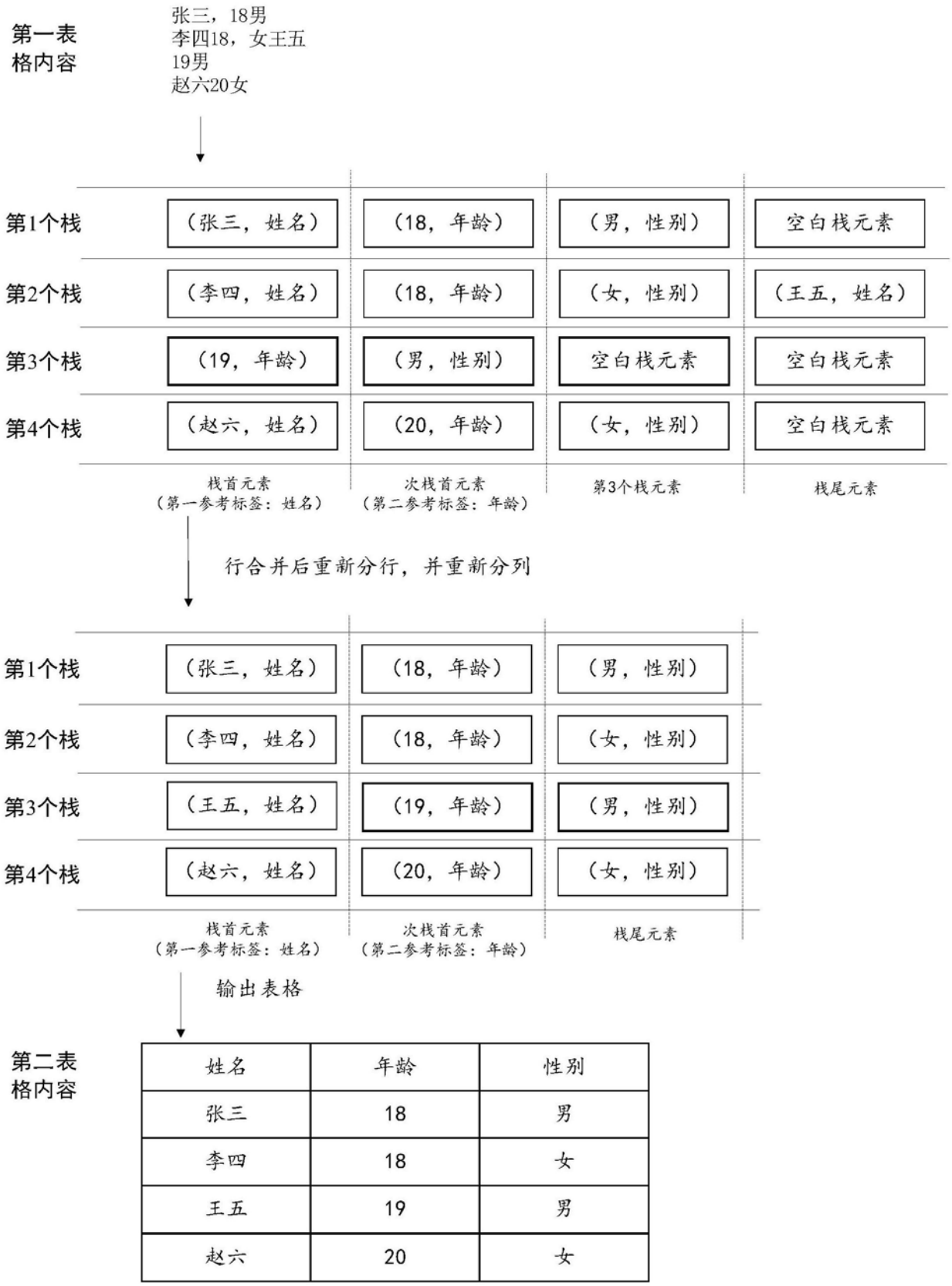

608
根据各个行文本数据中的实体信息标签的标签个数，统计出标签个数的正常值
609
当存在标签个数大于正常值的第一行文本数据，且第一行文本数据的相邻行的标签个数等于正常值时，将第一行文本数据进行重新分行，并对再次分行后的行文本数据进行重新分列，得到校正后的第二表格内容
610
当存在标签个数小于正常值的第二行文本数据，且与第二行文本数据相邻的第三行文本数据的标签个数大于正常值时，将第二行文本数据和第三行文本数据进行重新分行，并对再次分行后的行文本数据进行重新分列，得到校正后的第二表格内容
611
当存在标签个数大于正常值的最后一行文本数据，且最后一行文本数据的上一行的标签个数等于正常值时，在最后一行文本数据中确定干扰文字串；根据去除干扰文字串的最后一行文本数据，得到校正后的第二表格内容

图15



姓名	年龄	性别
张三	18	男
李四	18	女
王五	19	男
赵六	20	女

图16



第二表格内容

姓名	年龄	性别
张三	18	男
李四	18	女
王五	19	男
赵六	20	女

图17

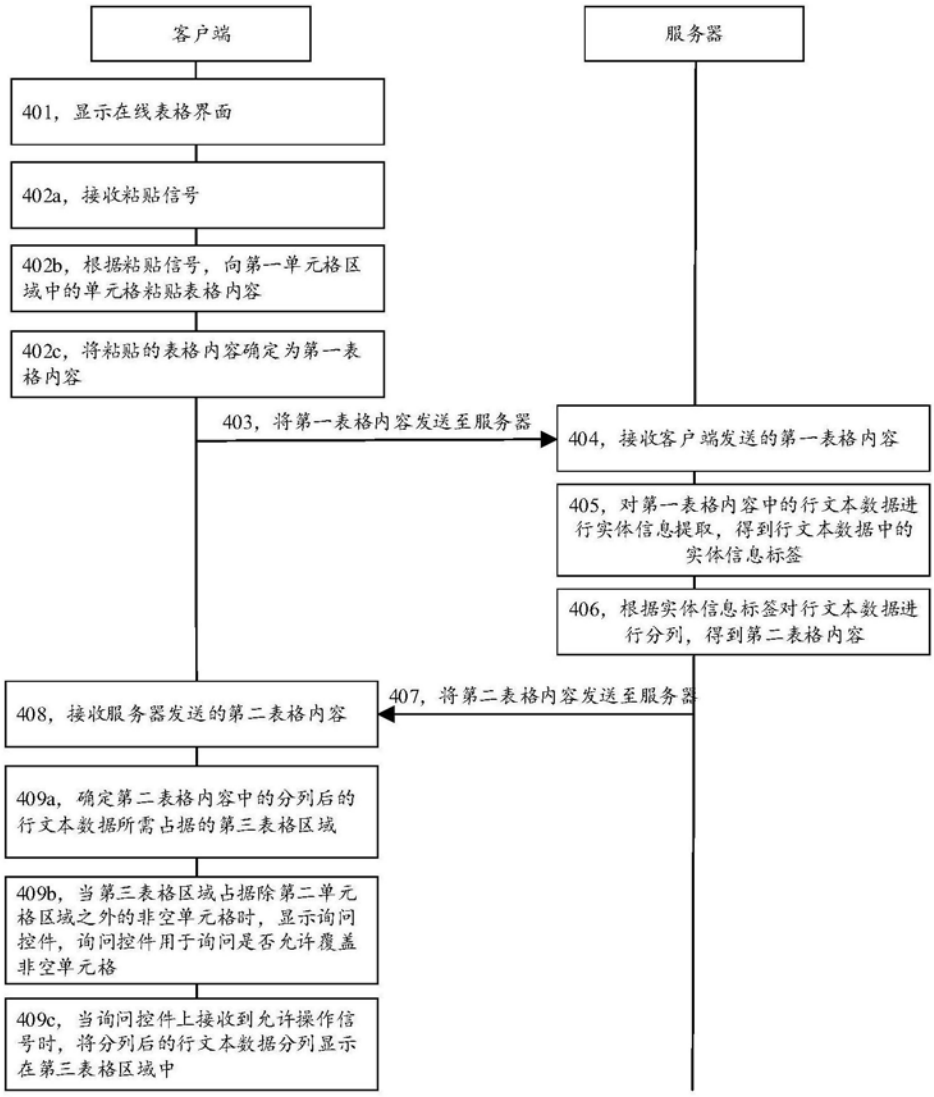

客户端
服务器
401，显示在线表格界面
402a，接收粘贴信号
402b，根据粘贴信号，向第一单元格区域中的单元格粘贴表格内容
402c，将粘贴的表格内容确定为第一表格内容
403，将第一表格内容发送至服务器
404，接收客户端发送的第一表格内容
405，对第一表格内容中的行文本数据进行实体信息提取，得到行文本数据中的实体信息标签
406，根据实体信息标签对行文本数据进行分列，得到第二表格内容
407，将第二表格内容发送至服务器
408，接收服务器发送的第二表格内容
409a，确定第二表格内容中的分列后的行文本数据所需占据的第三表格区域
409b，当第三表格区域占据除第二单元格区域之外的非空单元格时，显示询问控件，询问控件用于询问是否允许覆盖非空单元格
409c，当询问控件上接收到允许操作信号时，将分列后的行文本数据分列显示在第三表格区域中

图18

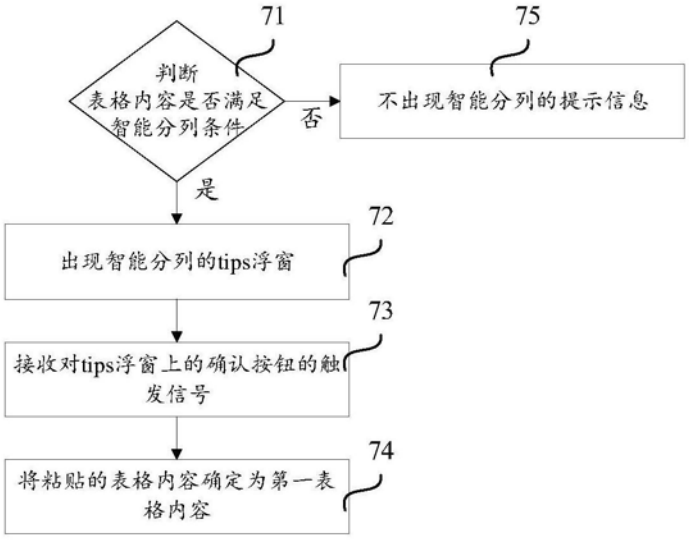

71
判断
表格内容是否满足
智能分列条件
否
75
不出现智能分列的提示信息
是
72
出现智能分列的tips浮窗
73
接收对tips浮窗上的确认按钮的触发信号
74
将粘贴的表格内容确定为第一表格内容

图19

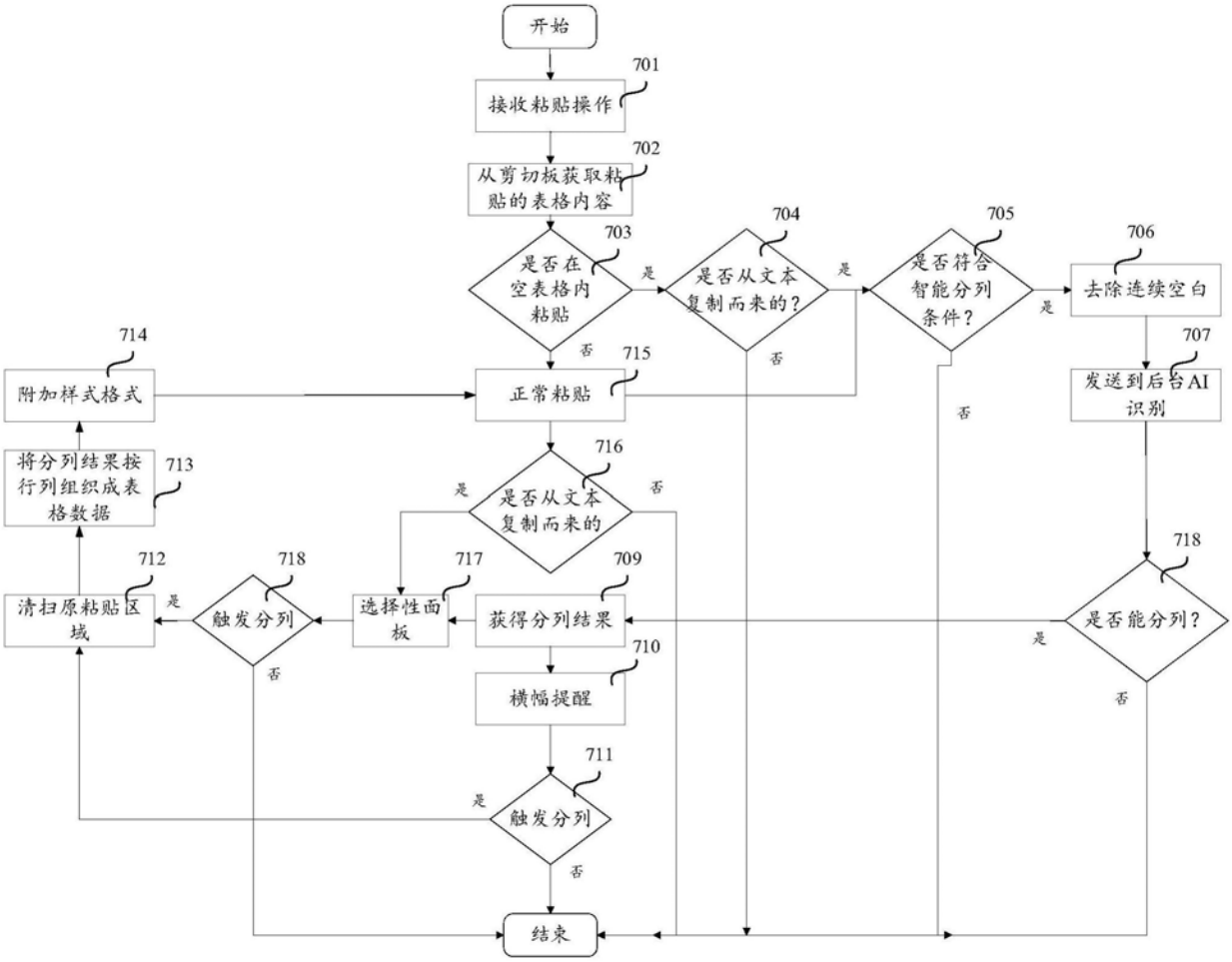

开始
701
接收粘贴操作
702
从剪切板获取粘贴的表格内容
703
是否在空表格内粘贴
是
704
是否从文本复制而来的？
是
705
是否符合智能分列条件？
是
706
去除连续空白
707
发送到后台AI识别
否
否
否
714
附加样式格式
715
正常粘贴
713
将分列结果按行列组织成表格数据
716
是否从文本复制而来的
是
否
712
清扫原粘贴区域
是
718
触发分列
否
717
选择性面板
709
获得分列结果
718
是否能分列？
是
否
710
横幅提醒
711
触发分列
是
否
结束

图20

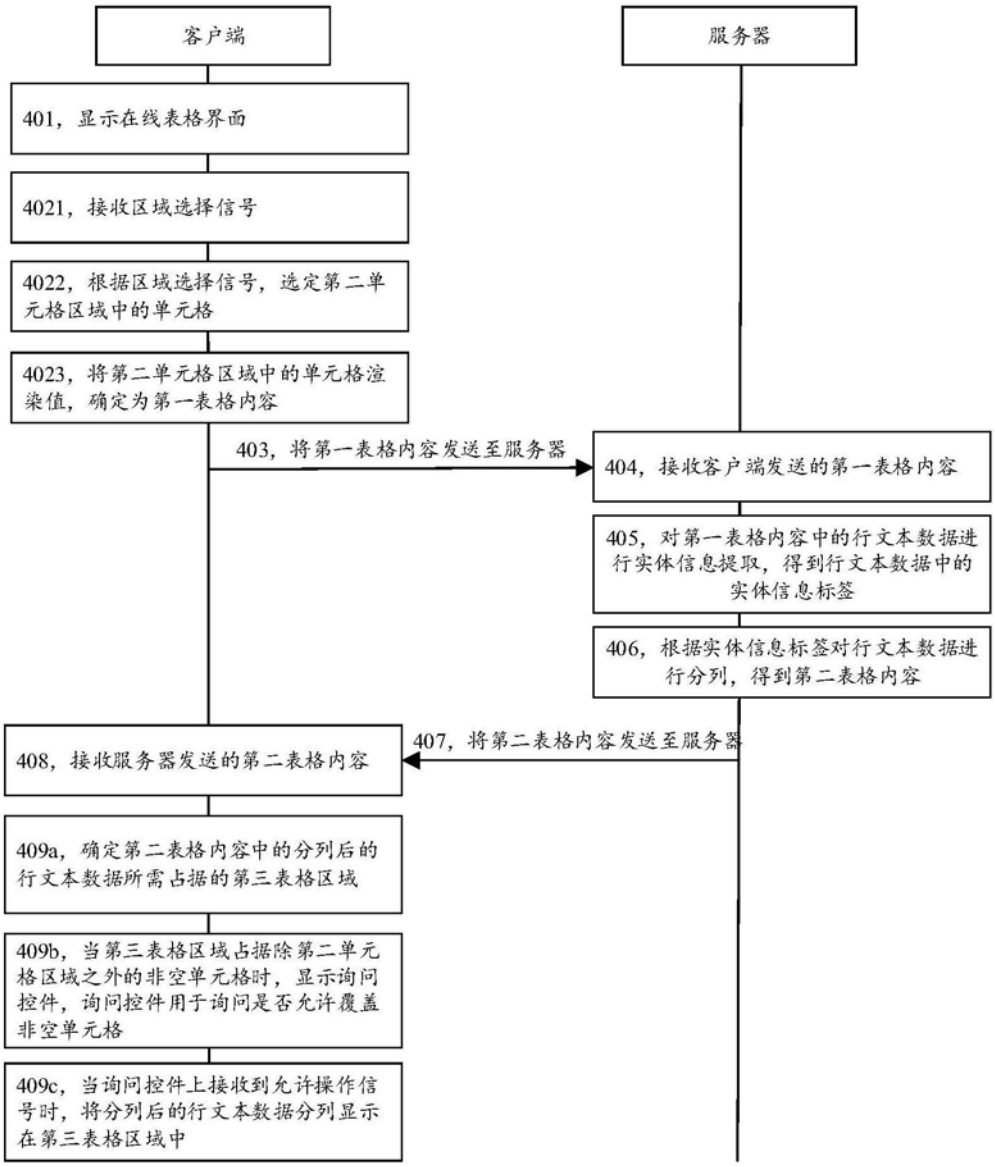

客户端
服务器
401，显示在线表格界面
4021，接收区域选择信号
4022，根据区域选择信号，选定第二单元格区域中的单元格
4023，将第二单元格区域中的单元格渲染值，确定为第一表格内容
403，将第一表格内容发送至服务器
404，接收客户端发送的第一表格内容
405，对第一表格内容中的行文本数据进行实体信息提取，得到行文本数据中的实体信息标签
406，根据实体信息标签对行文本数据进行分列，得到第二表格内容
407，将第二表格内容发送至服务器
408，接收服务器发送的第二表格内容
409a，确定第二表格内容中的分列后的行文本数据所需占据的第三表格区域
409b，当第三表格区域占据除第二单元格区域之外的非空单元格时，显示询问控件，询问控件用于询问是否允许覆盖非空单元格
409c，当询问控件上接收到允许操作信号时，将分列后的行文本数据分列显示在第三表格区域中

图21

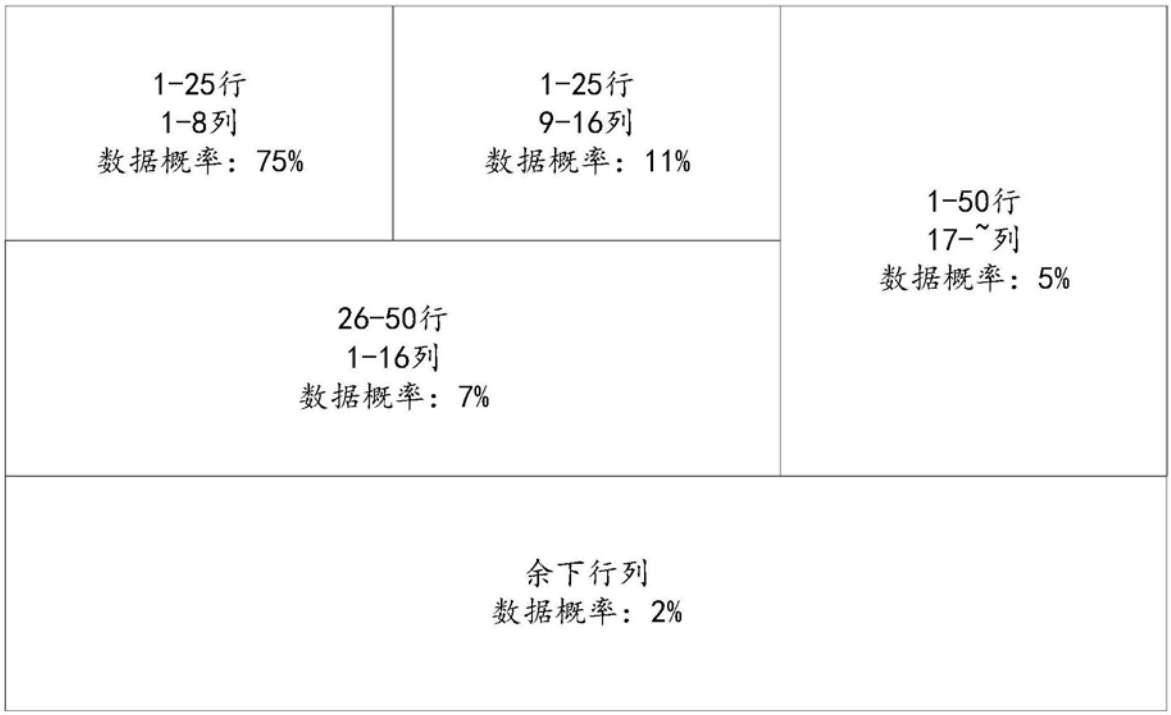

1-25行
1-8列
数据概率：75%
1-25行
9-16列
数据概率：11%
1-50行
17-~列
数据概率：5%
26-50行
1-16列
数据概率：7%
余下行列
数据概率：2%

图22

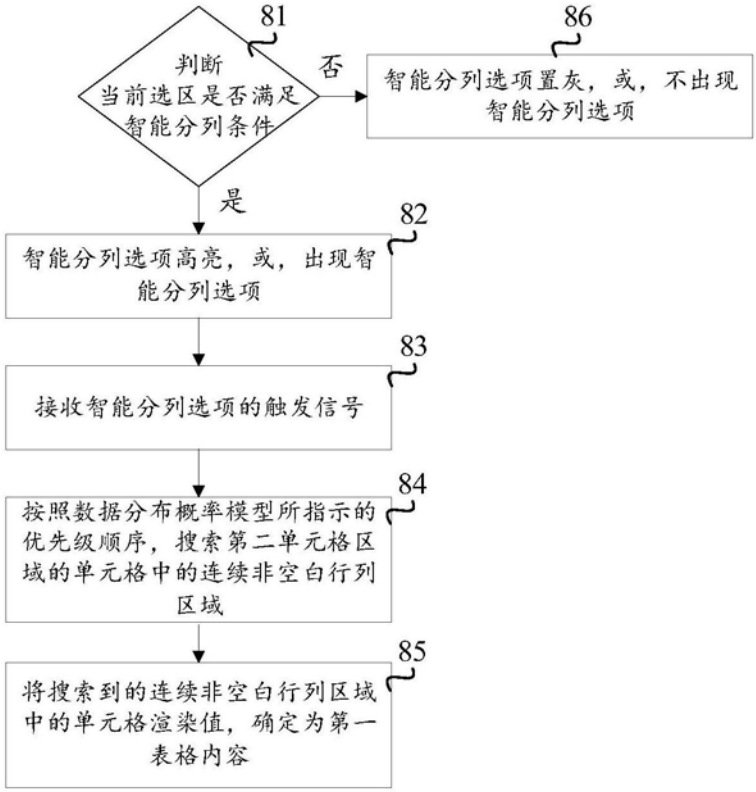

81
判断
当前选区是否满足
智能分列条件
否
86
智能分列选项置灰，或，不出现
智能分列选项
是
82
智能分列选项高亮，或，出现智
能分列选项
83
接收智能分列选项的触发信号
84
按照数据分布概率模型所指示的
优先级顺序，搜索第二单元格区
域的单元格中的连续非空白行列
区域
85
将搜索到的连续非空白行列区域
中的单元格渲染值，确定为第一
表格内容

图23

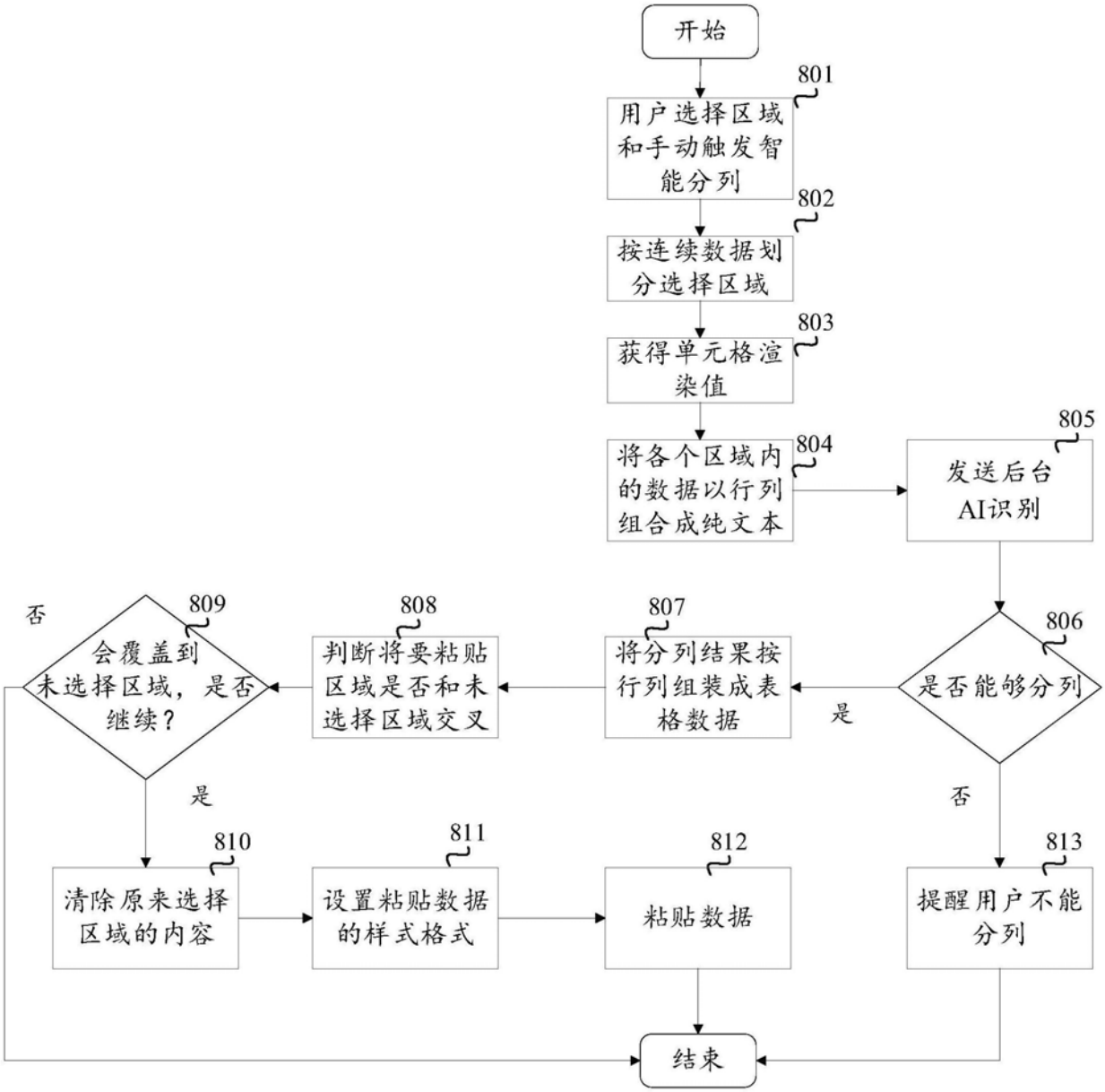

开始
801
用户选择区域和手动触发智能分列
802
按连续数据划分选择区域
803
获得单元格渲染值
804
将各个区域内的数据以行列组合成纯文本
805
发送后台AI识别
806
是否能够分列
是
否
807
将分列结果按行列组装成表格数据
808
判断将要粘贴区域是否和未选择区域交叉
809
会覆盖到未选择区域，是否继续？
否
是
810
清除原来选择区域的内容
811
设置粘贴数据的样式格式
812
粘贴数据
813
提醒用户不能分列
结束

图24

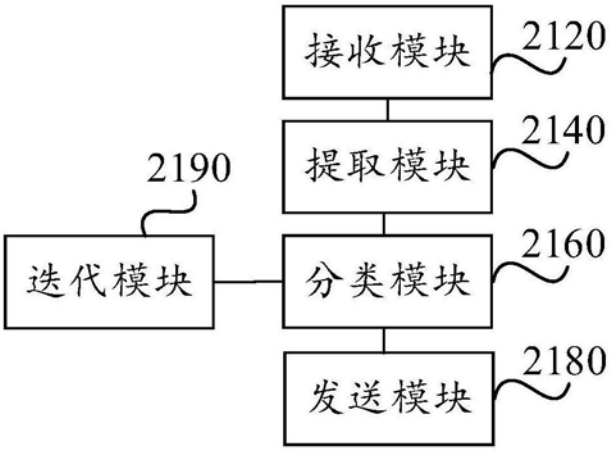

接收模块
2120
提取模块
2140
2190
迭代模块
分类模块
2160
发送模块
2180

图25

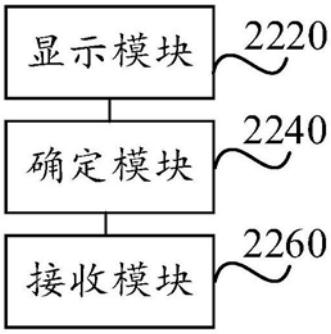

显示模块
2220
确定模块
2240
接收模块
2260

图26

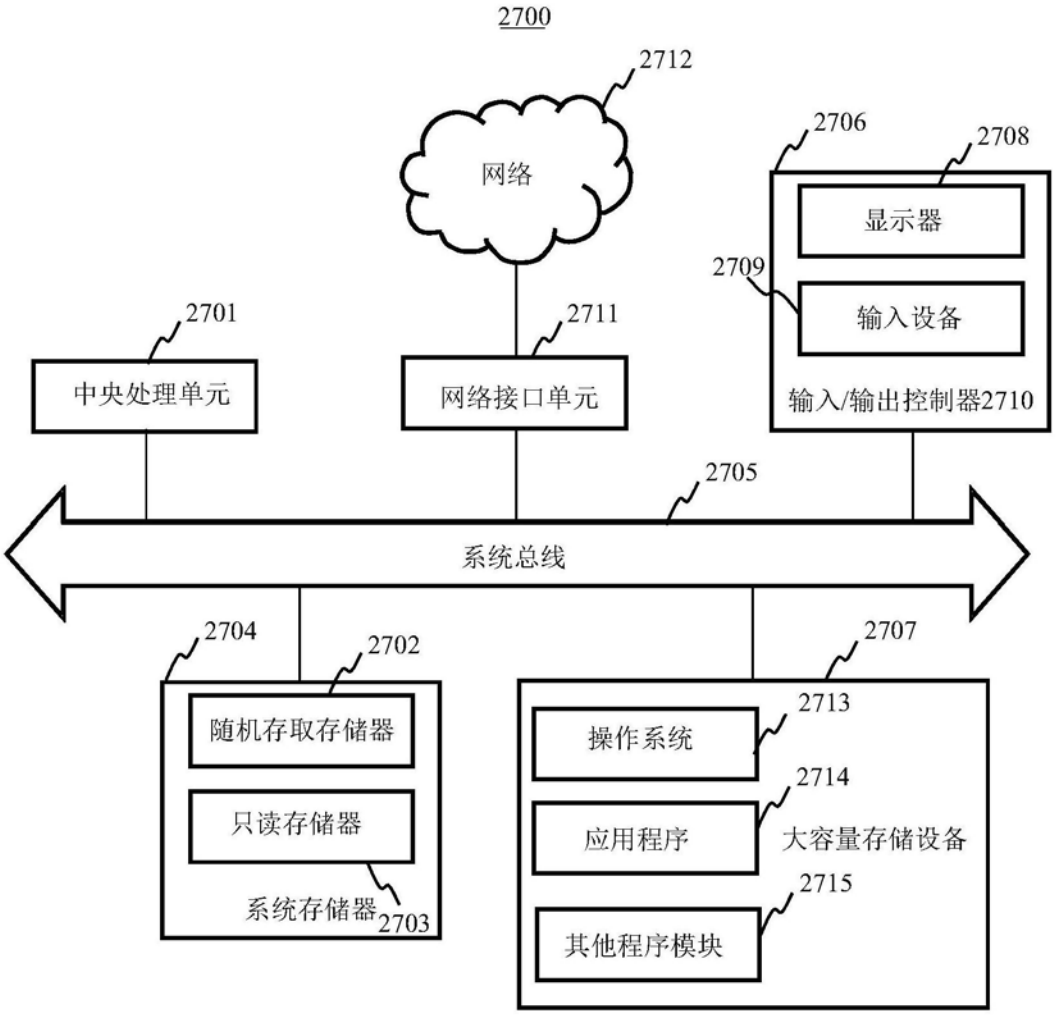

2700
2712
网络
2706
2708
显示器
2709
输入设备
2701
2711
中央处理单元
网络接口单元
输入/输出控制器2710
2705
系统总线
2704
2702
2707
随机存取存储器
2713
操作系统
2714
只读存储器
应用程序
大容量存储设备
系统存储器
2703
2715
其他程序模块

图27

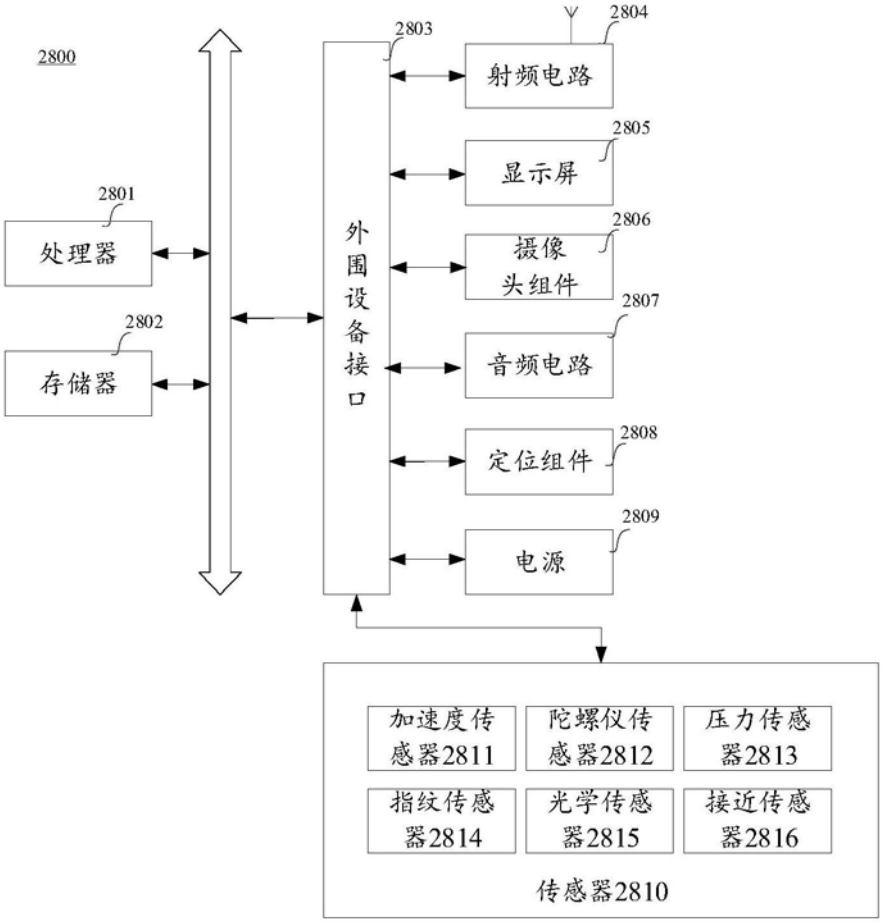

2800
2801
处理器
2802
存储器
2803
外围设备接口
2804
射频电路
2805
显示屏
2806
摄像头组件
2807
音频电路
2808
定位组件
2809
电源
加速度传感器2811
陀螺仪传感器2812
压力传感器2813
指纹传感器2814
光学传感器2815
接近传感器2816
传感器2810

图28