



(12) 发明专利

(10) 授权公告号 CN 110851506 B

(45) 授权公告日 2021.12.03

(21) 申请号 201810826049.0

(22) 申请日 2018.07.25

(65) 同一申请的已公布的文献号  
申请公布号 CN 110851506 A

(43) 申请公布日 2020.02.28

(73) 专利权人 上海柯林布瑞信息技术有限公司  
地址 200233 上海市徐汇区宜山路700号普  
天信息产业园B2栋1504A

(72) 发明人 秦晓宏 刘焕春

(74) 专利代理机构 上海思捷知识产权代理有限  
公司 31295

代理人 王宏婧

(51) Int. Cl.

G06F 16/248 (2019.01)

G16H 50/70 (2018.01)

(56) 对比文件

CN 104123390 A, 2014.10.29

CN 104123390 A, 2014.10.29

CN 102289467 A, 2011.12.21

CN 105045848 A, 2015.11.11

CN 106709652 A, 2017.05.24

CN 106503274 A, 2017.03.15

US 2004006555 A1, 2004.01.08

US 2004267765 A1, 2004.12.30

KR 20180046763 A, 2018.05.09

审查员 王玮

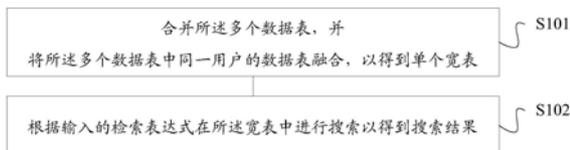
权利要求书2页 说明书8页 附图3页

(54) 发明名称

临床大数据的搜索方法及装置、存储介质、  
服务器

(57) 摘要

一种临床大数据的搜索方法及装置、存储介  
质、服务器,所述临床大数据包含多个数据表,所  
述临床大数据的搜索方法包括:合并所述多个数  
据表,并将所述多个数据表中同一用户的数据表  
融合,以得到单个宽表;根据输入的检索表达式  
在所述宽表中进行搜索以得到搜索结果。通过本  
发明提供的技术方案,可以加快搜索速度,提高  
数据查询效率。



1. 一种临床大数据的搜索方法,所述临床大数据包含多个数据表,其特征在于,所述临床大数据为关系型数据库,所述数据表为可以互相联接的二维行列表格,所述临床大数据的搜索方法包括:

合并所述多个数据表,所述多个数据表是以病人为中心因医疗就诊而被分别生成,并将所述多个数据表中同一用户的数据表融合,以得到单个宽表,包括:为用户建立主索引;对所述多个数据表中的用户标识数据进行分析,确定关联至同一用户的各个所述用户标识数据所在的数据表;将属于同一用户的所述多个数据表融合;删除所述多个数据表中表示用户的各个所述用户标识数据,采用所述主索引作为唯一所述用户标识;对数据进行处理,滤除各个用户相关标识数据中的重复数据,包括多余的姓名、性别和年龄信息,将所述多个数据表中的多个值填入所述宽表的单个字段中,包括疾病信息相同,但诊断信息、用药信息不同的相关数据;

根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果,所述检索表达式包括检索关键词,或者,所述检索表达式包括检索关键词和逻辑运算符。

2. 根据权利要求1所述的搜索方法,其特征在于,所述对所述多个数据表中的数据进行分析,以确定属于同一用户的各个数据表包括:

对所述多个数据表中的用户标识数据进行分析,如果不同的数据表中的用户标识数据关联至同一用户,则将所述不同的数据表确定为属于同一用户。

3. 根据权利要求2所述的搜索方法,其特征在于,所述将属于同一用户的各个数据表融合包括:

将属于同一用户的各个数据表融合,以得到原始宽表;

对所述原始宽表进行过滤,去除重复的数据,以得到所述宽表。

4. 根据权利要求1至3任一项所述的搜索方法,其特征在于,还包括:

根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果时,判断所述检索表达式中的检索关键词是否包含在用于分词的分词词典中;

如果所述检索关键词未包含于所述分词词典中,则将所述检索关键词的输入次数进行递增;

如果所述检索关键词的输入次数达到预设阈值,则将所述检索关键词加入所述分词词典。

5. 根据权利要求1至3任一项所述的搜索方法,其特征在于,还包括:

对所述搜索结果进行统计,以得到统计结果;

在呈现所述搜索结果时,以图表的形式一并呈现所述统计结果。

6. 一种临床大数据的搜索装置,所述临床大数据包含多个数据表,其特征在于,所述临床大数据为关系型数据库,所述数据表为可以互相联接的二维行列表格,所述临床大数据的搜索装置包括:

确定模块,适于合并所述多个数据表,所述多个数据表是以病人为中心因医疗就诊而被分别生成,并将所述多个数据表中同一用户的数据表融合,以得到单个宽表,包括:为用户建立主索引;所述确定模块包括分析子模块和融合子模块,所述分析子模块用于对所述多个数据表中的用户标识数据进行分析,确定关联至同一用户的各个所述用户标识数据所在的数据表;所述融合子模块用于将属于同一用户的所述多个数据表融合;删除所述多个

数据表中表示用户的各个所述用户标识数据,采用所述主索引作为唯一所述用户标识;所述融合子模块包括清洗单元,所述清洗单元用于对数据进行处理,滤除各个用户相关标识数据中的重复数据,包括多余的姓名、性别和年龄信息,将所述多个数据表中的多个值填入所述宽表的单个字段中,包括疾病信息相同,但诊断信息、用药信息不同的相关数据;

搜索模块,适于根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果,所述检索表达式包括检索关键词,或者,所述检索表达式包括检索关键词和逻辑运算符。

7. 根据权利要求6所述的搜索装置,其特征在于,所述分析子模块包括:

分析单元,适于对所述多个数据表中的用户标识数据进行分析,如果不同的数据表中的用户标识关联至同一用户,则将所述不同的数据表确定为属于同一用户。

8. 根据权利要求6所述的搜索装置,其特征在于,所述融合子模块还包括融合单元;

所述融合单元,适于将属于同一用户的各个数据表融合,以得到原始宽表;

所述清洗单元,还适于对所述原始宽表进行过滤,去除重复的数据,以得到所述宽表。

9. 根据权利要求6至8任一项所述的搜索装置,其特征在于,还包括:

判断模块,适于根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果时,判断所述检索表达式中的检索关键词是否包含在用于分词的分词词典中;

计算模块,如果所述检索关键词未包含于所述分词词典中,则所述计算模块适于将所述检索关键词的输入次数进行递增;

加入模块,如果所述检索关键词的输入次数达到预设阈值,则所述加入模块适于将所述检索关键词加入所述分词词典。

10. 根据权利要求6至8任一项所述的搜索装置,其特征在于,还包括:

统计模块,对所述搜索结果进行统计,以得到统计结果;

呈现模块,适于在呈现所述搜索结果时,以图表的形式一并呈现所述统计结果。

11. 一种存储介质,其上存储有计算机指令,其特征在于,所述计算机指令运行时执行权利要求1至5任一项所述的临床大数据的搜索方法的步骤。

12. 一种服务器,包括存储器和处理器,所述存储器上存储有可在所述处理器上运行的计算机指令,其特征在于,所述处理器运行所述计算机指令时执行权利要求1至5中任一项所述的临床大数据的搜索方法的步骤。

## 临床大数据的搜索方法及装置、存储介质、服务器

### 技术领域

[0001] 本发明涉及大数据搜索技术领域,具体地涉及一种临床大数据的搜索方法及装置、存储介质、服务器。

### 背景技术

[0002] 全文搜索引擎是目前广泛应用的主流搜索引擎。其工作原理是计算机索引程序通过扫描数据库存储的文本中的每个词语,并对每个词语建立索引,指明该词语在所述文本中出现的次数和位置。当用户搜索时,可以根据事先建立的索引进行查找,并将查找结果反馈给用户。

[0003] 现有的数据库(例如,关系型数据库)对全文搜索的支持不够完善,例如,对语义处理的支持不完善,往往导致搜索时间过长,降低搜索性能。即使简单的文本查询,也可能持续1至2分钟,甚至出现检索不到结果的情况。

[0004] 对于医院系统而言,其中存储有大量以病人为中心的临床数据。如果海量临床数据基于现有的数据库存储,则采用全文搜索引擎搜索时,其搜索速度缓慢,搜索效率低下。

### 发明内容

[0005] 本发明解决的技术问题是如何快速从临床大数据中得到搜索结果,以提高数据查询效率。

[0006] 为解决上述技术问题,本发明实施例提供一种临床大数据的搜索方法,所述临床大数据包含多个数据表,所述临床大数据的搜索方法包括:合并所述多个数据表,并将所述多个数据表中同一用户的数据表融合,以得到单个宽表;根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果。

[0007] 可选的,所述将所述多个数据表中同一用户的数据表融合包括:对所述多个数据表中的数据进行分析,以确定属于同一用户的各个数据表;将属于同一用户的各个数据表融合。

[0008] 可选的,所述对所述多个数据表中的数据进行分析,以确定属于同一用户的各个数据表包括:对所述多个数据表中的用户标识数据进行分析,如果不同的数据表中的用户标识数据关联至同一用户,则将所述不同的数据表确定为属于同一用户。

[0009] 可选的,所述将属于同一用户的各个数据表融合包括:将属于同一用户的各个数据表融合,以得到原始宽表;对所述原始宽表进行过滤,去除重复的数据,以得到所述宽表。

[0010] 可选的,所述临床大数据的搜索方法还包括:根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果时,判断所述检索表达式中的检索关键词是否包含在用于分词的分词词典中;如果所述检索关键词未包含于所述分词词典中,则将所述检索关键词的输入次数进行递增;如果所述检索关键词的输入次数达到预设阈值,则将所述检索关键词加入所述分词词典。

[0011] 可选的,所述临床大数据的搜索方法还包括:对所述搜索结果进行统计,以得到统

计结果;在呈现所述搜索结果时,以图表的形式一并呈现所述统计结果。

[0012] 为解决上述技术问题,本发明实施例还提供一种临床大数据的搜索装置,所述临床大数据包含多个数据表,所述搜索装置包括:确定模块,适于合并所述多个数据表,并将所述多个数据表中同一用户的数据表融合,以得到单个宽表;搜索模块,适于根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果。

[0013] 可选的,所述确定模块包括:分析子模块,对所述多个数据表中的数据进行分析,以确定属于同一用户的各个数据表;融合子模块,将属于同一用户的各个数据表融合。

[0014] 可选的,所述分析子模块包括:分析单元,适于对所述多个数据表中的用户标识数据进行分析,如果不同数据表中的用户标识关联至同一用户,则将所述不同的数据表确定为属于同一用户。

[0015] 可选的,所述融合子模块包括:融合单元,适于将属于同一用户的各个数据表融合,以得到原始宽表;清洗单元,适于对所述原始宽表进行过滤,去除重复的数据,以得到所述宽表。

[0016] 可选的,所述临床大数据的搜索装置还包括:判断模块,适于根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果时,判断所述检索表达式中的检索关键词是否包含在用于分词的分词词典中;计算模块,如果所述检索关键词未包含于所述分词词典中,则所述计算模块适于将所述检索关键词的输入次数进行递增;加入模块,如果所述检索关键词的输入次数达到预设阈值,则所述加入模块适于将所述检索关键词加入所述分词词典。

[0017] 可选的,所述临床大数据的搜索装置还包括:统计模块,对所述搜索结果进行统计,以得到统计结果;呈现模块,适于在呈现所述搜索结果时,以图表的形式一并呈现所述统计结果。

[0018] 为解决上述技术问题,本发明实施例还提供一种存储介质,其上存储有计算机指令,所述计算机指令运行时执行上述临床大数据的搜索方法的步骤。

[0019] 为解决上述技术问题,本发明实施例还提供一种服务器,包括存储器和处理器,所述存储器上存储有可在所述处理器上运行的计算机指令,所述处理器运行所述计算机指令时执行上述临床大数据的搜索方法的步骤。

[0020] 与现有技术相比,本发明实施例的技术方案具有以下有益效果:

[0021] 本发明实施例提供一种临床大数据的搜索方法,所述临床大数据包含多个数据表,所述搜索方法包括:将所述多个数据表中同一用户的数据表融合,以得到单个宽表;根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果。通过本发明实施例提供的技术方案,可以将属于同一用户的多个数据表融合,之后可以在单个宽表中搜索,搜索时无需查找多个不同的数据表,可以加快搜索速度,提高数据查询效率。

[0022] 进一步,将属于同一用户的各个数据表融合,以得到原始宽表;对所述原始宽表进行过滤,去除重复的数据,以得到所述宽表。通过本发明实施例提供的技术方案,可以去掉重复的无用数据,既节省存储空间,又有利于加快搜索速度。

[0023] 进一步,根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果时,判断所述检索表达式中的检索关键词是否包含在用于分词的分词词典中;如果所述检索关键词未包含于所述分词词典中,则将所述检索关键词的输入次数进行递增;如果所述检索关键词的输入次数达到预设阈值,则将所述检索关键词加入所述分词词典。通过本发明实施例

提供的技术方案,可以快速发现新词语,并将所述新词语加入至所述分词词典中,能够为后续搜索节约时间成本,加快搜索速度。

[0024] 进一步,对所述搜索结果进行统计,以得到统计结果;在呈现所述搜索结果时,以图表的形式一并呈现所述统计结果。通过本发明实施例提供的技术方案,可以将搜索结果的统计结果以图表形式呈现,可以使搜索结果的显示更加直观、清晰,进而提高用户体验。

### 附图说明

[0025] 图1是本发明实施例的一种临床大数据的搜索方法的流程示意图;

[0026] 图2是本发明实施例的又一种临床大数据的搜索方法的流程示意图;

[0027] 图3是本发明实施例的再一种临床大数据的搜索方法的流程示意图;

[0028] 图4是本发明实施例的一种临床大数据的搜索装置的结构示意图。

### 具体实施方式

[0029] 本领域技术人员理解,如背景技术所言,现有搜索方法对海量数据搜索的支持力度较差,难以满足快速搜索的用户需求。

[0030] 目前的全文搜索方法主要包括按字检索和按词检索。按字检索指的是,对于给定的检索内容,为每个字建立索引,并根据字的组合进行搜索。按词检索指的是,对于给定的检索内容,按照语义单位(即词语)建立索引,并基于语义单位进行搜索。

[0031] 对于英文等文字,字与词是相同的,可以按照空白切分子或词语,按字检索与按词检索的处理方式没有区别。对于中文等文字,则需要根据语义分词确定词语,按字检索与按词检索的处理方式区别很大。一般而言,按词检索的技术难度偏大。

[0032] 通常情况下,临床大数据往往采用关系型数据库存储。所述关系型数据库指的是建立在关系模型基础上的数据库,一般由多张可以互相联接的二维行列表格组成。其中,每行表格可以包含一条临床数据。每条临床数据可以包括病人的姓名、年龄、患病信息、化验项目、用药记录等。如果临床大数据存储在关系型数据库中,则搜索其中一条临床数据,意味着在关系型数据库中搜索,其搜索时间往往较长。

[0033] 本发明实施例提供一种临床大数据的搜索方法,所述临床大数据包含多个数据表,所述搜索方法包括:合并所述多个数据表,并将所述多个数据表中同一用户的数据表融合,以得到单个宽表;根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果。通过本发明实施例提供的技术方案,可以将属于同一用户的多个数据表融合,之后可以在所述宽表中搜索,搜索时无需查找多个不同的数据表,可以加快搜索速度,提高数据查询效率。

[0034] 为使本发明的上述目的、特征和有益效果能够更为明显易懂,下面结合附图对本发明的具体实施例做详细的说明。

[0035] 图1是本发明实施例的一种临床大数据的搜索方法的流程示意图。所述临床大数据包含多个数据表,每个数据表中包含多条临床数据。参考图1,所述搜索方法可以包括以下步骤:

[0036] 步骤S101:合并所述多个数据表,并将所述多个数据表中同一用户的数据表融合,以得到单个宽表;

[0037] 步骤S102:根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果。

[0038] 具体而言,在步骤S101中,可以融合多个数据表中属于同一用户的数据表,以得到该用户的宽表。

[0039] 更具体而言,数据表中可以包括用户标识数据。每一用户标识数据可以指向唯一的一个用户。不过,单个用户可以具有多种不同的用户标识数据,也即不同的用户标识数据可以属于同一用户。当不同的用户标识数据属于同一用户时,可以将不同的用户标识数据关联至同一用户。例如,用户可以和主索引一一对应,不同用户可以采用不同主索引表示,同一主索引可以对应同一用户的多个用户标识数据。同一主索引中,不同的用户标识数据属于同一用户。

[0040] 本领域技术人员理解,通过主索引,可以将不同的用户标识数据关联至同一用户。其中,所述主索引可以采用身份证号码、手机号码或其他可以唯一标识该用户的字符表示。例如,主索引可以是从小1开始的自然数,不同自然数表示不同用户。

[0041] 作为一个非限制性的实施例,病人A就诊时可以采用医保卡挂号,该病人A的用户标识数据可以为医保卡号码。或者,该病人A就诊时也可以采用临时病历卡挂号,该病人A的用户标识数据可以为临时病历卡号码。又或者,该病人A就诊时也可以采用身份证挂号,该病人A的用户标识数据可以为身份证号码。通常情况下,尽管病人A的医保卡号码与病人A的临时病例卡号码不同,但是如果已知该病人A的身份证号码,且通过身份证号码可以表明采用医保卡和/或临时病例卡就诊的病人即为病人A,则可以为病人A建立主索引(例如,采用身份证号码作为主索引),以将不同用户标识数据关联至病人A。

[0042] 作为又一个非限制性的实施例,病人A就诊时可以采用医保卡挂号,该病人A的用户标识数据可以为医保卡号码;或者,病人A也可以采用身份证挂号,此时病人A的用户标识数据为身份证号码。尽管医保卡号码与身份证号码不同,但假设可以确定采用医保卡挂号和采用身份证挂号的病人均为病人A,则可以为病人A建立主索引,以使不同用户标识数据关联至病人A。例如,以手机号作为主索引,为病人A建立主索引。

[0043] 在为用户建立主索引之后,可以对各个数据表中的数据进行分析。例如,可以根据所述数据中包含的用户标识数据进行分析,确定关联至同一用户的各个用户标识数据所在的数据表。如果多个数据表中的用户标识数据关联至同一主索引,则可以确定所述多个数据表属于同一用户。

[0044] 进一步,可以将属于同一用户的多个数据表进行融合,得到原始宽表。原始宽表可能包含多个重复数据,如果单纯地整合多张数据表,则可能形成笛卡尔积。优选地,可以对数据进行处理,例如,将所述原始宽表进行过滤,去除重复、无用的数据,以节省原始宽表的存储空间。

[0045] 作为一个非限制性实施例,参考表1、表2和表3,在建立宽表之前,为用户A建立主索引IndexA。如果根据用户A的唯一标识信息(例如,身份证号码)可以确定表1、表2和表3均为用户A的数据表,则可以对属于用户A的3个数据表进行融合,得到用户A的原始宽表(例如,表4)。

[0046] 表1

[0047]	用户标识	姓名	性别	年龄	疾病
	身份证号	用户A	女	32	肺炎

[0048] 表2

[0049]

用户标识	姓名	性别	年龄	疾病
医保卡号	用户A	女	32	咳嗽

[0050] 表3

[0051]

用户标识	姓名	性别	年龄	疾病
临时病历卡号	用户A	女	32	高烧

[0052] 表4

[0053]

用户标识	姓名	性别	年龄	疾病
身份证号	用户A	女	32	肺炎
医保卡号	用户A	女	32	咳嗽
临时病历卡号	用户A	女	32	高烧

[0054] 之后,可以对原始宽表进行过滤,去除重复的数据以得到如表5所示的宽表,该宽表中,单个字段可以对应多个值(例如,主索引IndexA对应3种疾病(肺炎、咳嗽、高烧))。

[0055] 表5

[0056]

用户标识	姓名	性别	年龄	疾病
主索引 IndexA	用户 A	女	32	肺炎
				咳嗽
				高烧

[0057] 具体地,可以删除表1、表2和表3中表示用户A的各个标识数据,并采用主索引IndexA作为用户A的唯一用户标识;之后,可以滤除各个用户标识数据中的重复数据,例如删除多余的姓名、性别和年龄信息。当临床数据的数据量很大(例如,超过数亿条数据)时,可以极大地节约存储成本,并提高数据搜索效率。

[0058] 本领域技术人员理解,具体实施时,表1至表5中的临床数据还可以包含病人住址信息、就诊日期以及疾病的具体诊断信息、化验信息、用药信息等等。其中,病人住址信息也可以合并后记录在所述宽表中。如果疾病信息相同,但是诊断信息、用药信息等不同,则可以表示疾病信息不是重复的数据,不可以合并,将保留相关数据。

[0059] 在步骤S102中,可以响应于用户输入的检索表达式,在所述宽表中搜索,以得到搜索结果。

[0060] 其中,所述检索表达式可以包括检索关键词和各种逻辑运算符,并可以根据检索系统规定的各种运算符将检索词之间的逻辑关系、位置关系,形成计算机可以识别和执行的检索命令式。需要说明的是,所述检索表达式也可以未包含逻辑运算符,仅包含检索词。

[0061] 例如,所述检索表达式可以为“胰头癌and HbsAg”。又例如,所述检索表达式可以为“肺炎not高烧”。再例如,所述检索表达式可以为“血常规or糖尿病”。另例如,“(胰头癌and HbsAg) or (胰十二指肠切除术andca199>1200)”。还例如,“HbsAg and导管腺癌”。

[0062] 作为一个变化实施例,参考图2,所述搜索方法可以包括以下步骤:

[0063] 步骤S201:合并所述多个数据表,并将所述多个数据表中同一用户的数据表融合,

以得到单个宽表；

[0064] 步骤S202:根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果；

[0065] 步骤S203:根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果时,判断所述检索表达式中的检索关键词是否包含在用于分词的分词词典中；

[0066] 步骤S204:如果所述检索关键词未包含于所述分词词典中,则将所述检索关键词的输入次数进行递增；

[0067] 步骤S205:如果所述检索关键词的输入次数达到预设阈值,则将所述检索关键词加入所述分词词典。

[0068] 具体而言,首先步骤S201,即合并所述多个数据表,并将所述多个数据表中同一用户的数据表融合,以得到单个宽表;之后,执行步骤S202,即根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果。所述步骤S201和步骤S202可以参照图1所示的步骤S101和步骤S102,这里不再赘述。

[0069] 在步骤S203中,当用户输入检索表达式之后,可以根据输入的检索表达式,在所述宽表中进行搜索。在搜索时,可以确定所述检索表达式中的检索关键词。一旦确定检索关键词,就可以判断所述检索关键词是否出现在所述分词词典中。所述检索关键词可以是检索表达式中的检索词原词,也可以是对所述检索表达式中的检索词进行分词得到的分词词语。

[0070] 通常情况下,可以根据分词词典进行分词。所述分词词典可以是行业标准词库。所述行业标准词库可以包括国际疾病分类与健康问题(International Statistical Classification of Diseases and Related Health Problems 10<sup>th</sup>,简称ICD-10)、临床医学学术语标准术语集(Systematized Nomenclature of Medicine-Clinical Terms,简称SNOMED-CT)、卫生信息交换标准(Health Level 7,简称HL7)等;本领域技术人员理解,所述分词词典还可以是通过人工的方式收集、整理得到的用户自定义的分词词典或适于医学分词的其他词典,这里不再赘述。

[0071] 在步骤S204中,在对用户输入的检索表达式进行拉取整合分析时,可以确定所述检索关键词是否在现有的分词词典中出现。如果所述分词词典未包含所述检索关键词,则可以在用户输入该检索关键词之后,对该检索关键词的输入次数进行累加。或者,如果用户后续输入的检索词在分词后可以再次得到所述检索关键词,则可以对该检索关键词的输入次数进行累加。

[0072] 进一步,在步骤S205中,如果该检索关键词的输入次数达到预设阈值(例如,预设阈值为10),则可以将该检索关键词作为新词汇加入分词词典中,既可以丰富分词词典的词汇量,又可以加快后续搜索速度,提高搜索效率。

[0073] 作为又一个变化例,参考图3,所述搜索方法可以包括以下步骤:

[0074] 步骤S301:合并所述多个数据表,并将所述多个数据表中同一用户的数据表融合,以得到单个宽表;

[0075] 步骤S302:根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果;

[0076] 步骤S303:对所述搜索结果进行统计,以得到统计结果;

[0077] 步骤S304:在呈现所述搜索结果时,以图表的形式一并呈现所述统计结果。

[0078] 具体而言,首先执行步骤S301,即合并所述多个数据表,并将所述多个数据表中同

一用户的数据表融合,以得到单个宽表;其次执行步骤S302,即根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果。所述步骤S301和步骤S302可以参照图1所示的步骤S101和步骤S102,这里不再赘述。

[0079] 之后,在步骤S303中,可以对所述搜索结果进行统计,以得到统计结果。例如,可以统计搜索结果的数量,也可以统计不同条件下的搜索结果的数量比例等。

[0080] 在步骤S304中,可以在呈现所述搜索结果时,一并呈现所述统计结果。所述统计结果可以采用图表形式呈现。

[0081] 作为一个非限制性的实施例,所述搜索结果包括指定的性别、年龄下的各种检验项目(例如,血常规检验、尿常规检验、粪便常规检验),此时,可以对所述搜索结果进行统计,以分别确定血常规检验、尿常规检验、粪便常规检验的搜索结果数量。之后,可以对统计结果进行表格绘制,向用户直观展示搜索结果。

[0082] 作为又一个非限制性的实施例,用户搜索血常规检验结果的数据,可以逐条显示血常规检验结果的相关搜索结果。此外,还可以对搜索结果进行统计,例如,统计白细胞 $>1200$ 、白细胞 $\leq 1200$ 的患者数量,之后,可以采用柱状图形式呈现所述统计结果,所述柱状图中可以包含患者的数量值。或者也可以统计白细胞 $>1200$ 的每一患者的属性信息(例如,性别或者年龄段等)的搜索结果的各自占比,以饼图形式呈现。

[0083] 本领域技术人员理解,还可以采用其他图表形式呈现所述统计结果,统计结果类别也不受本发明实施例限制。为简便,其他图表形式和/或统计结果类别等不再一一列举。

[0084] 由上,采用本发明实施例提供的搜索方法,可以加快搜索速度,提高搜索效率。

[0085] 图4是本发明实施例的一种临床大数据的搜索装置的结构示意图。参考图4,所述临床大数据的搜索装置3(以下简称为搜索装置3)可以用于实施上述图1至图3所示的搜索方法技术方案。

[0086] 具体而言,所述临床大数据包含多个数据表,所述搜索装置3可以包括:确定模块31,适于合并所述多个数据表,并将所述多个数据表中同一用户的数据表融合,以得到单个宽表;搜索模块32,适于根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果。

[0087] 作为一个非限制性实施例,所述确定模块31包括:分析子模块311,对所述多个数据表中的数据进行分析,以确定属于同一用户的各个数据表;融合子模块,将属于同一用户的各个数据表融合。

[0088] 进一步,所述分析子模块311可以包括:分析单元3111,适于对所述多个数据表中的用户标识数据进行分析,如果不同数据表中的用户标识关联至同一用户,则将所述不同的数据表确定为属于同一用户。

[0089] 进一步,所述融合子模块312可以包括:融合单元3121,适于将属于同一用户的各个数据表融合,以得到原始宽表;清洗单元3122,适于对所述原始宽表进行过滤,去除重复的数据,以得到所述宽表。

[0090] 进一步,所述搜索装置3还可以包括:判断模块33,适于根据输入的检索表达式在所述宽表中进行搜索以得到搜索结果时,判断所述检索表达式中的检索关键词是否包含在用于分词的分词词典中;计算模块34,如果所述检索关键词未包含于所述分词词典中,则所述计算模块34适于将所述检索关键词的输入次数进行递增;加入模块35,如果所述检索关键词的输入次数达到预设阈值,则所述加入模块35适于将所述检索关键词加入所述分词词

典。

[0091] 优选地,所述搜索装置3还可以包括:统计模块36,对所述搜索结果进行统计,以得到统计结果;呈现模块37,适于在呈现所述搜索结果时,以图表的形式一并呈现所述统计结果。

[0092] 关于图4所示的搜索装置3的工作原理、工作方式的更多内容,可以一并参照上述图1至图3中的相关描述,这里不再赘述。

[0093] 进一步地,本发明实施例还公开一种存储介质,其上存储有计算机指令,所述计算机指令运行时执行上述图1至图3所示实施例中所述的临床大数据的搜索方法技术方案。优选地,所述存储介质可以包括计算机可读存储介质。所述存储介质可以包括ROM、RAM、磁盘或光盘等。

[0094] 进一步地,本发明实施例还公开一种服务器,包括存储器和处理器,所述存储器上存储有能够在所述处理器上运行的计算机指令,所述处理器运行所述计算机指令时执行上述图1至图3所示实施例中所述的临床大数据的搜索方法技术方案。

[0095] 虽然本发明披露如上,但本发明并非限于于此。任何本领域技术人员,在不脱离本发明的精神和范围内,均可作各种更动与修改,因此本发明的保护范围应当以权利要求所限定的范围为准。

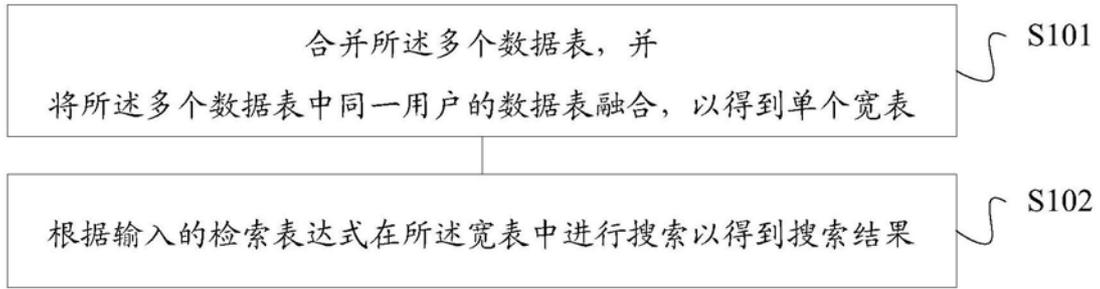


图1

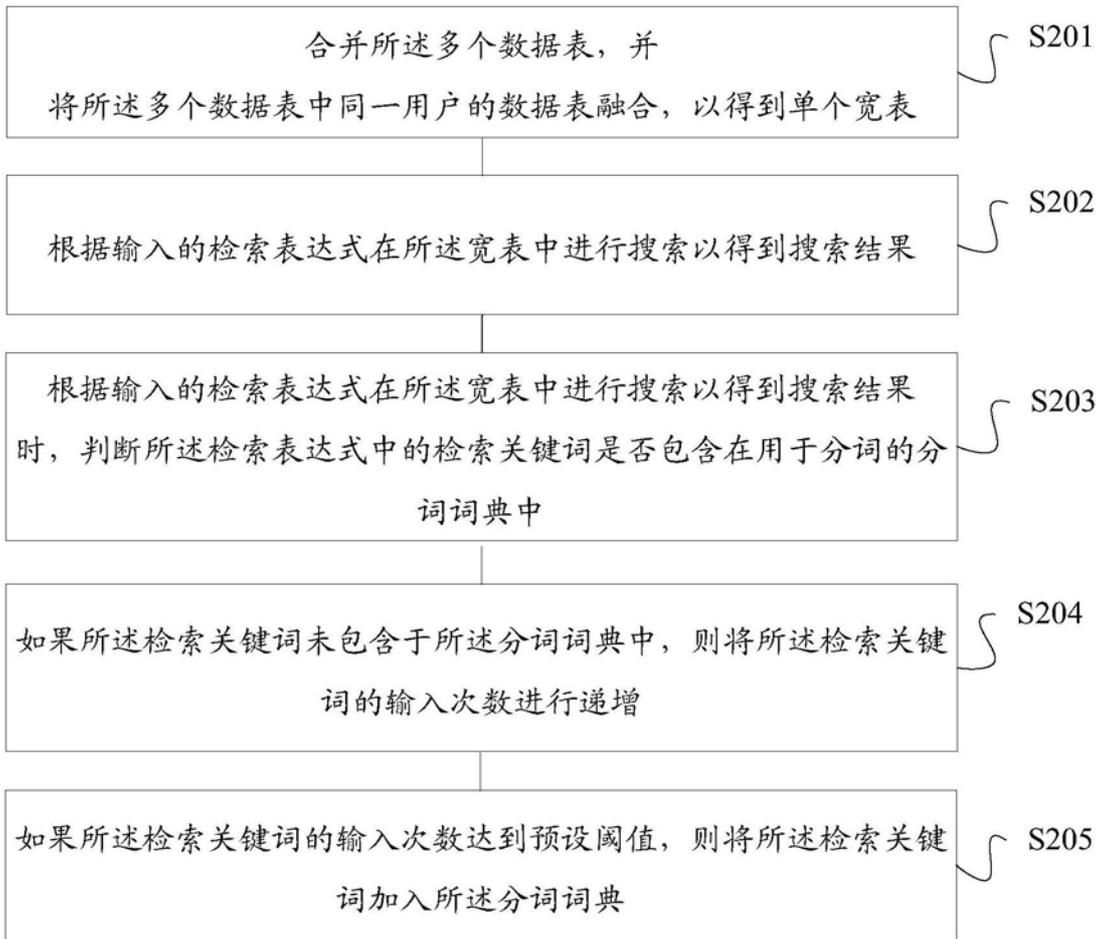


图2

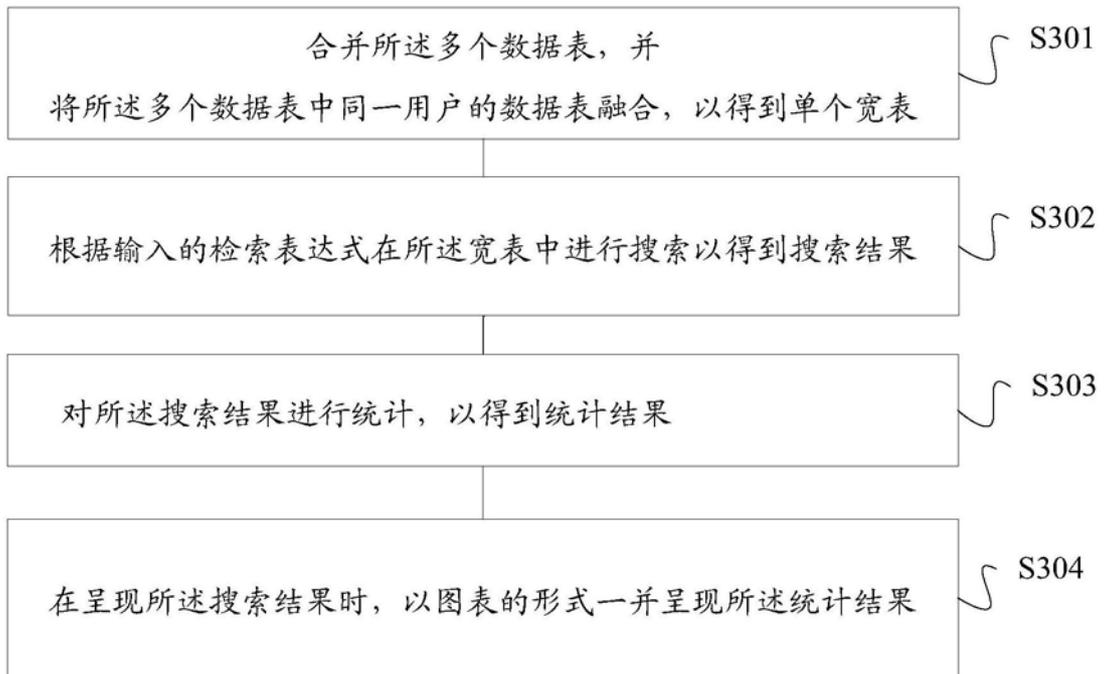


图3

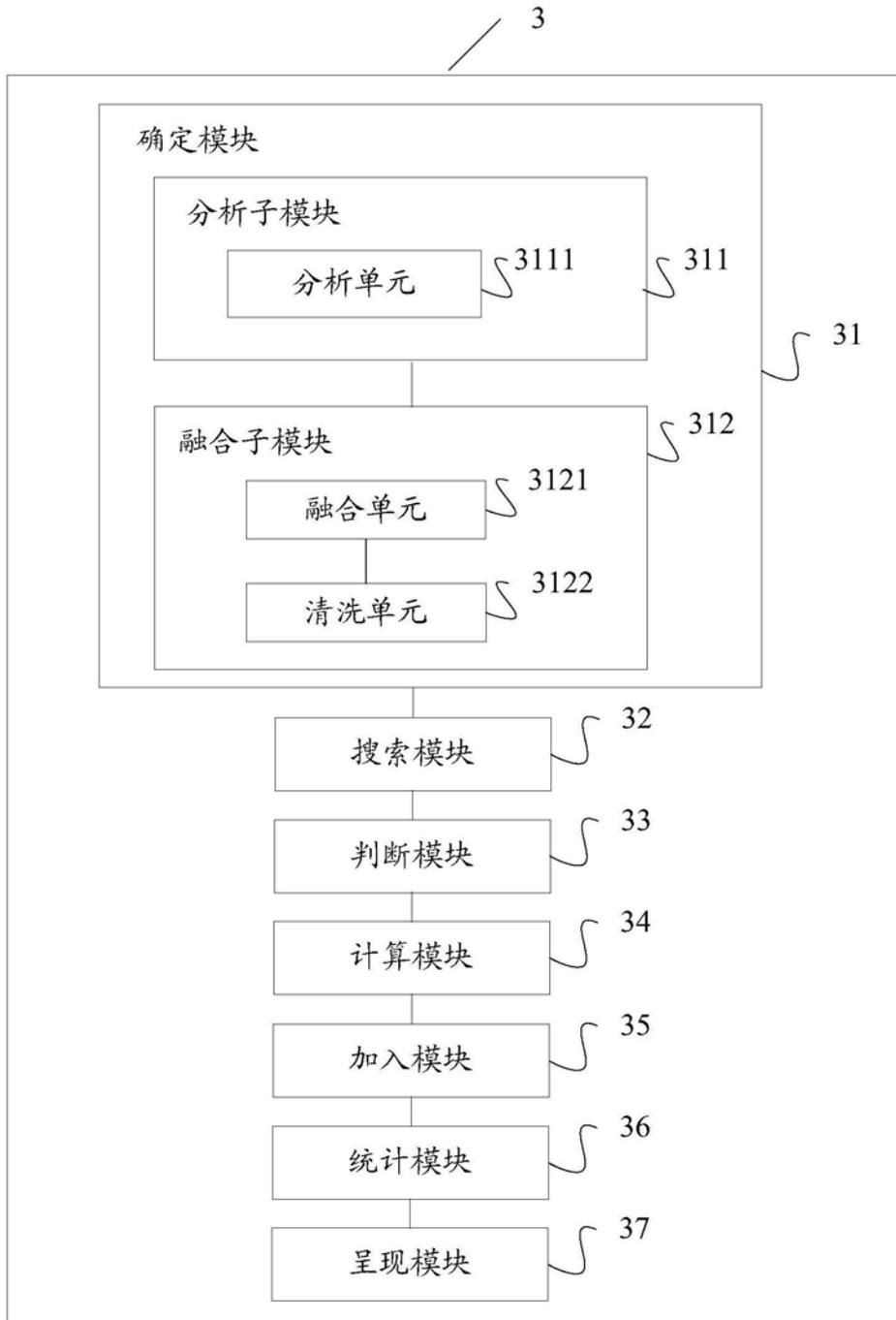


图4