



# (12) 发明专利

(10) 授权公告号 CN 115841312 B

(45) 授权公告日 2023.05.02

(21) 申请号 202310169142.X

(22) 申请日 2023.02.27

(65) 同一申请的已公布的文献号

申请公布号 CN 115841312 A

(43) 申请公布日 2023.03.24

(73) 专利权人 武汉耳东信息科技有限公司

地址 430058 湖北省武汉市经济技术开发区  
海棠路55号联创科技中心1号楼乐  
客工场孵化器(集-LKGC-B63)

专利权人 武汉彤新科技有限公司  
北京轻唯科技有限公司  
上海现安科技有限公司  
子为科技武汉市有限公司

(72) 发明人 彭小含 陈帅 秦中想

(74) 专利代理机构 安徽智鼎华诚专利代理事务  
所(普通合伙) 34242

专利代理师 赵春海

(51) Int.Cl.

G06Q 10/10 (2023.01)

G06F 16/16 (2019.01)

G06F 16/13 (2019.01)

G06F 16/172 (2019.01)

审查员 李云

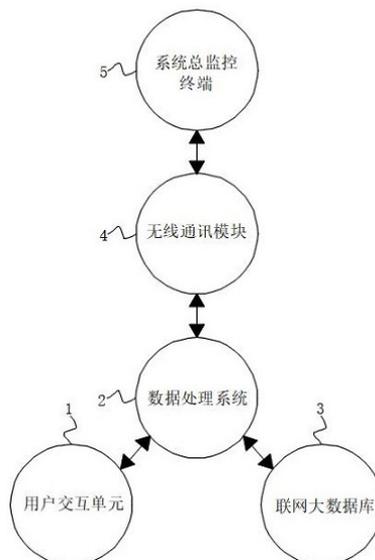
权利要求书3页 说明书9页 附图7页

## (54) 发明名称

用于企业管理文件数据实时采集分析系统  
及其采集终端

## (57) 摘要

本发明公开了用于企业管理文件数据实时采集分析系统及其采集终端,包括用户交互单元、数据处理系统、联网大数据库以及通过无线通讯模块与数据处理系统实现数据无线交互的系统总监控终端,本发明涉及企业管理数据处理技术领域。该用于企业管理文件数据实时采集分析系统及其采集终端,可实现通过将企业用户与第三方聊天应用进行对接,实时采集和分析企业员工在第三方应用上与客户或者与官方来往邮箱的文件数据以及聊天内容,同时很好的避免客户沟通处理的及时性,大大提高了整个公司运营效率,可实现通过对子文件数据模型与母文件数据模型的索引符进行智能对比,来确定新文件数据模型的存储位置。



1. 用于企业管理文件数据实时采集分析系统,包括用户交互单元(1)、数据处理系统(2)、联网大数据库(3)以及通过无线通讯模块(4)与数据处理系统(2)实现数据无线交互的系统总监控终端(5),其特征在于:用户通过所述用户交互单元(1)与数据处理系统(2)进行交互,且数据处理系统(2)能够向联网大数据库(3)内提取处理算法;

所述数据处理系统(2)包括系统后台服务器(21)、第三方数据导入单元(22)、数据采集池(23)、母文件数据库单元(24)、子文件数据模型创建单元(25)和文件数据对比分析单元(26),通过所述第三方数据导入单元(22)将企业工作人员与客户沟通的聊天记录和来往文件数据导入数据处理系统(2)内,再通过数据采集池(23)对导入的数据进行实时采集,采集的数据通过子文件数据模型创建单元(25)创建出带有索引符的文件数据子模型,然后通过系统后台服务器(21)将创建的带有索引符的文件数据子模型发送至文件数据对比分析单元(26)内进行对比分析,来对导入的数据进行归纳存储或创建新项目进行存储;

所述文件数据对比分析单元(26)包括微处理模块(261)、数据对比算法提取模块(262)、文件数据模型分析单元(263)、新文件数据模型生成模块(264)、初始文件类型判别模块(265)、数据实时更新模块(266)和数据收发模块(267);

所述微处理模块(261)控制数据收发模块(267)分别接受系统后台服务器(21)发来的子文件数据模型以及母文件数据库单元(24)内存储的母文件数据模型,先通过初始文件类型判别模块(265)对接收的子文件数据模型的索引符与母文件数据库单元(24)内存储的母文件数据模型索引符进行对比,来判定该子文件数据模型存储的聚焦位置,初始文件类型判别模块(265)会向联网大数据库(3)内提取索引符处理算法,来进行索引符对比处理,若初始判别出母文件数据库单元(24)内无对应存储位置,则直接通过新文件数据模型生成模块(264)在母文件数据库单元(24)内生成新的项目存储模块来存储该子文件数据模型,并将该子文件数据模型的索引符前增加新创建新的存储编码,升级成母文件数据模型,即生成新的索引符,当初始判别出母文件数据库单元(24)内有对应存储位置,则通过文件数据模型分析单元(263)进行同文件数据模型内的同项目判断;

所述索引符处理算法具体包括以下步骤:

S1、子文件数据模型的索引符由以下三个标识组成:企业工作人员编号 $a=\{0,1,2,3,\dots,n\}$ 、该工作人员名下立项的客户编号 $b=\{0,1,2,3,\dots,n\}$ 以及该子文件数据模型内采集整合的数据标识代码 $X=\{X_1,X_2,X_3,\dots,X_n\}$ ,这三个标识进行分类归纳处理得到索引符 $abX$ ,在生成子文件数据模型时,企业工作人员编号 $a$ 是系统可知的确定数值,该工作人员名下立项的客户编号 $b$ 被随机赋值,但赋值过程中,其辅标识标记客户名称转换后的二级制代码,即为 $b(\Omega)$ ,而该子文件数据模型内采集整合的数据表代码 $X$ 直接随机赋值,即组合成该子文件数据模型的索引符为 $ab(\Omega)X$ ;

S2、初始文件类型判别模块(265)会将步骤S1生成的子文件数据模型索引符发送至母文件数据库单元(24)内与母文件数据库单元(24)内各个项目存储模块上的索引符进行对比,先通过步骤S1中确定好的编号 $a$ 筛选出母文件数据库单元(24)内与该企业工作人员所对应的母文件数据模型;

S3、根据步骤S1中该工作人员名下立项的客户编号 $b(\Omega)$ 中的辅标识标记 $\Omega$ 在步骤S2筛选出的母文件数据模型中进行搜索对比,若匹配到对应母文件数据模型,则通过文件数据模型分析单元(263)内的数据对比算法进行立项识别;

S4、若未配到对应母文件数据模型,则直接通过新文件数据模型生成模块(264)在母文件数据库单元(24)内生成新的项目存储模块来存储该子文件数据模型,并将该子文件数据模型的索引符前增加新创建新的存储编码 $U=\{U_1, U_2, U_3, \dots, U_n\}$ ,升级成母文件数据模型,即生成新的索引符 $UabX$ 。

2. 根据权利要求1所述的用于企业管理文件数据实时采集分析系统,其特征在于:所述第三方数据导入单元(22)包括客户聊天数据导入模块(221)、官方文件导入模块(222)和客户发文导入模块(223);

所述第三方数据导入单元(22)是将企业工作人员的第三方聊天应用软件均应用于该用于企业管理文件数据实时采集分析系统内,并使企业工作人员的第三方聊天应用软件通过企业的局域网内进行客户的沟通,所述客户聊天数据导入模块(221)和客户发文导入模块(223)用于将企业工作人员与客户之间在第三方应用沟通的聊天内容实时导入数据处理系统(2)内,且官方文件导入模块(222)用于将企业工作人员与官方之间在第三方应用来往的文件实时导入数据处理系统(2)内。

3. 根据权利要求1所述的用于企业管理文件数据实时采集分析系统,其特征在于:所述数据采集池(23)是由n个数据采集模块组成,且n个数据采集模块能够对第三方数据导入单元(22)导入的文件数据进行实时采集,所述母文件数据库单元(24)是由n个项目存储模块组成,且每个项目存储模块内均存储有一个数据处理系统(2)内预先立项创建好的文件数据模型,即为一个母文件数据模型。

4. 根据权利要求1所述的用于企业管理文件数据实时采集分析系统,其特征在于:所述子文件数据模型创建单元(25)包括文件类型归纳模块(251)、文件逻辑关系生成模块(252)、文件索引标识创建模块(253)、文件索引标识整合模块(254)和文件模型生成模块(255);

通过所述文件类型归纳模块(251)将采集模块采集的数据以企业工作人员编号和该工作人员名下立项的客户编号两个标识进行分类归纳处理得到若干个数据包,再通过文件逻辑关系生成模块(252)识别与该文件相关联的其他关系人,并生成关系图谱,将生成的关系图谱数据合并至生成的对应数据包内,然后沟通过文件索引标识创建模块(253)根据识别的企业工作人员的编号和该工作人员名下立项的客户编号创建成二进制代码,即生成索引符,再通过文件索引标识整合模块(254)将生成的索引符整合到与之对应的该数据包上,之后通过文件模型生成模块(255)生成一个子文件数据模型,之后相同操作进行实时数据处理生成若干个子文件数据模型,并将生成的若干个子文件数据模型依次发送至系统后台服务器(21)。

5. 根据权利要求1所述的用于企业管理文件数据实时采集分析系统,其特征在于:所述文件数据模型分析单元(263)包括子文件数据模型提取模块(2631)、母文件数据模型提取模块(2632)、数据类型对比分析模块(2633)和数据类型判定模块(2634);

通过所述子文件数据模型提取模块(2631)和母文件数据模型提取模块(2632)分别提取待对比的子文件数据模型和母文件数据模型,通过数据类型对比分析模块(2633)进行数据对比分析,数据类型对比分析模块在对比分析过程中会通过数据对比算法提取模块(262)向联网大数据数据库(3)内提取数据对比算法,来进行数据对比分析处理,然后通过数据类型判定模块(2634)判定该子文件数据模型是否为新立项项目。

6. 根据权利要求5所述的用于企业管理文件数据实时采集分析系统,其特征在于:所述数据对比算法具体包括以下步骤:

T1、先通过子文件数据模型提取模块(2631)和母文件数据模型提取模块(2632)分别提取待对比的子文件数据模型和母文件数据模型,再将子文件数据模型内采集整合的数据通过文字和图片识别算法进行特征提取,即对子文件数据模型内的标识代码X内的内容进行处理,分别提取对应的标识特征 $z^1, z^2, z^3, \dots, z^n$ ;

T2、将步骤T2提取的标识特征与筛选出的母文件数据模型内各个立项内容特征进行匹配,匹配过程中计算匹配度W,当匹配度 $W \geq 85\%$ 时,则判定匹配成功,表示该子文件模型为母文件模型中已经立项过的补充文件数据,直接将该子文件模型内的文件数据存储到对应的立项存储模块内,并将该子文件模型的数据标识代码X变更为与该立项存储模块相对于的代码;

T3、当匹配度 $W < 85\%$ 时,则判定匹配失败,表示该子文件模型为母文件模型中新立项的补充文件数据,在该母文件数据模型内创建新的项目存储模块,该新的项目标识代码为现有标识码的相邻的下一位,并将该子文件模型内的文件数据存储到对应的新创建的立项存储模块内,并将该子文件模型的数据标识代码X变更为与新立项存储模块相对于的数据标识代码。

7. 根据权利要求1所述的用于企业管理文件数据实时采集分析系统,其特征在于:所述用户交互单元(1)包括用户登录模块(11)、文件模型删除模块(12)、数据显示模块(13)、文件模型修订模块(14)、文件模型查询模块(15)和文件模型转移模块(16)。

8. 一种采集终端,其特征在于:所述采集终端存储有计算机可运行指令,所述计算机可运行指令在被处理器调用和运行时,所述计算机可运行指令促使所述处理器运行权利要求1-7任一项所述用于企业管理文件数据实时采集分析系统内的运行方法。

## 用于企业管理文件数据实时采集分析系统及其采集终端

### 技术领域

[0001] 本发明涉及企业管理数据处理技术领域,具体为用于企业管理文件数据实时采集分析系统及其采集终端。

### 背景技术

[0002] 企业管理是对企业生产经营活动进行计划、组织、指挥、协调和控制等一系列活动的总称,而在企业管理过程中针对企业文件数据管理尤为繁琐且重要,例如企业业务员与客户沟通数据以及沟通过程中来往的文件(文件材料、合同或修改文件等),还例如企业的财务或流程人员与官方的来往文件(官方通知书、回执或官方证明材料等),这些企业文件一般需要制定的公司职员进行整理上报和存储。

[0003] 参考中国专利公开号为CN110738464A的一种基于实时变动数据的企业数据化管理系统及方法,通过综合各个部门的历史数据和实时数据,进行数据分析,相对于现有的企业管理方式,数据分析的结果更为准确,从而引导管理层做出更有利于企业发展的管理决策。

[0004] 参考中国专利公开号为CN112214536B的一种企业账目数据实时导入管理系统,通过设置分部门出纳管理平台,所述分部门出纳管理平台的数量为若干个,所述分部门出纳管理平台用于部门提交账目数据信息进行汇总,所述财务管理平台用于企业各部门提交账目数据信息进行处理,分析、报表生成、以及对账审核,实现了企业分部门系统化账目数据管理,便于企业高效快速处理账目数据信息。

[0005] 综合分析以上参考专利可知,具体包括以下缺陷:

[0006] 现有的企业文件数据采集系统虽然能够实现对企业数据的采集和分析,但都是采用人为整理导入的数据,例如参考专利CN110738464A的一种基于实时变动数据的企业数据化管理系统及方法和CN112214536B的一种企业账目数据实时导入管理系统,需要工作人员先在系统上建立数据订单,即创建立项数据包,然后导入处理系统内进行数据的分类和处理,但是这样的数据采集分析方式使用较为繁琐,并且不方便公司系统对员工的工作状态进行监控,例如企业员工在公司平台与客户有私下交易等问题,同时无法避免客户沟通处理的及时性,由于业务员可能同时沟通多个客户,不能及时对沟通完成的客户的文件数据进行立项处理的问题,从而大大降低了整个公司运营效率,不能实现通过将企业用户与第三方聊天应用进行对接,实时采集和分析企业员工在第三方应用上与客户或者与官方来往邮箱的文件数据以及聊天内容,无法达到在短时间内进行文件数据立项分析运营的目的,从而给企业的文件数据管理带来极大的不便。

[0007] 现有企业文件数据在进行立项过程中,不能实现自动判断立项内容是否为前立项的关联项目进行智能归档存储,例如参考专利CN110738464A的一种基于实时变动数据的企业数据化管理系统及方法和CN112214536B的一种企业账目数据实时导入管理系统,虽然能够变动数据的识别,但不能根据具体变动内容进行自动更改或补充新数据,不能实现通过对子文件数据模型与母文件数据模型的索引符进行智能对比,来确定新文件数据模型的存

储位置,无法达到既快速又方便的进行企业文件数据的智能整理的目的,从而对企业文官数据管理工作十分不利。

### 发明内容

[0008] 针对现有技术的不足,本发明提供了用于企业管理文件数据实时采集分析系统及其采集终端,解决了现有的企业文件数据采集系统虽然能够实现对企业数据的采集和分析,但都是采用人为整理导入的数据,需要工作人员先在系统上建立数据订单,即创建立项数据包,然后导入处理系统内进行数据的分类和处理,但是这样的数据采集分析方式使用较为繁琐,并且不方便公司系统对员工的工作状态进行监控,同时无法避免客户沟通处理的及时性,不能实现通过将企业用户与第三方聊天应用进行对接,实时采集和分析企业员工在第三方应用上与客户或者与官方来往邮箱的文件数据以及聊天内容,无法达到在短时间内进行文件数据立项分析运营的目的,现有企业文件数据在进行立项过程中,不能实现自动判断立项内容是否为前立项的关联项目进行智能归档存储,但不能根据具体变动内容进行自动更改或补充新数据,不能实现通过对子文件数据模型与母文件数据模型的索引符进行智能对比,来确定新文件数据模型的存储位置,无法达到既快速又方便的进行企业文件数据的智能整理目的的问题。

[0009] 为实现以上目的,本发明通过以下技术方案予以实现:用于企业管理文件数据实时采集分析系统,包括用户交互单元、数据处理系统、联网大数据库以及通过无线通讯模块与数据处理系统实现数据无线交互的系统总监控终端,用户通过所述用户交互单元与数据处理系统进行交互,且数据处理系统能够向联网大数据库内提取处理算法;

[0010] 所述数据处理系统包括系统后台服务器、第三方数据导入单元、数据采集池、母文件数据库单元、子文件数据模型创建单元和文件数据对比分析单元,通过所述第三方数据导入单元将企业工作人员与客户沟通的聊天记录和来往文件数据导入数据处理系统内,再通过数据采集池对导入的数据进行实时采集,采集的数据通过子文件数据模型创建单元创建出带有索引符的文件数据子模型,然后通过系统后台服务器将创建的带有索引符的文件数据子模型发送至文件数据对比分析单元内进行对比分析,来对导入的数据进行归纳存储或创建新项目进行存储。

[0011] 优选的,所述第三方数据导入单元包括客户聊天数据导入模块、官方文件导入模块、客户发文导入模块和修改文件分类导入模块;

[0012] 所述第三方数据导入单元是将企业工作人员的第三方聊天应用软件均应用于该用于企业管理文件数据实时采集分析系统内,并使企业工作人员的第三方聊天应用软件通过企业的局域网内进行客户的沟通,所述客户聊天数据导入模块和客户发文导入模块用于将企业工作人员与客户之间在第三方应用沟通的聊天内容实时导入数据处理系统内,且官方文件导入模块用于将企业工作人员与官方之间在第三方应用来往的文件实时导入数据处理系统内。

[0013] 优选的,所述数据采集池是由n个数据采集模块组成,且n个数据采集模块能够对第三方数据导入单元导入的文件数据进行实时采集,所述母文件数据库单元是由n个项目存储模块组成,且每个项目存储模块内均存储有一个数据处理系统内预先立项创建好的文件数据模型,即为一个母文件数据模型。

[0014] 优选的,所述子文件数据模型创建单元包括文件类型归纳模块、文件逻辑关系生成模块、文件索引标识创建模块、文件索引标识整合模块和文件模型生成模块;

[0015] 通过所述文件类型归纳模块将采集模块采集的数据以企业工作人员编号和该工作人员名下立项的客户编号两个标识进行分类归纳处理得到若干个数据包,再通过文件逻辑关系生成模块识别与该文件相关联的其他关系人,并生成关系图谱,将生成的关系图谱数据合并至生成的对应数据包内,然后沟通过文件索引标识创建模块根据识别的企业工作人员的编号和该工作人员名下立项的客户编号创建成二进制代码,即生成索引符,再通过文件索引标识整合模块将生成的索引符整合到与之对应的该数据包上,之后通过文件模型生成模块生成一个子文件数据模型,之后相同操作进行实时数据处理生成若干个子文件数据模型,并将生成的若干个子文件数据模型依次发送至系统后台服务器。

[0016] 优选的,所述文件数据对比分析单元包括微处理模块、数据对比算法提取模块、文件数据模型分析单元、新文件数据模型生成模块、初始文件类型判别模块、数据实时更新模块和数据收发模块;

[0017] 所述微处理模块控制数据收发模块分别接受系统后台服务器发来的子文件数据模型以及母文件数据库单元内存储的母文件数据模型,先通过初始文件类型判别模块对接收的子文件数据模型的索引符与母文件数据库单元内存储的母文件数据模型索引符进行对比,来判定该子文件数据模型存储的聚焦位置,初始文件类型判别模块会向联网大数据数据库内提取索引符处理算法,来进行索引符对比处理,若初始判别出母文件数据库单元内无对应存储位置,则直接通过新文件数据模型生成模块在母文件数据库单元内生成新的项目存储模块来存储该子文件数据模型,并将该子文件数据模型的索引符前增加新创建新的存储编码,升级成母文件数据模型,即生成新的索引符,当初始判别出母文件数据库单元内有对应存储位置,则通过文件数据模型分析单元进行同文件数据模型内的同项目判断。

[0018] 优选的,所述文件数据模型分析单元包括子文件数据模型提取模块、母文件数据模型提取模块、数据类型对比分析模块和数据类型判定模块;

[0019] 通过所述子文件数据模型提取模块和母文件数据模型提取模块分别提取待对比的子文件数据模型和母文件数据模型,通过数据类型对比分析模块进行数据对比分析,数据类型对比分析模块在对比分析过程中会通过数据对比算法提取模块向联网大数据数据库内提取数据对比算法,来进行数据对比分析处理,然后通过数据类型判定模块判定该子文件数据模型是否为新立项项目。

[0020] 优选的,所述索引符处理算法具体包括以下步骤:

[0021] S1、子文件数据模型的索引符由以下三个标识组成:企业工作人员编号 $a=\{0,1,2,3,\dots,n\}$ 、该工作人员名下立项的客户编号 $b=\{0,1,2,3,\dots,n\}$ 以及该子文件数据模型内采集整合的数据标识代码 $X=\{X_1,X_2,X_3,\dots,X_n\}$ ,这三个标识进行分类归纳处理得到索引符 $abX$ ,在生成子文件数据模型时,企业工作人员编号 $a$ 是系统可知的确定数值,该工作人员名下立项的客户编号 $b$ 被随机赋值,但赋值过程中,其辅标识标记客户名称转换后的二级制代码,即为 $b(\Omega)$ ,而该子文件数据模型内采集整合的数据表代码 $X$ 直接随机赋值,即组合成该子文件数据模型的索引符为 $ab(\Omega)X$ ;

[0022] S2、初始文件类型判别模块会将步骤S1生成的子文件数据模型索引符发送至母文件数据库单元内与母文件数据库单元内各个项目存储模块上的索引符进行对比,先通过步

骤S1中确定好的编号a筛选出母文件数据库单元内与该企业工作人员所对应的母文件数据模型；

[0023] S3、根据步骤S1中该工作人员名下立项的客户编号 $b(\Omega)$ 中的辅标识标记 $\Omega$ 在步骤S2筛选出的母文件数据模型中进行搜索对比，若匹配到对应母文件数据模型，则通过文件数据模型分析单元内的数据对比算法进行立项识别；

[0024] S4、若未配到对应母文件数据模型，则直接通过新文件数据模型生成模块在母文件数据库单元内生成新的项目存储模块来存储该子文件数据模型，并将该子文件数据模型的索引符前增加新创建新的存储编码 $U=\{U_1, U_2, U_3, \dots, U_n\}$ ，升级成母文件数据模型，即生成新的索引符 $UabX$ 。

[0025] 优选的，所述数据对比算法具体包括以下步骤：

[0026] T1、先通过子文件数据模型提取模块和母文件数据模型提取模块分别提取待对比的子文件数据模型和母文件数据模型，再将子文件数据模型内采集整合的数据通过文字和图片识别算法进行特征提取，即对子文件数据模型内的标识代码 $X$ 内的内容进行处理，分别提取对应的标识特征 $z^1, z^2, z^3, \dots, z^n$ ；

[0027] T2、将步骤T2提取的标识特征与筛选出的母文件数据模型内各个立项内容特征进行匹配，匹配过程中计算匹配度 $W$ ，当匹配度 $W \geq 85\%$ 时，则判定匹配成功，表示该子文件模型为母文件模型中已经立项过的补充文件数据，直接将该子文件模型内的文件数据存储到对应的立项存储模块内，并将该子文件模型的数据标识代码 $X$ 变更为与该立项存储模块相对于的代码；

[0028] T3、当匹配度 $W < 85\%$ 时，则判定匹配失败，表示该子文件模型为母文件模型中新立项的补充文件数据，在该母文件数据模型内创建新的项目存储模块，该新的项目标识代码为现有标识码的相邻的下一位，并将该子文件模型内的文件数据存储到对应的新创建的立项存储模块内，并将该子文件模型的数据标识代码 $X$ 变更为与新立项存储模块相对于的数据标识代码。

[0029] 优选的，所述用户交互单元包括用户登录模块、文件模型删除模块、数据显示模块、文件模型修订模块、文件模型查询模块和文件模型转移模块。

[0030] 本发明还提供了一种采集终端，所述采集终端存储有机器可运行指令，所述计算机可运行指令在被处理器调用和运行时，所述计算机可运行指令促使所述处理器运行所述用于企业管理文件数据实时采集分析系统内的运行方法。

[0031] 本发明提供了用于企业管理文件数据实时采集分析系统及其采集终端。与现有技术相比具备以下有益效果：

[0032] (1)、该用于企业管理文件数据实时采集分析系统及其采集终端，可实现通过将企业用户与第三方聊天应用进行对接，实时采集和分析企业员工在第三方应用上与客户或者与官方来往邮箱的文件数据以及聊天内容，很好的达到了在短时间内进行文件数据立项分析运营的目的，无需采用人为整理导入的数据，也无需工作人员先在系统上建立数据订单，即创建立项数据包，然后导入处理系统内进行数据的分类和处理，使用简单，并且方便公司系统对员工的工作状态进行监控，同时很好的避免客户沟通处理的及时性，即使遇到业务员同时沟通多个客户，也能及时对沟通完成的客户的文件数据进行立项处理，从而大大提高了整个公司运营效率，从而大大方便了企业的文件数据管理。

[0033] (2)、该用于企业管理文件数据实时采集分析系统及其采集终端,可实现通过对子文件数据模型与母文件数据模型的索引符进行智能对比,来确定新文件数据模型的存储位置,很好的达到了既快速又方便的进行企业文件数据的智能整理的目的,实现自动判断立项内容是否为前立项的关联项目进行智能归档存储,可以根据具体变动内容进行自动更改或补充新数据,从而对企业文官数据管理工作十分有益。

## 附图说明

[0034] 图1为本发明系统的结构原理框图;

[0035] 图2为本发明数据处理系统的结构原理框图;

[0036] 图3为本发明用户交互单元的结构原理框图;

[0037] 图4为本发明第三方数据导入单元的结构原理框图;

[0038] 图5为本发明子文件数据模型创建单元的结构原理框图;

[0039] 图6为本发明文件数据对比分析单元的结构原理框图;

[0040] 图7为本发明文件数据模型分析单元的结构原理框图;

[0041] 图8为本发明系统工作方法的逻辑图。

[0042] 图中,1用户交互单元、11用户登录模块、12文件模型删除模块、13数据显示模块、14文件模型修订模块、15文件模型查询模块、16文件模型转移模块、2数据处理系统、21系统后台服务器、22第三方数据导入单元、221客户聊天数据导入模块、222官方文件导入模块、223客户发文导入模块、23数据采集池、24母文件数据库单元、25子文件数据模型创建单元、261文件类型归纳模块、252文件逻辑关系生成模块、253文件索引标识创建模块、254文件索引标识整合模块、255文件模型生成模块、26文件数据对比分析单元、261微处理模块、262数据对比算法提取模块、263文件数据模型分析单元、2631子文件数据模型提取模块、2632母文件数据模型提取模块、2633数据类型对比分析模块、2634数据类型判定模块、264新文件数据模型生成模块、265初始文件类型判别模块、266数据实时更新模块、267数据收发模块、3联网大数据库、4无线通讯模块、5系统总监控终端。

## 实施方式

[0043] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0044] 请参阅图1-8,本发明实施例提供两种技术方案:用于企业管理文件数据实时采集分析系统及其采集终端,具体包括以下实施例:

## 实施例

[0045] 用于企业管理文件数据实时采集分析系统,包括用户交互单元1、数据处理系统2、联网大数据库3以及通过无线通讯模块4与数据处理系统2实现数据无线交互的系统总监控终端5,系统总监控终端5能够通过监控终端进行远程监控各企业人员的工作内容和工作状态,用户通过用户交互单元1与数据处理系统2进行交互,且数据处理系统2能够向联网大数

数据库3内提取处理算法；

[0046] 数据处理系统2包括系统后台服务器21、第三方数据导入单元22、数据采集池23、母文件数据库单元24、子文件数据模型创建单元25和文件数据对比分析单元26,通过第三方数据导入单元22将企业工作人员与客户沟通的聊天记录和来往文件数据导入数据处理系统2内,再通过数据采集池23对导入的数据进行实时采集,采集的数据通过子文件数据模型创建单元25创建出带有索引符的文件数据子模型,然后通过系统后台服务器21将创建的带有索引符的文件数据子模型发送至文件数据对比分析单元26内进行对比分析,来对导入的数据进行归纳存储或创建新项目进行存储。

[0047] 由图4所示,本发明实施例中,第三方数据导入单元22包括客户聊天数据导入模块221、官方文件导入模块222和客户发文导入模块223;

[0048] 第三方数据导入单元22是将企业工作人员的第三方聊天应用软件均应用于该用于企业管理文件数据实时采集分析系统内,并使企业工作人员的第三方聊天应用软件通过企业的局域网内进行客户的沟通,客户聊天数据导入模块221和客户发文导入模块223用于将企业工作人员与客户之间在第三方应用沟通的聊天内容实时导入数据处理系统2内,且官方文件导入模块222用于将企业工作人员与官方之间在第三方应用来往的文件实时导入数据处理系统2内。

[0049] 本实施例中,企业工作人员使用的第三方聊天应用为QQ、微信、ICQ或MSN中的一种,直接让这些第三方应用软件在本发明的用于企业管理文件数据实时采集分析系统上登录和运行,并且能够获得这些第三方聊天应用软件的数据获取权限和通讯录权限,明显运行于本发明系统上的这些第三方聊天软件所登录的帐号为企业的工作帐号,不能使企业人员的个人隐私帐号,并且本发明中的第三方数据导入单元22是内置有数据自动获取上传模块,能够设置上传时间,例如一天或一周等,进行一段时间内数据的上传,从而能够将企业人员与客户沟通的聊天记录和文件数据一起自动打包上传。

[0050] 由图2所示,本发明实施例中,数据采集池23是由n个数据采集模块组成,且n个数据采集模块能够对第三方数据导入单元22导入的文件数据进行实时采集,母文件数据库单元24是由n个项目存储模块组成,且每个项目存储模块内均存储有一个数据处理系统2内预先立项创建好的文件数据模型,即为一个母文件数据模型,一个母文件数据模型内的有若干个立项存储模块,每个立项存储模块内均存储一个该企业人员对应该客户的一个立项订单。

[0051] 由图5所示,本发明实施例中,子文件数据模型创建单元25包括文件类型归纳模块251、文件逻辑关系生成模块252、文件索引标识创建模块253、文件索引标识整合模块254和文件模型生成模块255;

[0052] 通过文件类型归纳模块251将采集模块采集的数据以企业工作人员编号和该工作人员名下立项的客户编号两个标识进行分类归纳处理得到若干个数据包,再通过文件逻辑关系生成模块252识别与该文件相关联的其他关系人,并生成关系图谱,将生成的关系图谱数据合并至生成的对应数据包内,然后沟通过文件索引标识创建模块253根据识别的企业工作人员的编号和该工作人员名下立项的客户编号创建成二进制代码,即生成索引符,再通过文件索引标识整合模块254将生成的索引符整合到与之对应的该数据包上,之后通过文件模型生成模块255生成一个子文件数据模型,之后相同操作进行实时数据处理生成若

干个子文件数据模型,并将生成的若干个子文件数据模型依次发送至系统后台服务器21。

[0053] 由图6所示,本发明实施例中,文件数据对比分析单元26包括微处理模块261、数据对比算法提取模块262、文件数据模型分析单元263、新文件数据模型生成模块264、初始文件类型判别模块265、数据实时更新模块266和数据收发模块267,数据实时更新模块266能够对整个文件数据对比分析单元26内的工作数据进行实时更新;

[0054] 微处理模块261控制数据收发模块267分别接受系统后台服务器21发来的子文件数据模型以及母文件数据库单元24内存储的母文件数据模型,先通过初始文件类型判别模块265对接收的子文件数据模型的索引符与母文件数据库单元24内存储的母文件数据模型索引符进行对比,来判定该子文件数据模型存储的聚焦位置,聚焦位置为仅仅包含企业工作人员编号a和该工作人员名下立项的客户编号b组成的索引符的位置,初始文件类型判别模块265会向联网大数据库3内提取索引符处理算法,来进行索引符对比处理,若初始判别出母文件数据库单元24内无对应存储位置,则直接通过新文件数据模型生成模块264在母文件数据库单元24内生成新的项目存储模块来存储该子文件数据模型,并将该子文件数据模型的索引符前增加新创建新的存储编码,升级成母文件数据模型,即生成新的索引符,当初始判别出母文件数据库单元24内有对应存储位置,则通过文件数据模型分析单元263进行同文件数据模型内的同项目判断。

[0055] 由图7所示,本发明实施例中,文件数据模型分析单元263包括子文件数据模型提取模块2631、母文件数据模型提取模块2632、数据类型对比分析模块2633和数据类型判定模块2634;

[0056] 通过子文件数据模型提取模块2631和母文件数据模型提取模块2632分别提取待对比的子文件数据模型和母文件数据模型,通过数据类型对比分析模块2633进行数据对比分析,数据类型对比分析模块在对比分析过程中会通过数据对比算法提取模块262向联网大数据库3内提取数据对比算法,来进行数据对比分析处理,然后通过数据类型判定模块2634判定该子文件数据模型是否为新立项项目。

[0057] 本发明实施例中,索引符处理算法具体包括以下步骤:

[0058] S1、子文件数据模型的索引符由以下三个标识组成:企业工作人员编号 $a=\{0,1,2,3,\dots,n\}$ 、该工作人员名下立项的客户编号 $b=\{0,1,2,3,\dots,n\}$ 以及该子文件数据模型内采集整合的数据标识代码 $X=\{X_1,X_2,X_3,\dots,X_n\}$ ,这三个标识进行分类归纳处理得到索引符 $abX$ ,在生成子文件数据模型时,企业工作人员编号a是系统可知的确定数值,该工作人员名下立项的客户编号b被随机赋值,但赋值过程中,其辅标识标记客户名称转换后的二级制代码,即为 $b(\Omega)$ ,而该子文件数据模型内采集整合的数据表代码X直接随机赋值,即组合成该子文件数据模型的索引符为 $ab(\Omega)X$ ;

[0059] S2、初始文件类型判别模块265会将步骤S1生成的子文件数据模型索引符发送至母文件数据库单元24内与母文件数据库单元24内各个项目存储模块上的索引符进行对比,先通过步骤S1中确定好的编号a筛选出母文件数据库单元24内与该企业工作人员所对应的母文件数据模型;

[0060] S3、根据步骤S1中该工作人员名下立项的客户编号 $b(\Omega)$ 中的辅标识标记 $\Omega$ 在步骤S2筛选出的母文件数据模型中进行搜索对比,若匹配到对应母文件数据模型,则通过文件数据模型分析单元263内的数据对比算法进行立项识别;

[0061] S4、若未配到对应母文件数据模型,则直接通过新文件数据模型生成模块264在母文件数据库单元24内生成新的项目存储模块来存储该子文件数据模型,并将该子文件数据模型的索引符前增加新创建新的存储编码 $U=\{U_1, U_2, U_3, \dots, U_n\}$ ,升级成母文件数据模型,即生成新的索引符 $UabX$ 。

[0062] 本发明实施例中,数据对比算法具体包括以下步骤:

[0063] T1、先通过子文件数据模型提取模块2631和母文件数据模型提取模块2632分别提取待对比的子文件数据模型和母文件数据模型,再将子文件数据模型内采集整合的数据通过文字和图片识别算法进行特征提取,即对子文件数据模型内的标识代码 $X$ 内的内容进行处理,分别提取对应的标识特征 $z^1, z^2, z^3, \dots, z^n$ 。

[0064] T2、将步骤T2提取的标识特征与筛选出的母文件数据模型内各个立项内容特征进行匹配,匹配过程中计算匹配度 $W$ ,当匹配度 $W \geq 85\%$ 时,则判定匹配成功,表示该子文件模型为母文件模型中已经立项过的补充文件数据,直接将该子文件模型内的文件数据存储到对应的立项存储模块内,并将该子文件模型的数据标识代码 $X$ 变更为与该立项存储模块相对于的代码;

[0065] T3、当匹配度 $W < 85\%$ 时,则判定匹配失败,表示该子文件模型为母文件模型中新立项的补充文件数据,在该母文件数据模型内创建新的项目存储模块,该新的项目标识代码为现有标识码的相邻的下一位,并将该子文件模型内的文件数据存储到对应的新创建的立项存储模块内,并将该子文件模型的数据标识代码 $X$ 变更为与新立项存储模块相对于的数据标识代码。

[0066] 本发明实施例还提供了一种采集终端,所述采集终端存储有机器可运行指令,所述计算机可运行指令在被处理器调用和运行时,所述计算机可运行指令促使所述处理器运行用于企业管理文件数据实时采集分析系统内的运行方法。

[0067] 本发明实施例中,使用的文字和图片识别算法为现有的OCR算法,具体包括以下步骤:

[0068] 1、图像输入、预处理:图像输入:对于不同的图像格式,有着不同的存储格式,不同的压缩方式。预处理:主要包括二值化,噪声去除,倾斜较正等

[0069] 二值化:对摄像头拍摄的图片,大多数是彩色图像,彩色图像所含信息量巨大,对于图片的内容,我们可以简单的分为前景与背景,为了让计算机更快的,更好的识别文字,我们需要先对彩色图进行处理,使图片只前景信息与背景信息,可以简单的定义前景信息为黑色,背景信息为白色,这就是二值化图了。

[0070] 噪声去除:对于不同的文档,我们对燥声的定义可以不同,根据燥声的特征进行去燥,就叫做噪声去除。

[0071] 倾斜较正:由于一般用户,在拍照文档时,都比较随意,因此拍照出来的图片不可避免的产生倾斜,这就需要文字识别软件进行较正。

[0072] 版面分析:将文档图片分段落,分行的过程就叫做版面分析,由于实际文档的多样性,复杂性,因此,目前还没有一个固定的,最优的切割模型。

[0073] 字符切割:由于拍照条件的限制,经常造成字符粘连,断笔,因此极大限制了识别系统的性能,这就需要文字识别软件有字符切割功能。

[0074] 字符识别:已经是很早的事情了,比较早有模板匹配,后来以特征提取为主,由于

文字的位移,笔画的粗细,断笔,粘连,旋转等因素的影响,极大影响特征的提取的难度。

[0075] 文件的识别:对数据中非图片格式的其他文件进行识别和提取。

[0076] 9、后处理、校对:根据特定的语言上下文的关系,对识别结果进行校正,就是后处理。

### 实施例

[0077] 由图3所示,本发明实施例的技术方案区别于实施例1的技术方案包括:用户交互单元1包括用户登录模块11、文件模型删除模块12、数据显示模块13、文件模型修订模块14、文件模型查询模块15和文件模型转移模块16。

[0078] 用户登录模块11用于用户登录整个系统,用户进入系统后可分别通过文件模型删除模块12、数据显示模块13、文件模型修订模块14、文件模型查询模块15和文件模型转移模块16进行文件的删除、显示、修订、查询和文件模型转移至其他非高于本用户权限级别的用户。

[0079] 以某知识产权服务企业为例,该企业的员工A在本发明的用于企业管理文件数据实时采集分析系统上注册有员工编号3,即 $a=3$ ,员工A使用公司的企业微信帐号登录,并与微信中的客户B进行知识产权业务沟通,沟通的内容以及客户B从微信上发给员工A的文件均会上传至系统内,若客户B之前并未与本公司合作,并且为员工A的第5个客户,即 $b=5$ ,则系统内会在母文件数据库单元24内新建母文件数据模型 $U_1$ 并建立一个立项 $X_1$ ,并给员工A生成索引符 $U_135X_1$ ,当该客户B后期准备想新增立项时,通过与员工A微信沟通后,系统自动实时在员工A母文件数据库单元24的建母文件数据模型 $U_1$ 内新建立一个立项 $X_2$ ,即新建的索引符为 $U_135X_2$ 。

[0080] 同时本说明书中未作详细描述的内容均属于本领域技术人员公知的现有技术。

[0081] 需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。

[0082] 尽管已经示出和描述了本发明的实施例,对于本领域的普通技术人员而言,可以理解在不脱离本发明的原理和精神的情况下可以对这些实施例进行多种变化、修改、替换和变型,本发明的范围由所附权利要求及其等同物限定。

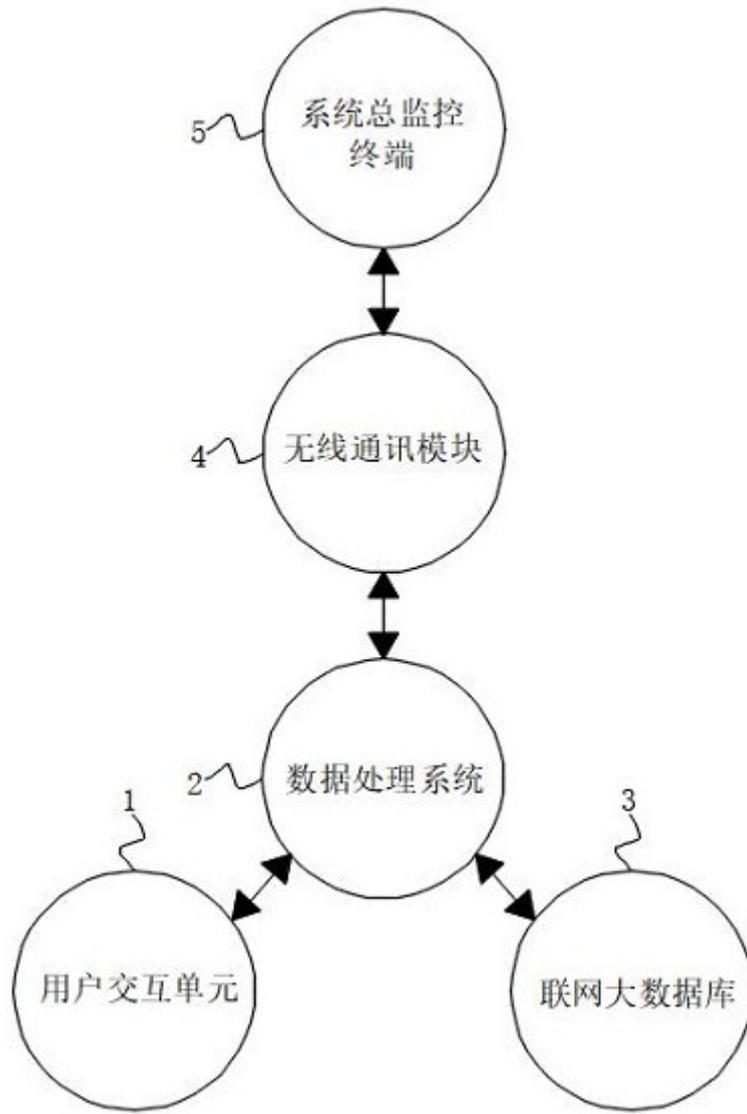


图1

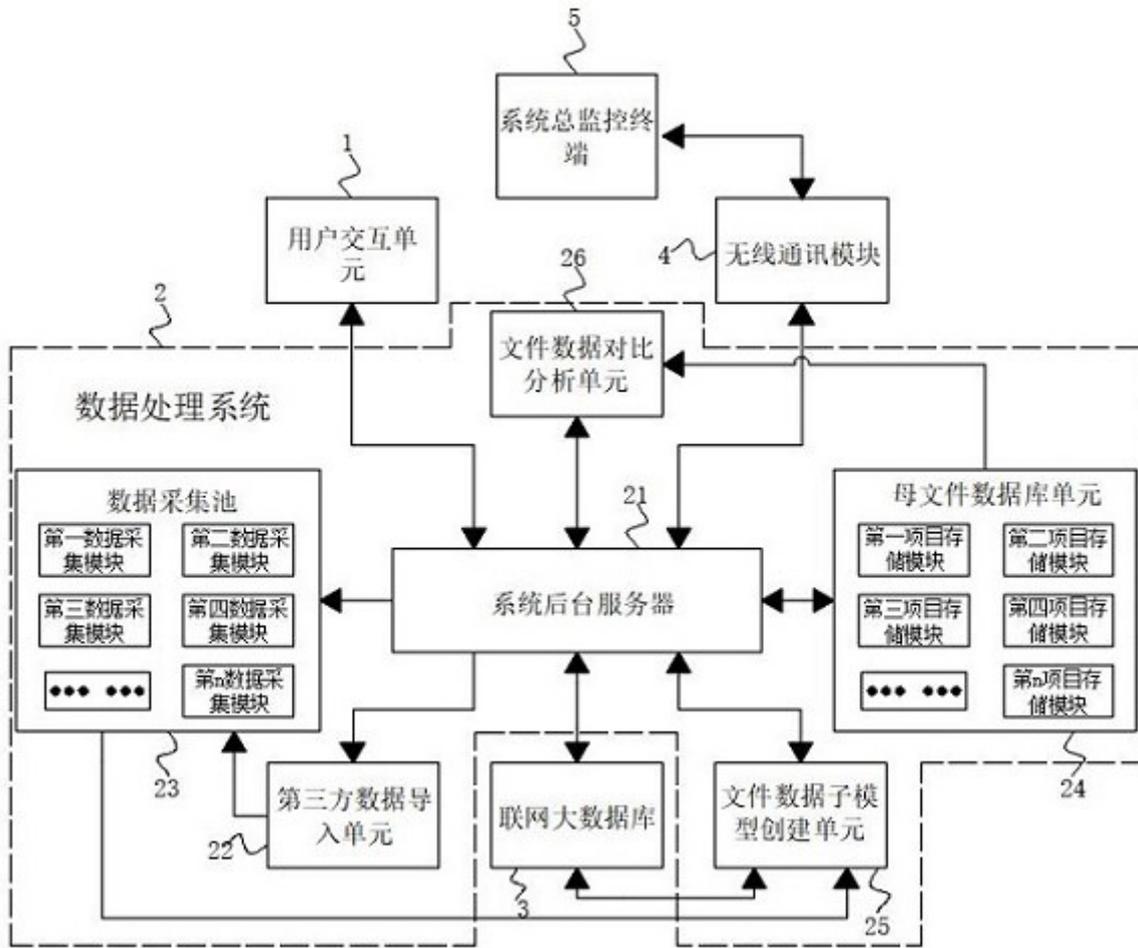


图2

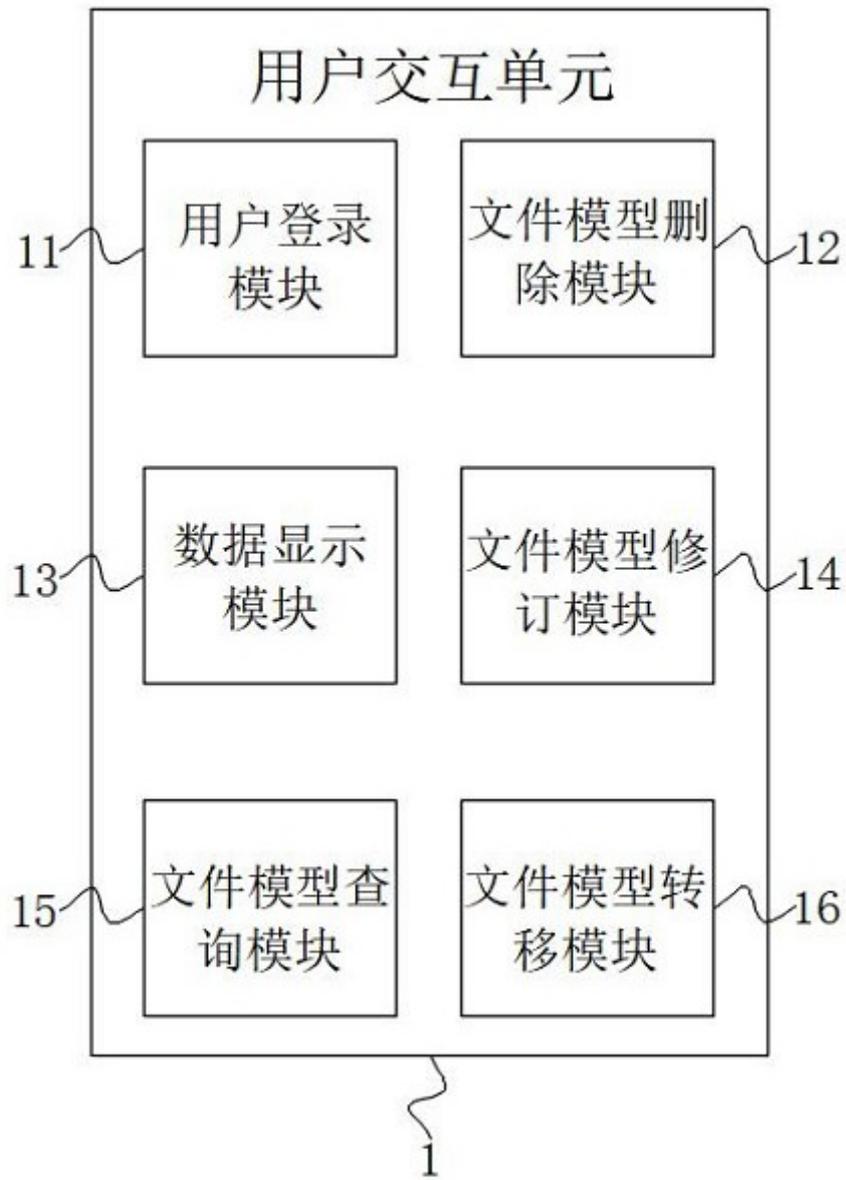


图3



图4

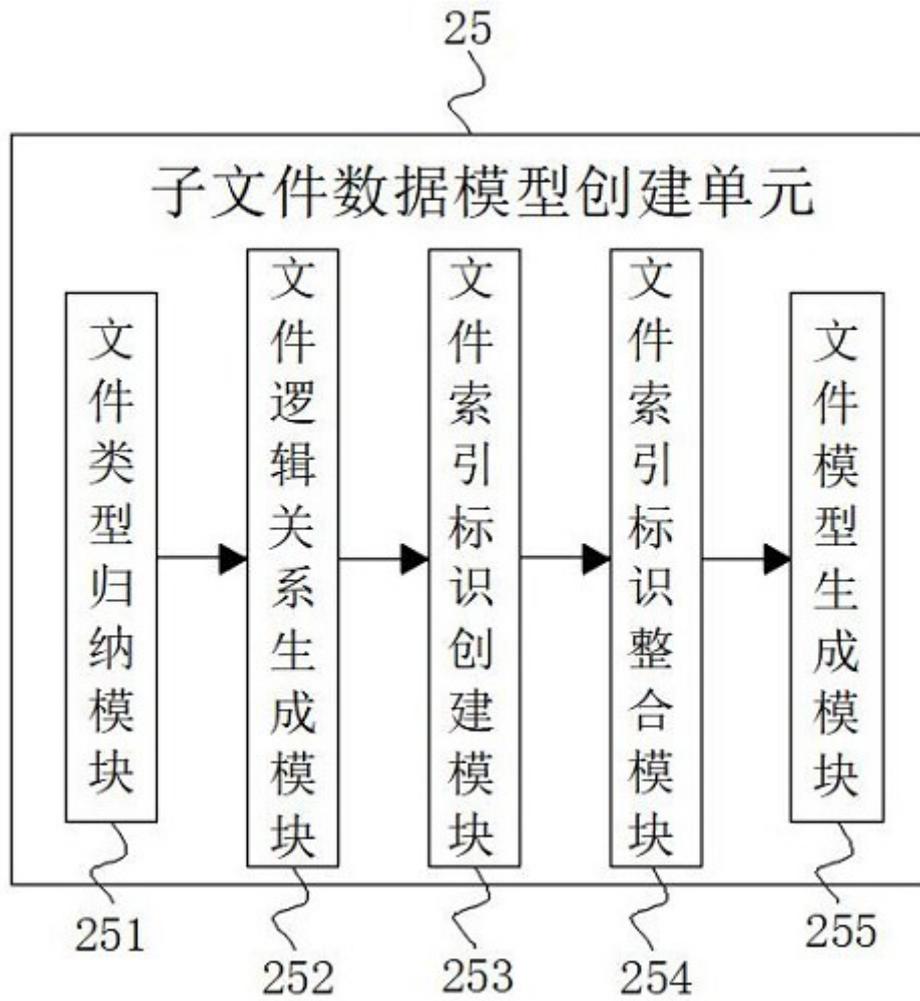


图5

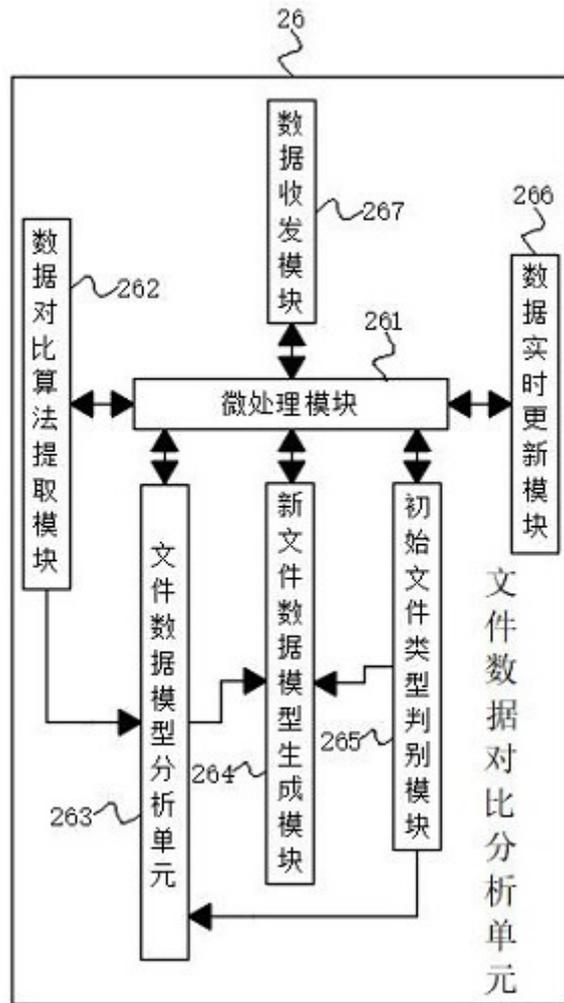


图6

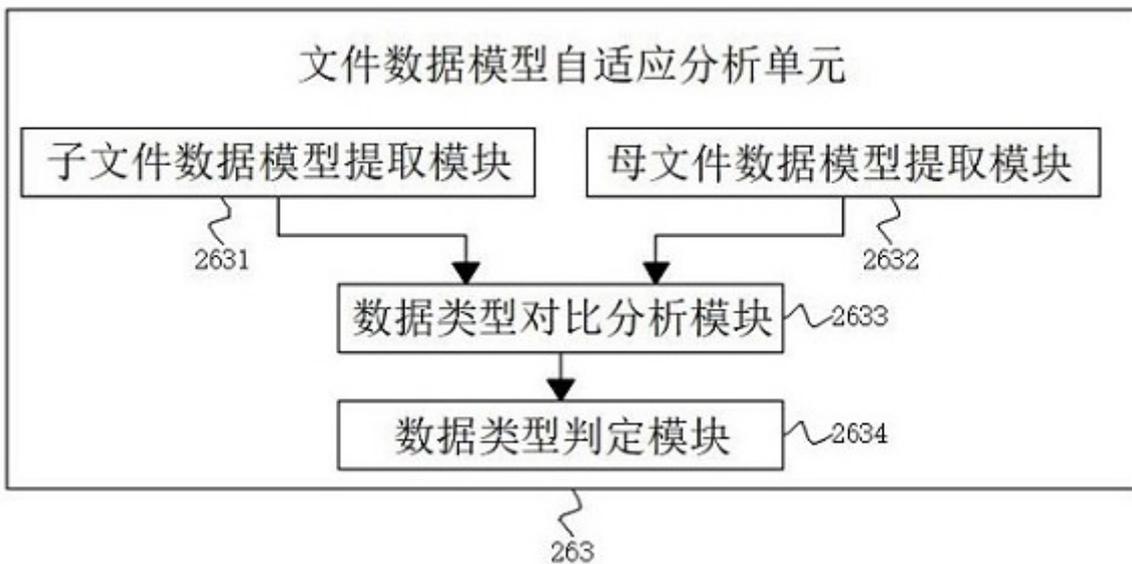


图7

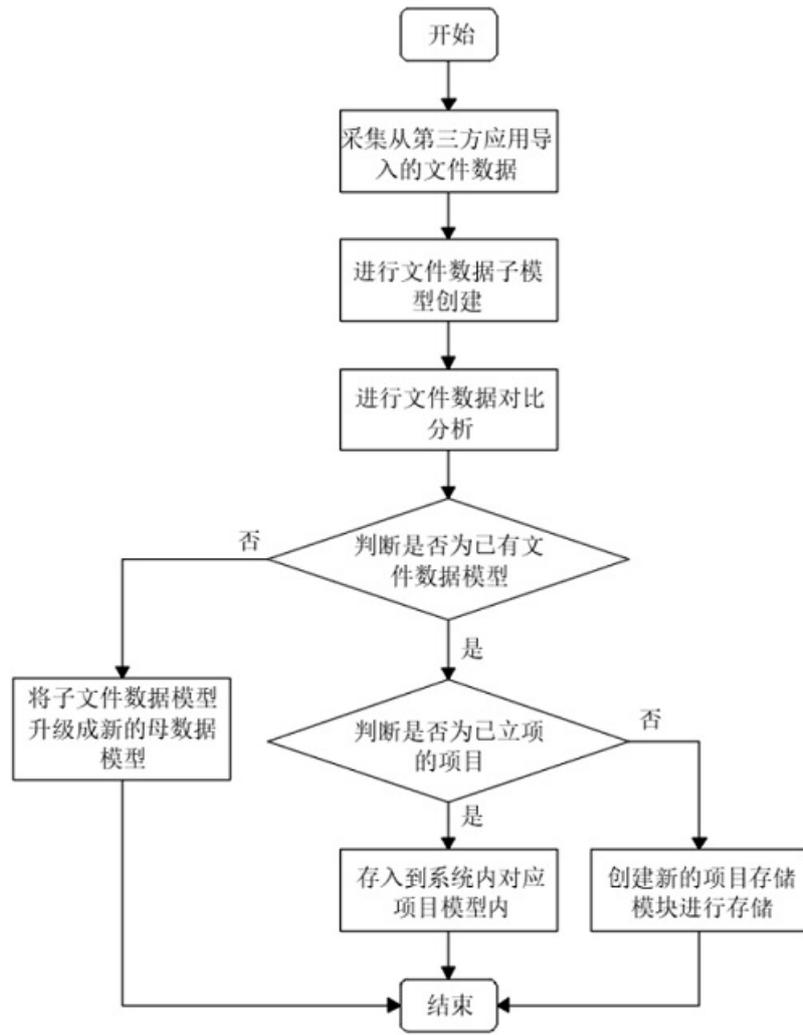


图8