



(12) 发明专利申请

(10) 申请公布号 CN 104636485 A

(43) 申请公布日 2015. 05. 20

(21) 申请号 201510085493. 8

(22) 申请日 2015. 02. 16

(71) 申请人 广东省公安厅

地址 510000 广东省广州市越秀区黄华路
97 号

申请人 珠海市公安局

安徽讯飞智元信息科技有限公司

(72) 发明人 高永强 母宝

(74) 专利代理机构 广州市越秀区哲力专利商标
事务所(普通合伙) 44288

代理人 齐文剑

(51) Int. Cl.

G06F 17/30(2006. 01)

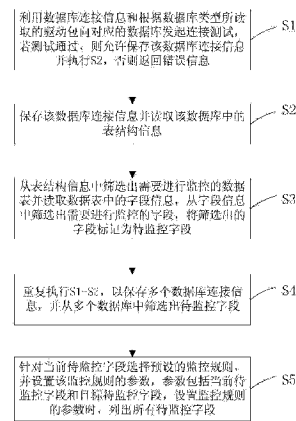
权利要求书2页 说明书5页 附图1页

(54) 发明名称

基于数据监控的多数据源数据关联方法和装置

(57) 摘要

本发明公开了基于数据监控的多数据源数据关联方法和装置,方法包括:利用数据库连接信息和根据数据库类型所读取的驱动包向对应的数据库发起连接测试;保存该数据库连接信息并读取该数据库中的表结构信息;从表结构信息中筛选出需要进行监控的数据表并读取数据表中的字段信息,从字段信息中筛选出需要进行监控的字段,将筛选出的字段标记为待监控字段;重复执行 S1-S3,以保存多个数据库连接信息,并从多个数据库中筛选出待监控字段;针对当前待监控字段选择预设的监控规则,并设置该监控规则的参数。本发明通过读取多个数据库中的数据,将需要被监控的字段标记为待监控字段,在设置监控规则时,只列出待监控字段,有利于用户选择。



1. 基于数据监控的多数据源数据关联方法,其特征在於,包括以下步骤:

S1:利用数据库连接信息和根据数据库类型所读取的驱动包向对应的数据库发起连接测试,若测试通过,则允许保存该数据库连接信息并执行 S2,否则返回错误信息;

S2:保存该数据库连接信息并读取该数据库中的表结构信息;

S3:从表结构信息中筛选出需要进行监控的数据表并读取数据表中的字段信息,从字段信息中筛选出需要进行监控的字段,将筛选出的字段标记为待监控字段;

S4:重复执行 S1-S3,以保存多个数据库连接信息,并从多个数据库中筛选出待监控字段;

S5:针对当前待监控字段选择预设的监控规则,并设置该监控规则的参数,参数包括当前待监控字段和目标待监控字段,设置监控规则的参数时,列出所有待监控字段,以供用户选择。

2. 根据权利要求 1 所述的基于数据监控的多数据源数据关联方法,其特征在於,数据库连接信息包括数据库 URL、数据库用户名和数据库密码。

3. 根据权利要求 1 所述的基于数据监控的多数据源数据关联方法,其特征在於,在 S3 中,被标记为待监控字段后,针对该待监控字段添加字段描述框,以供用户对该待监控字段进行描述,当保存该待监控字段时,判断其对应的字段描述框的内容是否为空,若是,则显示空值错误,否则,该待监控字段保存成功。

4. 根据权利要求 1 所述的基于数据监控的多数据源数据关联方法,其特征在於,在 S5 中,列出所有待监控字段时,同时将待监控字段所对应的字段描述列出。

5. 根据权利要求 1 所述的基于数据监控的多数据源数据关联方法,其特征在於,在 S5 中,在同一监控规则中所列出的待监控字段中,被设为参数的待监控字段会被剔除。

6. 基于数据监控的多数据源数据关联装置,其特征在於,包括以下模块:

数据库连接测试模块:用于利用数据库连接信息和根据数据库类型所读取的驱动包向对应的数据库发起连接测试,若测试通过,则允许保存该数据库连接信息并执行保存数据库模块,否则返回错误信息;

保存数据库模块:用于保存该数据库连接信息并读取该数据库中的表结构信息;

筛选字段模块:用于从表结构信息中筛选出需要进行监控的数据表并读取数据表中的字段信息,从字段信息中筛选出需要进行监控的字段,将筛选出的字段标记为待监控字段;

多数据库筛选字段模块:用于重复执行数据库连接测试模块、保存数据库模块和筛选字段模块,以保存多个数据库连接信息,并从多个数据库中筛选出待监控字段;

监控规则参数设置模块:用于针对当前待监控字段选择预设的监控规则,并设置该监控规则的参数,参数包括当前待监控字段和目标待监控字段,设置监控规则的参数时,列出所有待监控字段,以供用户选择。

7. 根据权利要求 6 所述的基于数据监控的多数据源数据关联装置,其特征在於,数据库连接信息包括数据库 URL、数据库用户名和数据库密码。

8. 根据权利要求 6 所述的基于数据监控的多数据源数据关联装置,其特征在於,在筛选字段模块中,被标记为待监控字段后,针对该待监控字段添加字段描述框,以供用户对该待监控字段进行描述,当保存该待监控字段时,判断其对应的字段描述框的内容是否为空,

若是,则显示空值错误,否则,该待监控字段保存成功。

9. 根据权利要求 6 所述的基于数据监控的多数据源数据关联装置,其特征在于,在监控规则参数设置模块中,列出所有待监控字段时,同时将待监控字段所对应的字段描述列出。

10. 根据权利要求 6 所述的基于数据监控的多数据源数据关联装置,其特征在于,在监控规则参数设置模块中,在同一监控规则中所列出的待监控字段中,被设为参数的待监控字段会被剔除。

基于数据监控的多数据源数据关联方法和装置

技术领域

[0001] 本发明涉及基于数据监控的多数据源数据关联方法。

背景技术

[0002] 随着公安信息化工程的不断建设,以及实际应用的成效表明,数据质量至关重要。

[0003] 众所周知,信息资源是开展信息化侦查最为重要的支撑条件。但是目前公安网系统里所存储的信息资源质量并不高,公安网上的犯罪信息良莠不齐,质量高的和质量低的并存。

[0004] 简而言之就是信息污染较严重,即信息资源混入了干扰性、欺骗性和误导性的信息。虽然近几年公安信息化建设取得了很大进步,各地区各警种均建立各种各样的信息系统,公安工作的基础数据也在日益完善。如:常(暂)住人口信息,机动车信息,各种证照信息,旅馆业信息,以及刑侦部门收集的大量案件、人员、物品信息等。所有这些,都为信息资源的利用奠定了坚实的基础。但是,在工作中也发现大量虚假、夸大、重复、和过时的无用信息,这些垃圾信息与有效信息混杂在一起,降低了网络信息资源的可靠性和对有效信息的判断,严重影响了民警对有效信息的吸收和利用。民警为了获得有用信息,不得不花大量的时间去阅读那些无关的信息,容易导致无功而返。

[0005] 主要表现如下:

[0006] 1、数据录入随意性大,应采必采、当录必录的要求不严,致使缺项多、差错多。

[0007] 以案件信息为例,案件信息的不报、漏报、错报问题就十分突出。如有一个抓如立案的典型单位,一年间补立案前案件和隐漏案件就多达 1000 多起。

[0008] 2、不严格遵循数据标准规范,或标准修订滞后,给跨地区、跨系统数据整合、清理带来很多困难,大量信息冗余。

[0009] 目前,随着各单位各部门之间都开发了一些专业软件系统。但由于不严格遵循数据标准规范,数据共享性差,数据格式不同,重复和多余现象经常出现。在跨系统数据整合时会产生大量的信息冗余,这些多余重复的无用信息,非但不能给人们以新的认识,反而还成为人们认识领域的障碍,给信息的甄别带来困难,让民警在分析研判时无从着手。

[0010] 3、缺乏责任机制确保数据质量的约束机制和激励机制尚不健全。

[0011] 特别是前几年,“各级加快推进信息化建设,经常组织开展集中突击式的信息资源采集,在短时间内完成海量数据的采集录入,但由于部分民警的信息化应用技能较低、培训学习不够,又缺乏科学合理的约束机制、考核标准来保证信息采集的质量和录入的规范化,致使部分单位和个人片面追求考核指标中的数据数量指标,忽视数据质量,甚至采取一些投机取巧的办法加大数据量,在系统中录入了大量的垃圾信息;少数单位仍然存在辅警或聘用人员录入数据的情况,也可能造成录入数据与原始数据有所偏差”。

[0012] 基于上述的实际情况,对不同数据库中的数据表中存在关联关系的字段进行关联监控则显得尤为重要。

发明内容

[0013] 为了克服现有技术的不足,本发明的目的在于提供基于数据监控的多数据源数据关联方法和装置,读取多个数据库中的数据,将需要被监控的字段标记为待监控字段,在设置监控规则时,只列出待监控字段,去除其它不相关字段,有利于用户选择。

[0014] 为解决上述问题,本发明所采用的技术方案如下:

[0015] 方案一:

[0016] 基于数据监控的多数据源数据关联方法,包括以下步骤:

[0017] S1:利用数据库连接信息和根据数据库类型所读取的驱动包向对应的数据库发起连接测试,若测试通过,则允许保存该数据库连接信息并执行 S2,否则返回错误信息;

[0018] S2:保存该数据库连接信息并读取该数据库中的表结构信息;

[0019] S3:从表结构信息中筛选出需要进行监控的数据表并读取数据表中的字段信息,从字段信息中筛选出需要进行监控的字段,将筛选出的字段标记为待监控字段;

[0020] S4:重复执行 S1-S3,以保存多个数据库连接信息,并从多个数据库中筛选出待监控字段;

[0021] S5:针对当前待监控字段选择预设的监控规则,并设置该监控规则的参数,参数包括当前待监控字段和目标待监控字段,设置监控规则的参数时,列出所有待监控字段,以供用户选择。

[0022] 进一步地,数据库连接信息包括数据库 URL、数据库用户名和数据库密码。

[0023] 进一步地,在 S3 中,被标记为待监控字段后,针对该待监控字段添加字段描述框,以供用户对该待监控字段进行描述,当保存该待监控字段时,判断其对应的字段描述框的内容是否为空,若是,则显示空值错误,否则,该待监控字段保存成功。

[0024] 进一步地,在 S5 中,列出所有待监控字段时,同时将待监控字段所对应的字段描述列出。

[0025] 进一步地,在 S5 中,在同一监控规则中所列出的待监控字段中,被设为参数的待监控字段会被剔除。

[0026] 方案二:

[0027] 基于数据监控的多数据源数据关联装置,包括以下模块:

[0028] 数据库连接测试模块:用于利用数据库连接信息和根据数据库类型所读取的驱动包向对应的数据库发起连接测试,若测试通过,则允许保存该数据库连接信息并执行保存数据库模块,否则返回错误信息;

[0029] 保存数据库模块:用于保存该数据库连接信息并读取该数据库中的表结构信息;

[0030] 筛选字段模块:用于从表结构信息中筛选出需要进行监控的数据表并读取数据表中的字段信息,从字段信息中筛选出需要进行监控的字段,将筛选出的字段标记为待监控字段;

[0031] 多数据库筛选字段模块:用于重复执行数据库连接测试模块、保存数据库模块和筛选字段模块,以保存多个数据库连接信息,并从多个数据库中筛选出待监控字段;

[0032] 监控规则参数设置模块:用于针对当前待监控字段选择预设的监控规则,并设置该监控规则的参数,参数包括当前待监控字段和目标待监控字段,设置监控规则的参数时,列出所有待监控字段,以供用户选择。

[0033] 进一步地,数据库连接信息包括数据库 URL、数据库用户名和数据库密码。

[0034] 进一步地,在筛选字段模块中,被标记为待监控字段后,针对该待监控字段添加字段描述框,以供用户对该待监控字段进行描述,当保存该待监控字段时,判断其对应的字段描述框的内容是否为空,若是,则显示空值错误,否则,该待监控字段保存成功。

[0035] 进一步地,在监控规则参数设置模块中,列出所有待监控字段时,同时将待监控字段所对应的字段描述列出。

[0036] 进一步地,在监控规则参数设置模块中,在同一监控规则中所列出的待监控字段中,被设为参数的待监控字段会被剔除。

[0037] 相比现有技术,本发明的有益效果在于:读取多个数据库中的数据,将需要被监控的字段标记为待监控字段,在设置监控规则的参数时,只列出待监控字段,去除其它不相关字段,有利于用户进行选择。另外,在选择待监控字段时,同时将待监控字段所对应的字段描述列出,有利于用户理解该字段的含义。

附图说明

[0038] 图 1 为本发明的基于数据监控的多数据源数据关联方法的流程图。

具体实施方式

[0039] 下面,结合附图以及具体实施方式,对本发明做进一步描述:

[0040] 参考图 1 本发明的多数据源数据关联方法,主要针对不同地区的不同公安部门所建立的数据系统中的数据进行关联监控,目前的公安部门所建立的数据系统中的数据均由不同的数据负责人进行录入和管理,会存在数据质量高低不平的情况,其中可能存在一些虚假、重复或者过期数据,这些数据可以理解为垃圾数据,与有效数据混在一起将大大降低数据系统的数据可靠性和有效性,严重影响了有效数据的利用率,或者有些数据并没有录入系统,因此公安人员在办案时,会出现读取不到相关数据的情况,或者需要花大量时间在混杂了垃圾数据的数据库中读取有效数据的过程。因此需要建立一个数据监控系统,利用其中的数据监控方法对多个数据库之间的数据表及其数据进行关联监控,通过具体的监控规则达到监控数据是否异常的目的,并且能在监控报告中直观地查看是否有异常告警。其中,基于数据监控的多数据源数据关联方法的步骤如下:

[0041] S1:利用数据库连接信息和根据数据库类型所读取的驱动包向对应的数据库发起连接测试,若测试通过,则允许保存该数据库连接信息并执行 S2,否则返回错误信息。

[0042] 数据库连接信息包括数据库 URL、数据库用户名和数据库密码。数据库类型包括 DB2 数据库、ORACLE 数据库和 MYSQL 数据库,不同类型的数据库具有其对应的驱动包,驱动包为预设的,在检测到数据库类型时,会自动获取相应的驱动包配合数据连接信息进行连接测试。

[0043] S2:保存该数据库连接信息并读取该数据库中的表结构信息。连接测试成功后,用户可以对该数据库连接信息进行保存,并保持对该数据库的连通状态,配合前面读取的驱动包可以解析该数据库中的数据。

[0044] S3:从表结构信息中筛选出需要进行监控的数据表并读取数据表中的字段信息,从字段信息中筛选出需要进行监控的字段,将筛选出的字段标记为待监控字段。

[0045] 由于每个数据库中都有大量的数据表,每个数据表中也可能包括大量的字段,因此在该步骤中需要对需要监控的字段进行筛选,并标记为待监控字段,在后面的步骤中,不是待监控字段的字段将不会被显示,为用户设置监控规则时带来很大的便利。

[0046] 进一步地,被标记为待监控字段后,针对该待监控字段添加字段描述框,以供用户对该待监控字段进行描述,当保存该待监控字段时,判断其对应的字段描述框的内容是否为空,若是,则显示空值错误,否则,该待监控字段保存成功。由于字段来自不同的数据库,而字段名称是在数据库设计时确定的,在不同的数据库中,字段含义相同的字段,其字段名称可能是不同的,因此,为字段添加字段描述是很重要的。而且将字段描述强制设为必选项,只要在用户对该字段进行含义描述后才能将其保存为待监控字段。

[0047] S4:重复执行 S1-S3,以保存多个数据库连接信息,并从多个数据库中筛选出待监控字段。经过该步骤就将多个数据库中的需要监控的字段筛选出来。

[0048] S5:针对当前待监控字段选择预设的监控规则,并设置该监控规则的参数,参数包括当前待监控字段和目标待监控字段,设置监控规则的参数时,列出所有待监控字段,以供用户选择。

[0049] 在该步骤中,主要对监控规则进行参数设置,当前待监控字段指当前被选择的待监控字段且默认为监控规则中的其中一个参数,除了当前待监控字段,还要选择目标待监控字段,目标待监控字段指其他数据表中的待监控字段,其他数据表可以是同一数据库中的数据表,也可以是其他数据库中的数据表。在选择待监控字段时,同时将待监控字段所对应的字段描述列出,以便于用户理解该字段的含义。而且,在同一监控规则中所列出的待监控字段中,被设为参数的待监控字段会被剔除,即一个数据表中的一个字段在一条监控规则中只能用一次。

[0050] 该步骤中的监控规则具体指异表联查规则,异表联查规则为对当前数据表中被选定的字段中的数据与另一数据表中被选定的字段中的数据之间的关系进行监控,从而得到不同数据表字段之间完整性约束或者一致性约束,以发现目标数据表中的错误数据,例如,异表联查规则可以是异表身份证重号检查,即针对当前选择的身份证字段中的数据与另一指定的表中的身份证字段中的数据相同而对应的姓名不同的情况进行检查。监控规则分为预设监控规则和自定义监控规则,预设监控规则保存于监控规则库中,使用时只需要调用即可,若预设监控规则不能满足监控任务的需求,用户可选择自定义监控规则,自定义监控规则支持 sql 语句和正则表达式两种规则设置方式,用户根据实际需求设定了监控规则的逻辑后,可通过测试判断该规则是否达到预期的目的。

[0051] 进一步地,在完成监控规则的参数设置后,会对该监控规则中的参数进行非空检查,若某个参数为空则会提示对应的参数未设置,若全部参数均为非空则完成监控规则的参数设置。

[0052] 本发明的基于数据监控的多数据源数据关联方法的优点在于:读取多个数据库中的数据,将需要被监控的字段标记为待监控字段,在设置监控规则的参数时,只列出待监控字段,去除其它不相关字段,有利于用户进行选择。另外,在选择待监控字段时,同时将待监控字段所对应的字段描述列出,有利于用户理解该字段的含义。

[0053] 对应于上述的基于数据监控的多数据源数据关联方法,还公开了基于数据监控的多数据源数据关联装置,包括以下模块:

[0054] 数据库连接测试模块：用于利用数据库连接信息和根据数据库类型所读取的驱动包向对应的数据库发起连接测试，若测试通过，则允许保存该数据库连接信息并执行保存数据库模块，否则返回错误信息；

[0055] 保存数据库模块：用于保存该数据库连接信息并读取该数据库中的表结构信息；

[0056] 筛选字段模块：用于从表结构信息中筛选出需要进行监控的数据表并读取数据表中的字段信息，从字段信息中筛选出需要进行监控的字段，将筛选出的字段标记为待监控字段；

[0057] 多数据库筛选字段模块：用于重复执行数据库连接测试模块、保存数据库模块和筛选字段模块，以保存多个数据库连接信息，并从多个数据库中筛选出待监控字段；

[0058] 监控规则参数设置模块：用于针对当前待监控字段选择预设的监控规则，并设置该监控规则的参数，参数包括当前待监控字段和目标待监控字段，设置监控规则的参数时，列出所有待监控字段，以供用户选择。

[0059] 进一步地，数据库连接信息包括数据库 URL、数据库用户名和数据库密码。

[0060] 进一步地，在筛选字段模块中，被标记为待监控字段后，针对该待监控字段添加字段描述框，以供用户对该待监控字段进行描述，当保存该待监控字段时，判断其对应的字段描述框的内容是否为空，若是，则显示空值错误，否则，该待监控字段保存成功。

[0061] 进一步地，在监控规则参数设置模块中，列出所有待监控字段时，同时将待监控字段所对应的字段描述列出。

[0062] 进一步地，在监控规则参数设置模块中，在同一监控规则中所列出的待监控字段中，被设为参数的待监控字段会被剔除。

[0063] 对本领域的技术人员来说，可根据以上描述的技术方案以及构思，做出其它各种相应的改变以及形变，而所有的这些改变以及形变都应该属于本发明权利要求的保护范围之内。

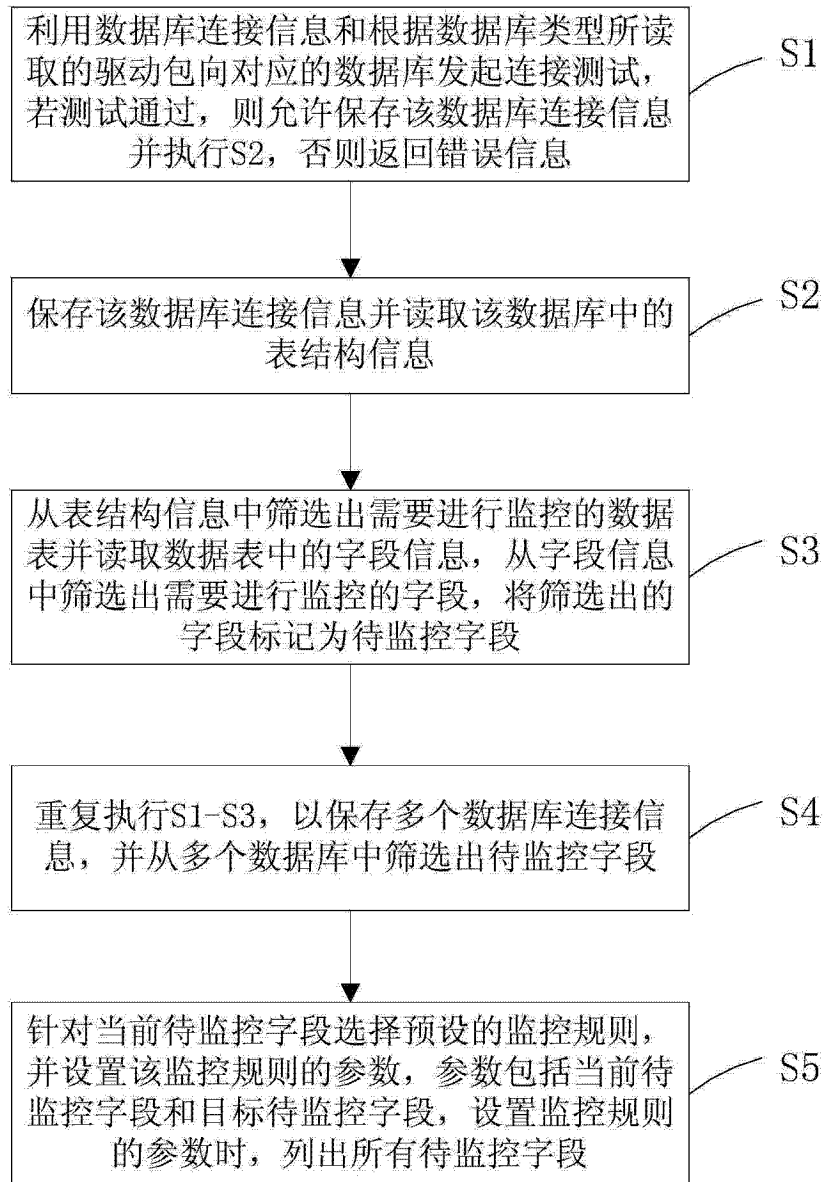


图 1