

基于智能匹配的计算机审计数据采集技术研究

叶焕倬, 何玉凝

(中南财经政法大学 信息与安全工程学院, 湖北 武汉 430060)

[摘要]当前计算机审计数据采集中所面临的主要问题是审计人员需要大量地手动匹配源表与目标表字段。基于字段的智能匹配技术同时利用了属性值和属性名的信息进行属性匹配,可以实现数据采集过程中源表与目标表字段的自动匹配,减少审计人员手动参与。实验表明,该技术适用于电子审计数据,具有较好的查准率。

[关键词]计算机审计;数据采集;智能匹配

[中图分类号]F239.1 **[文献标识码]**A **[文章编号]**1004-4833(2010)05-0040-07

一、引言

随着信息化进程的迅猛发展,信息技术推动审计朝着持续、实时的方向发展,传统审计模式将逐步被计算机审计模式所代替^[1]。在我国,“金审”工程的启动为审计信息化建设和应用技术研究提供了有利的条件。审计署组织计算机审计已有近10年的时间,近5年来,开发和推广应用了财政审计、金融审计、企业审计等方面的计算机审计软件,也结合实际需要研制了应用于数据采集与分析等方面的软件,在审计业务中发挥了较好的作用。

当前我国正在推广“金审一期”的重要成果——AO(Auditor Office)现场审计实施系统。AO系统主要为审计人员提供了现场审计实施的功能,提高了审计人员的工作效率。在审计电子数据采集方面,AO已具有三种方法:

1. GB/T19581-2004 数据采集

该方法采用了《信息技术会计核算软件数据接口》国家标准,使用基于文本格式和XML格式的数据接口规范,采集整理的数据需符合此国家标准才可导入AO中。此方法一般较少采用。

2. 财务软件备份数据采集

财务备份数据采集是指被审单位提供的数据是由所使用的财务软件备份导出。AO中已具有众多采集模板,对于符合模板的数据就易于采集。AO提供对多种财务软件备份及数据库的接口支持,选择正确的转换接口后可以由AO自动完成数据转换。

虽然在AO中具有众多采集模板,但目前适用于企业的财务软件越来越多,需要定时更新模板。2005年刚实施AO时并无如此众多模板,2008年审计署发布了一个模板升级程序,才具有现在较丰富的模板。但是,由于需要为每台采集转换的电脑进行升级,造成了工作量大、效率低的情况。所以,虽然这种方法当前可能够用,但从长期发展来看是不够灵活和稳定的,并且容易出错。

另外,该方法没有采用开放模板。开放模板在采集时,仍能观察到表、字段的对应情况。如果存在不对应的情况,可以手动修改。AO的采集方法为一次性完成,人为不能干预。

3. 财务软件数据库数据采集

数据库数据采集是指从被审单位后台数据库直接获取数据的情况,在财务软件备份数据采集不能

[收稿日期]2010-03-30

[基金项目]国家自然科学基金项目(70972138)

[作者简介]叶焕倬(1968—),男,湖北武汉人,中南财经政法大学信息与安全工程学院信息系副教授,博士,从事持续审计和人工智能研究;何玉凝(1984—),女,福建厦门人,中南财经政法大学信息与安全工程学院硕士研究生,从事持续审计和异构数据库研究。

使用时多采用此方法。此方法可制作新的财务软件备份数据采集模板,以供以后采集相同模板时使用,取得的数据为最原始数据。

该方法具有较高的稳定性,是三种方法中灵活性最高、适用性最广的。但是它的缺陷也很明显:需要手工完成数据库表与字段的匹配,过程复杂,还要求审计人员在采集前对数据库表与字段的经济含义进行明确。由此可见,如果能够实现采集过程中源数据库与目标数据库表与字段的自动匹配,那么将大大提高采集效率。

源数据库与目标数据库表与字段的匹配将使审计人员在数据采集过程中无需人为指定所需的字段,通过对相似字段的自动匹配,可以大大减少数据采集的工作量,节约大量人力成本。由此可以实现非现场数据采集,为实现在线式审计建立基础,为实现持续审计创造条件。数据转换是数据分析、处理的前提,明确地标识出每个字段的经济含义及其相互之间的关系是数据转换的内容之一,所以实现数据采集过程中的源数据库与目标数据库表与字段的匹配能为后续的数据分析、处理工作建立良好基础。

目前很多国家已经开展了计算机审计数据获取与处理工作,发达的西方国家已经拥有自己的一套标准的数据接口。近几年来,我国的审计手段也已经由传统的手工审计逐渐向计算机审计过渡。同时,国内外对匹配技术的研究已经取得了不少的成果^[2-4]。但是,在现有的查找方式下国外有关在计算机审计中运用智能技术的相关文献是比较少的,在国内的应用更是少见,所以研究基于智能匹配的计算机审计数据采集技术具有非常重要的理论与实际意义^[5-7]。

二、计算机审计数据采集内容与方法

(一) 数据采集基本流程

如果 AO 中有现成模板财务软件数据,使用“财务软件备份数据采集”将很方便,反之,则会非常困难和复杂,特别对于基层的业务部门,例如财政处,更是如此。所以,财务软件备份数据采集不是一个长久有效的方法。本文主要探讨的是第三种方法——财务软件数据库数据采集,只有改进此方法才能从根本上优化采集流程,提高采集效率。

财务软件数据库数据采集的基本流程为:

1. 添加数据源,手动连接源数据库表与目标数据库表

根据系统向导选择添加的数据源,然后手工连

接源数据库表与目标数据库表,匹配相应的字段。

2. 生成财务数据临时表

生成财务数据临时表是指将采集数据库数据后的结果数据,通过一系列转换生成审计分析使用的财务数据表。

(二) 数据采集过程面临的主要问题

财务软件数据库数据采集与前两种方法相比,优势比较明显。该方法灵活性高,适用性广,任何形式的数据源都可以通过此方法进行采集,导入至 AO 中。但是,该方法的不足之处同样明显,主要问题有以下几点:

1. 添加数据源过程中需要手动匹配表与字段

在添加数据源过程中,目标表名将被自动命名为“源_源表名”,如果已存在现有模板,可以从下拉列表中选择所需的模板。如果目标表名为“源_源表名”,则目标列名自动等于“源列”名,需要手工修改目标列名称。如果目标表名为模板名称,则软件会进行列名匹配,但是只能完成简单识别。例如,可识别拼音首字母缩写,KMBM 可识别为科目编码,但对于英文简写、中文则不能识别。

2. 生成财务数据临时表中需要大量的手动匹配表与字段

当 AO 系统需要采集凭证表、科目表、余额表三张表时,应先将这三张表整理为所需的财务表,再存入 AO 的数据库中。所以,需要从这三张表中手动指定科目余额表中的科目编码、本币期初余额、余额方向、会计月份,会计科目表中的科目编码、科目名称、余额方向、科目全称等;对于不同表不同字段需要一一指定,工作量大,耗时长。图 1 显示了生成财务数据临时表中所需的字段对应设置过程。



图 1 字段对应设置

导致这两个问题的主要原因是需要手动匹配源数据库与目标数据库的表与字段。在手动匹配前审

计人员还需了解财务软件的数据库表与字段的经济含义,只有明确数据库表与字段的经济含义,才能准确地进行匹配。

(三) 数据采集问题的当前解决方案与其不足之处

针对目前数据采集中面临的各种问题,各地审计人员也在积极探索,寻找各种解决办法。虽然当前已提出几种解决方案,采取了一些有效方法,但根本问题没有解决。主要方案有以下两种:

1. 在数据导入 AO 系统时使用开放模板

目前 AO 在财务软件备份数据采集时采用的模板不是开放模板。AO 财务软件备份数据采集的整个过程都由系统控制完成,采集过程中人为不能干预,只要存在小部分不匹配模板的数据,整个采集过程都将不能完成。而开放模板可在采集时,仍能观察到表、字段的对应情况。在存在不对应的情况下,可手动修改表和字段的匹配情况,避免整个采集过程中断。当前用友中审 GRP 的数据采集就是采用开放模板。如果采用开放模板,对于部分无法匹配的表与字段则可使用工作量相对较少的匹配工作完成导入,而无需重新制作模板。

假如已存在支持被审单位财务软件备份的模板,当被审单位财务软件升级后,使用开放模板对升级后小部分修改的财务软件备份依然可以采集,只需手动指定部分不匹配的表与字段即可。但是如果采集完全不支持模板的财务软件数据,开放模板也将失去作用,在采集过程中所有的表与字段还是需要重新指定与匹配,与财务软件数据库采集的方法是相同的。

2. 通过财务软件数据库数据采集制作 AO 采集模板

当采集的数据不符合财务软件备份数据采集模板时,只能使用财务软件数据库数据采集。如果每次都使用财务软件数据库数据采集,工作量会很大。对于这种情况,可以使用财务软件数据库数据采集方法制作新的财务软件模板,在以后采集相同格式数据时,就可使用新制作的模板,提高了采集效率。

但是,财务软件数据库数据采集还是有较大的工作量。审计人员采集前需要了解该财务软件的数据库表与字段的经济含义,一般一个数据库含有数十至上百张的表,而往往审计人员只需要其中的几张表,这样会导致工作量很大。

由此可见,这两种方法都需要匹配源数据库与目标数据库的表与字段,源数据库与目标数据库的表与字段的匹配问题才是数据采集的根本问题。如

果能解决这个问题,审计人员就不需要在采集前了解财务软件的数据库表与字段的经济含义,这样才能从根本上减少他们的工作量,提高数据采集效率。

三、基于智能匹配的数据采集技术

本文提出了一种基于字段的智能匹配技术,力求解决数据集中的主要问题。采用数据库相似字段匹配技术实现源数据库与目标数据库的表与字段的匹配,当前主要有三种方法可以进行相似字段的匹配:在数据字典级别上比较属性名的方法;在模式级别上比较模式信息的方法;在数据内容级别上比较属性值的方法^[8]。本文同时利用属性名和属性值的信息进行字段匹配。

(一) 采用基于属性名和属性值的字段匹配的原因

同时采用属性名和属性值的信息进行字段匹配,不仅考虑到其可以有效提高效率,还考虑到电子审计数据所具有的特点有利于同时采用这两种方法。其主要原因可以归纳为以下几点:

一是电子审计数据的属性命名具有一定规则可循。虽然各财务软件的表与字段均不相同,但字段的命名都是具有一定规则的,例如字段名“KMDM”,是科目代码的拼音首字缩写;字段名“FBEGIN-BALANCE”,是借方余额的英文表达方式;在部分财务软件数据库中,仍然存在用中文名称对字段进行命名的情况,如“灵狐 2000”财务软件用“科目名称”中文表达方式作为字段名称。

二是电子审计数据具有易统计的特点。电子审计数据都是从各被审单位数据库里获得的财务会计数据,这些数据多与金额、数量相关,具有易计算、易统计的特点,例如凭证表中的凭证编号、科目编号、借方金额、贷方金额、附件张数等字段,均可进行统计与计算。

使用比较属性值的方法,对所有属性值进行比较统计,获取这些属性值的各种统计值。在比较过程中只使用这些统计值进行属性匹配,可以避免出现因属性值的个数增多、效率变低的情况。

另外,电子审计数据一般多为海量数据,统计值特征明显且较为稳定,不易受到噪声数据的干扰,且不同属性的统计值容易根据其不同特征区分开来。

三是同时使用比较属性名和属性值的方法,可以弥补由于电子审计数据、金额数据内容容易雷同所带来的问题。同时采用属性名和属性值的信息进行属性

匹配,不仅可以有效提高匹配效率,还可以弥补由于电子审计数据、金额数据内容容易雷同所带来的问题。在比较属性值上,可能存在一个表具有多个金额字段的情况,金额字段多为两位小数数据,经过统计后也不易区别开,所以在比较属性值的基础上加入对属性名的比较,将有助于将金额字段区分开来。

(二) 基于属性名的匹配

基于属性命名规则的匹配,需要对属性的命名进行详细了解与分析。经过对电子审计数据的结构与内容的详细调查,我们了解到虽然各财务软件的表与字段均不相同,但字段的命名都是具有一定规则的。现有的比较属性名的方法较多是通过参考同义词字典的方式,具有简单、高效的特点。但是需要提前定义词根表,会额外占用系统空间,并且都需要在匹配前预先进行定义,动态适应性较差。

在比较属性名的过程中,我们为了克服使用同义词字典方式所带来的问题,同时定义了属性命名规则。在采用同义词字典的方式之前,先制定属性的命名规则,然后根据目标表的字段名称生成属性命名规则,并与源表字段名进行比较^[9-10]。整个基于属性名的匹配过程包括:制定属性命名规则、应用属性命名规则、应用启发式规则进行比较、相似度计算。图2给出了处理基于属性名的匹配问题的流程。下文将进行详细说明。

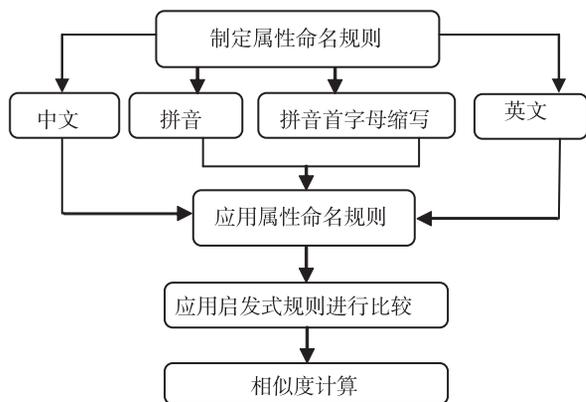


图2 基于属性名的匹配过程

1. 制定属性命名规则

经过对属性名称的分析,我们发现属性的名称具有可获取的规则。当前财务软件数据库的字段命名规则主要为中文、拼音、拼音首字母缩写、英文这四种方式,其中以拼音首字母缩写和英文居多。在匹配过程中,我们不建立词根表进行查询,而是直接使用C#API接口调用字典,所以对目标表的名称结构有一定的要求。为了使目标表符合源表的命名规

则,目标表的字段命名需符合预定义的规则。目标表字段名需由英文组成,在电子审计数据中,英文字段都是由中文名称翻译而成的,由于翻译的缘故,英文字段一般由多个单词组成。所以对目标表字段名称的定义,也是由多个英文单词组成的。为了区别单词,同时为了满足后面的匹配工作的需要,目标表字段名需符合如下形式:

单词 1_单词 2_单词 3

当字段名由多个单词组成时,单词与单词之间需要以“_”符号间隔;当字段名只由一个单词组成时,则不需要再写入“_”符号。

2. 应用属性命名规则

在确定了属性命名规则后,我们将根据目标表的字段名称生成中文、拼音、拼音首字母缩写、英文四种属性命名规则,然后将不同的命名规则与源表字段名称进行比较。属性命名规则的生成都是通过C#的API接口实现的,下面将进行具体说明:英文命名规则应用策略将使用WordNet来对同义词进行识别,利用WordNet提供的API接口,直接通过编程调用这个接口访问字典。中文命名规则应用策略需要先把目标表的英文名称转换为中文,使其符合中文命名规则。有两种方法可以实现英文名称向中文的转换,一是调用Google翻译API,二是在目标表的字段定义中将中文名称写入描述信息中。拼音与拼音首字母缩写命名规则应用策略将使用自行开发的C#中文拼音转换API接口。

3. 应用启发式规则进行比较

当前目标表字段名有四种命名规则,这四种命名规则都将与源表字段名进行比较。本步骤将利用辅助的启发式规则确定源表字段名与命名规则的匹配程度。

具体的启发式规则定义如下:

规则 1:如果源表字段名包含中文命名规则,则匹配。

规则 2:如果源表字段名包含拼音命名规则,则匹配。

规则 3:如果源表字段名包含拼音首字母缩写命名规则,则匹配。

规则 4:如果目标表字段名由单个单词组成,且源表字段名包含英文命名规则,则匹配;如果目标表字段名由 $n(n > 1)$ 个单词组成,当源表字段名包含 $\geq n$ 个英文命名规则,则匹配;当源表字段名包含 $m(0 < m < n)$ 个英文命名规则,则匹配程度需乘以 (m/n) 。

规则 5:如果源表字段符合规则 1、2、3、4 中的任

意一条,就无需再检查是否符合其他规则。

4. 相似度计算

源表字段与目标表字段比较的相似度计算方法为:当符合规则 1、2、3 中任意一条规则时,相似度为 1;当符合规则 4 时,相似度需根据规则 4 的具体情况进行计算;当对所有规则均不符合时,相似度为 0。

这就获得了目标表字段与源表字段的基于属性命名规则的匹配度,其取值范围为 0 到 1,当其值越接近 1 时,表明两个字段的匹配程度越高。

(三) 基于属性值的匹配

比较属性值的方法是用基于数据内容的方法来鉴别异构数据库间相等的字段。将源表的属性与目标表的属性进行一一比较,但随着需要比较的属性的个数增多,直接比较属性的值是否相等及其所来自的域是否相同来确定属性是否为同一属性的方法是低效的。为了克服这个问题,我们可以在比较前通过 sql 语句分别对源表与目标表的所有属性值进行统计,获取属性值的长度、小数位数、平均值、标准差、最大值、最小值等统计信息,在比较过程中使用这些统计信息进行属性匹配。

整个基于属性值的匹配过程包括:根据数据类型对属性分类、属性类型强制转换、建立描述属性的数据指标体系、相似度计算、权重设置、获取最终匹配度。

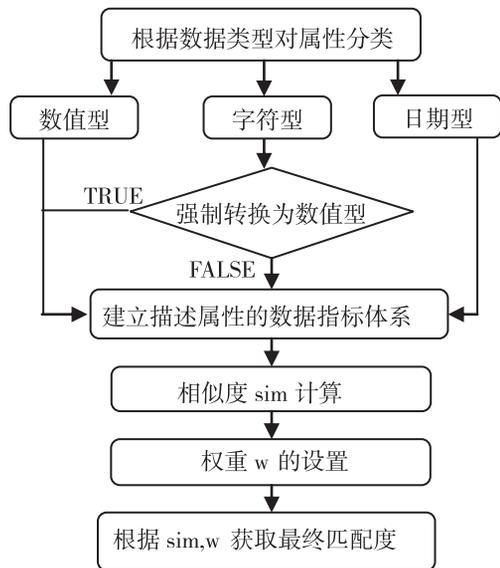


图 3 基于属性值的匹配过程

图 3 给出了处理基于属性值的匹配问题的流程。具体过程如下:

1. 根据数据类型对属性分类

根据描述属性的数据类型的异同,将数据类型分为三大类。本文采用 C#语言实现,数据类型为以

下所示:

数值型 (System. Int32, System. Int64, System. Int16, System. Byte, System. Decimal, System. Double, System. Single)

字符型 (System. String)

日期型 (System. DateTime)

由于其他稀有类型在电子审计数据中极少使用,所以本文暂不考虑其他稀有类型。

2. 属性类型强制转换

先将属性类型转换为相同的数据类型后再比较,即源表与目标表的属性类型须相同才能进行比较。我们先对属性做检测,如果字符中所含的数据全为数值,就将其转换为数值型,这就可以和其他数值型字段进行比较,而不再和字符型字段进行比较。

3. 建立描述属性的数据指标体系

我们将采用含有 7 个元素的一维数组装载指标值,具体的指标体系定义如下:

(1) 数值型

长度,小数位数,平均值,标准差,最大值,最小值,1 表示其为数值型。

(2) 字符型

0,0,字符串长度平均值,字符串长度标准差,字符串长度最大值,字符串长度最小值,-1 表示其为字符型。

(3) 日期型

日期型将不建立指标体系,数组的前 6 个元素均为 0,最后 1 个元素也置为 0,表示其为日期型。

4. 相似度计算

相似度主要是指目标表字段与源表字段的匹配程度。以下为具体计算过程,其中 $f_{ori}[]$ 为源表的字段指标值, $f_{aim}[]$ 为目标表的字段指标值。

if ($f_{ori}[] = 0$ and $f_{aim}[] = 0$) // 如果目标表、源表字段指标值均为 0, 则:

sim = 1

else if ($f_{ori}[] \neq 0$ and $f_{aim}[] = 0$) // 如果只有目标表字段的指标值为 0, 则:

sim = ∞

else // 如果目标表字段指标值不为 0, 则:

sim = $f_{ori}[] / f_{aim}[]$

5. 权重设置

本文采用 RC 方法计算指标的权重。RC 法是一种所有属性根据重要性排序的权值法^[11]。

在 RC 方法中,若所有属性的特征向量根据其重要性排序,则重要性为 S_k 特征向量的权重 w_k 计算如下:

$$w_k = \frac{1}{S} \sum_{t=1}^s \frac{1}{t}$$

计算所得权重如下： w_1 指标长度为0.4083， w_2 小数位数为0.2417， w_3 标准差为0.1583， w_4 平均值为0.1028， w_5 最大值为0.0611， w_6 最小值为0.0278。

6. 获取最终匹配度

计算出指标的权重后，再根据步骤4中获得的相似度获取目标表字段与源表字段的最终匹配度 Sim。采用的计算方法是：

$$\text{sim} = \sum_{i=1}^6 w_i \times \text{sim}_i$$

在这里，我们就获得了目标表字段与源表字段的基于属性值的匹配度。匹配度的取值范围为无穷小到无穷大，但是当匹配度越接近正1时，表明两个字段的匹配程度越高。基于属性值和基于属性名的匹配度的加权和最接近1的即为与目标表最匹配的源表。该匹配方法可实现从众多源表中挑选出最匹配目标表的一个源表，适用于具有多个源表的电子审计数据环境。

四、实验与结果分析

对基于属性值和属性命名规则的匹配技术的验证，主要通过实际系统的运行获得相应指标，以论证其有效性。除了考虑系统性能外，结果的准确度最

能反映其有效性。实验数据来源于 SQLServer2005 数据库。为了使实验符合审计环境，我们将采用电子审计数据作为数据源，以科目余额表作为实验数据库表对象。为了保证实验的客观性，使其更符合实际运用环境，本实验将模拟几种常用财务软件数据库表，建立4组源表数据：“中财信”、“用友”、“金蝶”和“灵狐2000”财务管理软件源数据。

经过实验发现，在4组数据中，基于值与规则的属性匹配对源表的选择全部正确，在字段查准率上也达到了75%。不准确字段主要集中在目标表的金额字段，其中 credit_amount 在4组数据中均没有得到正确的匹配，opening_balance 在3组数据中没有得到正确的匹配，说明其对金额字段识别的准确性有待提高，但对其他字段的匹配都表现良好。具体匹配结果如表1所示。

在与只基于规则的方法和只基于值的方法的比较过程中，所得到的表查准率、字段查准率的实验结果对比见表2。

可以看出，只基于规则的方法的字段查准率、表查准率都为最低，说明完全依赖于规则的效果较差；只基于值的方法的表查准率也有达到100%的，但其字段查准率略低于同时基于值与规则的属性匹配，说明基于命名规则的方法在匹配过程中发挥了作用，提

表1 基于值与规则的属性匹配结果

数据类型	字符型			数值型			
目标表属性	account_code	account_name	opening_balance	debit_amount	credit_amount	balance_year	balance_month
源表中财信“hz”属性	KMDM	KMMC	NCYE	JFE	(YEFX)	YEN	YEY
源表用友“GL_ACCSUM”属性	CCODE	CCODE_NAME	(MB)	MB	(CCODE)	IYEAR	IPERIOD
源表金蝶“T_BALANCE”属性	FNUMBER	FNAME	(FJAMOUNT)	FBEGINBALANCE (FNUMBER)		FYEAR	FPERIOD
源表灵狐2000“科目余额表”属性	科目编码	科目名称	(借方余额)	借方余额	(科目编码)	余额年度	余额月份

注：括号内为匹配错误的表或属性。

高了字段的查准率。总之，基于值与规则的属性匹配技术与其他方法相比，匹配结果的准确率最高。我们通过实验证明了基于值与规则的属性匹配技术是适用于电子审计数据的。

表2 查准率结果对比

方法	字段查准率	表查准率
只基于值的方法	67.8%	100%
只基于规则的方法	39.3%	75%
基于值与规则的方法	75%	100%

五、结束语

在计算机审计数据采集过程中,审计人员采集了电子审计数据后,需要手动匹配源数据库与目标数据库的表与字段,在手动匹配前审计人员还需了解财务软件的数据库表与字段的经济含义。本文提出的基于智能匹配的数据采集技术,可以解决计算机审计数据采集中存在的大量字段匹配问题,提高数据采集的效率。本文的研究内容只是计算机审计数据采集中的一部分,未来可以在此基础上扩展基于属性值与命名规则的匹配技术的适用性,综合解决计算机审计数据采集中存在的问题。未来可以深入开展的工作主要有如下两个方面:

一是数据采集过程的整体实现。目前国内计算机审计中还没有完善的数据采集方案,未来的工作可以考虑实现自动数据采集的整体方案。另外,在审计电子数据分析中,我们还可以考虑智能技术的使用,通过智能技术减少人为分析与判断,提高审计分析效率。

二是审计环境中的异构数据问题。实际审计环境中存在着大量的异构情况。异构数据库语义集成问题比较复杂,涉及数据库技术、人工智能理论等多门学科,所以需要在现有的研究基础上进一步做更深入的研究。

[参考文献]

[1]张娟,廖洪.基于IT技术的连续审计:综述与展望[J].审

计研究,2006(5):79-84.

- [2]Chou C L, Du T, Lai V S. Continuous auditing with a multi-agent system[J]. Decision Support Systems, 2007, 42(4): 2274-2292.
- [3]Flowerday S, Blundell A, Solms R V. Continuous auditing technologies and models: a discussion[J]. Computers & Security, 2006, 25(5):325-331.
- [4]Murthy U S, Groomer S M. A continuous auditing web services model for XML-based accounting systems[J]. International Journal of Accounting Information Systems, 2004, 5(2):139-163.
- [5]尹平,戚振东.国家治理视角下的中国政府审计特征研究[J].审计与经济研究,2010(3):9-14.
- [6]于玉林.会计规范研究方法的剖析与回归[J].审计与经济研究,2010(3):3-8.
- [7]王浙勤,陈瑶.信息特性和审计信息风险[J].审计与经济研究,2009(6):20-24.
- [8]Li W S, Clifton C, Liu S. Database integration using neural networks, implementation and experiences [J]. Knowledge and Information Systems, 2000, 2(1):73-96.
- [9]张正,左春,王裕国,等.基于语义的领域构件接口名称匹配方法[J].通信学报,2007(5):73-79.
- [10]黄广君,孙建国,罗俊丽.基于潜语义分析的概念名称相似度算法[J].计算机工程,2009(14):69-71.
- [11]Barron F H, Baertrt B E. Decision quality using ranked attribute weighs [J]. Management Science, 1996, 42(11): 1515-1523.

[责任编辑:马志娟]

A Research on Data Acquisition in Computer Audit Environment Based on Intelligent Matching Technologies

YE Huan-zhuo, HE Yu-ning

(School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430060, China)

Abstract: Nowadays, the main problem for data acquisition in computer audit environment is the huge work of matching the source and target databases' tables and fields. To solve these problems, we present a field-matching technology to improve data acquisition process, which can realize auto-matching of the source and target table's fields. It reduces manual involvement. The experimental results show that this field-matching technology has a good performance for correct audit data.

Key Words: computer auditing; data acquisition; intelligent matching