



(12) 发明专利申请

(10) 申请公布号 CN 122262210 A

(43) 申请公布日 2026. 06. 23

(21) 申请号 202610178432.4

(22) 申请日 2026.02.06

(71) 申请人 中煤科工开采研究院有限公司

地址 101399 北京市顺义区中关村科技园
区顺义园临空二路1号

(72) 发明人 吕依濛

(74) 专利代理机构 北京清亦华知识产权代理事

务所(普通合伙) 11201

专利代理师 单冠飞

(51) Int. Cl.

G06F 16/25 (2019.01)

G06F 16/2457 (2019.01)

G06F 18/22 (2023.01)

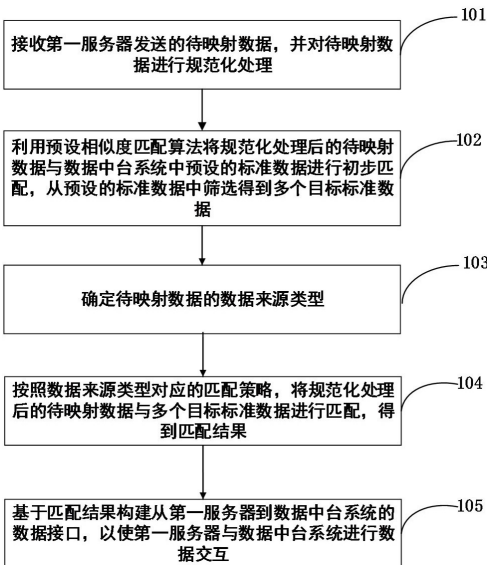
权利要求书4页 说明书15页 附图2页

(54) 发明名称

数据中台系统的接口数据自动映射方法及装置

(57) 摘要

本公开是关于一种数据中台系统的接口数据自动映射方法及装置。其中,方法包括:接收第一服务器发送的待映射数据,并对所述待映射数据进行规范化处理;利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配,从预设的标准数据中筛选得到多个目标标准数据;确定所述待映射数据的数据来源类型;按照所述数据来源类型对应的匹配策略,将规范化处理后的待映射数据与所述多个目标标准数据进行匹配,得到匹配结果;基于所述匹配结果构建从所述第一服务器到所述数据中台系统的数据接口,以使所述第一服务器与所述数据中台系统进行数据交互。本方案可以实现数据接口的快速生成。



1. 一种数据中台系统的接口数据自动映射方法,其特征在于,包括:

接收第一服务器发送的待映射数据,并对所述待映射数据进行规范化处理;所述第一服务器为数据接口服务的发起方服务器,或者数据接口服务的接收方服务器;

利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配,从所述预设的标准数据中筛选得到多个目标标准数据;

确定所述待映射数据的数据来源类型;其中,所述数据来源类型包括生产设备对应的第一类型和业务系统对应的第二类型;

按照所述数据来源类型对应的匹配策略,将规范化处理后的待映射数据与所述多个目标标准数据进行匹配,得到匹配结果;其中,所述第一类型对应基于统计特征的匹配策略;所述第二类型对应基于文本相似度的匹配策略;

基于所述匹配结果构建从所述第一服务器到所述数据中台系统的数据接口,以使所述第一服务器与所述数据中台系统进行数据交互。

2. 根据权利要求1所述的数据中台系统的接口数据自动映射方法,其特征在于,所述待映射数据包括待映射数据名和待映射数据值;所述目标标准数据包括标准数据名和标准数据值;

所述按照所述数据来源类型对应的匹配策略,将规范化处理后的待映射数据与所述多个目标标准数据进行匹配,得到匹配结果,包括:

针对每个目标标准数据,按照所述数据来源类型对应的匹配策略,分别确定所述目标标准数据的标准数据名与所述待映射数据名之间的名称匹配度,以及,分别确定所述目标标准数据的标准数据值与所述待映射数据值之间的数值匹配度;

采用以下公式计算所述待映射数据名与所述数据中台中全部标准数据名之间的名称综合匹配度Match(name):

$$\text{Match}(\text{name}) = \frac{\sum_{k=1}^n \frac{x(k)}{k}}{\ln(n)}$$

其中,x(k)为第k个目标标准数据的标准数据名与所述待映射数据名之间的名称匹配度,n为所述数据中台中预设的标准数据总数量;

采用以下公式计算所述待映射数据值与所述数据中台中全部标准数据值之间的数值综合匹配度Match(value):

$$\text{Match}(\text{value}) = \frac{\sum_{k=1}^n \frac{y(k)}{n+k}}{\ln 2}$$

其中,y(k)为第k个目标标准数据的目标标准数据值与所述待映射数据值之间的数值匹配度;

采用以下公式计算待映射数据匹配值,基于所述待映射数据匹配值确定所述匹配结果:

$$\text{Match} = \sqrt{c1 * \text{Match}(\text{name})^2 + c2 * \text{Match}(\text{value})^2}$$

其中,Match为待映射数据匹配值,c1为名称综合匹配度的调节系数,c2为数值综合匹

配度的调节系数。

3. 根据权利要求2所述的数据中台系统的接口数据自动映射方法,其特征在于,所述按照所述数据来源类型对应的匹配策略,分别确定所述目标标准数据的标准数据名与所述待映射数据名之间的名称匹配度,以及,分别确定所述目标标准数据的标准数据值与所述待映射数据值之间的数值匹配度,包括:

在所述数据来源类型为第一类型的情况下,采用以下公式计算目标标准数据的标准数据名与所述待映射数据名之间的名称匹配度:

$$x(k) = \frac{1}{2} (a1 * v(p_name) + a2 * \frac{m}{p})$$

其中, $x(k)$ 为第 k 个目标标准数据的标准数据名与所述待映射数据名之间的名称匹配度, $v(p_name)$ 为所述待映射数据的数据价值,是基于预设的针对煤炭开采领域的专业词汇数据库确定的, m 为预先存储的所述待映射数据所属生产设备的数据名称总量; p 为预先存储的全部生产设备的数据名称总量, $a1$ 为所述数据价值的调节系数, $a2$ 为第一总数据量在第二总数据量中占比值的调节系数;

采用以下公式计算目标标准数据的目标标准数据值与所述待映射数据值之间的数值匹配度:

$$y(k) = \frac{1}{2} (b1 * \text{Cosine}(f_value, m_value(k)) + b2 * \frac{r}{s})$$

其中, $\text{Cosine}(f_value, m_value(k))$ 为用于计算待映射数据值 f_value 与第 k 个标准数据值 $m_value(k)$ 之间相似度的余弦相似度函数, r 为预先存储的所述待映射数据所属生产设备的数据值总量; s 为预先存储的全部生产设备的数据值总量, $b1$ 为余弦相似度函数的调节系数, $b2$ 为 r/s 的调节系数。

4. 根据权利要求2所述的数据中台系统的接口数据自动映射方法,其特征在于,所述按照所述数据来源类型对应的匹配策略,分别确定所述目标标准数据的标准数据名与所述待映射数据名之间的名称匹配度,以及,分别确定所述目标标准数据的标准数据值与所述待映射数据值之间的数值匹配度,包括:

在所述数据来源类型为第二类型的情况下,采用以下公式计算目标标准数据的目标标准数据名与所述待映射数据名之间的名称匹配度:

$$x(k) = \frac{1}{3} (a1 * v(p_name) + a2 * \text{Levenshtein}(f_name, m_name(k)) + a3 * \text{Jaccard}(f_name, m_name(k)))$$

其中, $x(k)$ 为第 k 个目标标准数据的标准数据名与所述待映射数据名之间的名称匹配度, $v(p_name)$ 为所述待映射数据的数据价值,是基于预设的针对煤炭开采领域的专业词汇数据库确定的, $\text{Levenshtein}(f_name, m_name(k))$ 为用于计算待映射数据名 f_name 与第 k 个标准数据名 $m_name(k)$ 之间相似度的莱文斯坦距离函数, Jaccard 为杰卡德相似度函数, $a1$ 为数据价值的调节系数, $a2$ 为杰卡德相似度函数的调节系数, $a3$ 为杰卡德相似度函数的调节系数;

采用以下公式计算目标标准数据的标准数据值与所述待映射数据值之间的数值匹配度:

$$y(k) = \frac{1}{2} (b1 * \text{Cosine}(f_value, m_value(k)) + b2 * \text{TF-IDF}(f_value, m_value(k)))$$

其中, $y(k)$ 为第 k 个目标标准数据的标准数据值与所述待映射数据值之间的数值匹配度, Cosine 为余弦相似度函数, TF-IDF 为词频-逆文档频率相似度函数, b_1 为余弦相似度函数的调节系数, b_2 为 TF-IDF 函数的调节系数。

5. 根据权利要求3或4所述的数据中台系统的接口数据自动映射方法, 其特征在于, 所述 $v(p_name)$ 采用以下公式计算得到:

$$v(p_name) = \frac{\sum_{q=1}^Q \text{TF-IDF}(t(q)) * (\text{Num}(t(q)) + v - \text{eva})}{Q}$$

其中, p_name 为所述专业词汇数据库中存储的目标字段, 是通过将所述待映射数据名与所述专业词汇库中的全部字段进行相似度匹配后得到的, Q 为所述专业词汇数据库中与所述 p_name 关联字段的总数量, $\text{TF-IDF}(t(q))$ 为计算 p_name 与第 q 个关联字段之间相似度的 TF-IDF 函数, $\text{Num}(tn)$ 为所述 p_name 出现在所述专业词汇库内行业知识文本中的次数, $v - \text{eva}$ 为 p_name 的预设估计价值。

6. 根据权利要求2所述的数据中台系统的接口数据自动映射方法, 其特征在于, 所述基于所述匹配结果构建从所述第一服务器到所述数据中台系统的数据接口, 包括:

在待映射数据匹配值 Match 大于或等于预设匹配值的情况下, 确定所述第一服务器与所述数据中台系统匹配成功, 构建从所述第一服务器到所述数据中台系统的数据接口;

在待映射数据匹配值 Match 小于所述预设匹配值的情况下, 确定所述待映射数据与所述数据中台系统匹配失败, 输出匹配失败信息, 以使用户手动匹配所述待映射数据的数据接口。

7. 根据权利要求1所述的数据中台系统的接口数据自动映射方法, 其特征在于, 方法还包括:

在所述第一服务器为发起方服务器的情况下, 如果所述匹配结果为匹配成功, 则将每个目标标准数据的更新度加一, 如果所述匹配结果为匹配失败, 则将所述规范化处理后的待映射数据存储为新的标准数据, 将所述新的标准数据的初始更新度设置为0;

在所述第一服务器为接收方服务器的情况下, 如果所述匹配结果为匹配成功, 则将每个目标标准数据的匹配度加一, 如果所述匹配结果为匹配失败, 则将所述规范化处理后的待映射数据存储为新的标准数据, 将所述新的标准数据的初始匹配度设置为0;

在所述从所述预设的标准数据中筛选得到目标标准数据之后, 方法还包括:

计算每个目标标准数据的更新度与匹配度之和, 得到每个目标标准数据的第一和值;

在存在大于预设和值的目标第一和值的情况下, 确定所述匹配结果为匹配成功, 执行所述基于所述匹配结果构建从所述第一服务器到所述数据中台系统的数据接口的步骤;

在不存在大于预设和值的目标第一和值的情况下, 执行所述确定所述待映射数据的数据来源类型的步骤。

8. 根据权利要求1所述的数据中台系统的接口数据自动映射方法, 其特征在于, 所述方法还包括:

获取对所述数据接口中交互数据的人工干预修改次数, 在所述人工干预修改次数大于预设次数的情况下, 停用所述数据接口。

9. 根据权利要求1所述的数据中台系统的接口数据自动映射方法, 其特征在于, 所述方

法还包括：

获取所述数据接口的正确对接数据的正确数据量，以及所述数据接口的数据传输总量；

在所述正确数据量在所述数据传输总量中的占比值小于或者等于预设占比值的情况下，停用所述数据接口；

在停用所述数据接口预设时长之后，若所述数据接口的停用次数小于预设次数，重新启用所述数据接口。

10. 一种数据中台系统的接口数据自动映射装置，其特征在于，包括：

处理单元，用于接收第一服务器发送的待映射数据，并对所述待映射数据进行规范化处理；所述第一服务器为数据接口服务的发起方服务器，或者数据接口服务的接收方服务器；

筛选单元，用于利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配，从所述预设的标准数据中筛选得到多个目标标准数据；

确定单元，用于确定所述待映射数据的数据来源类型；其中，所述数据来源类型包括生产设备对应的第一类型和业务系统对应的第二类型；

匹配单元，用于按照所述数据来源类型对应的匹配策略，将规范化处理后的待映射数据与所述多个目标标准数据进行匹配，得到匹配结果；其中，所述第一类型对应基于统计特征的匹配策略；所述第二类型对应基于文本相似度的匹配策略；

构建单元，用于基于所述匹配结果构建从所述第一服务器到所述数据中台系统的数据接口，以使所述第一服务器与所述数据中台系统进行数据交互。

数据中台系统的接口数据自动映射方法及装置

技术领域

[0001] 本公开涉及煤炭开采技术领域,尤其涉及一种数据中台系统的接口数据自动映射方法及装置。

背景技术

[0002] 在煤炭开采领域,生产过程中涉及的采煤机、液压支架等设备种类繁多、品牌各异,实时产生海量多源异构数据。这些数据普遍存在标准不一、格式分散、一数多源等问题,导致各系统间难以实现精准、高效的接口对接。传统依赖人工配置或固定规则的映射方法,不仅工作量大、响应迟缓,且难以适应设备更新与业务扩展,形成了严重的数据孤岛,无法支撑生产协同、智能预警等精细化管控需求,制约了矿山数字化转型与智能化发展。

发明内容

[0003] 为克服相关技术中存在的问题,本公开提供一种数据中台系统的接口数据自动映射方法及装置。

[0004] 根据本公开实施例的第一方面,提供一种数据中台系统的接口数据自动映射方法,包括:

[0005] 接收第一服务器发送的待映射数据,并对所述待映射数据进行规范化处理;所述第一服务器为数据接口服务的发起方服务器,或者数据接口服务的接收方服务器;

利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配,从所述预设的标准数据中筛选得到多个目标标准数据;

确定所述待映射数据的数据来源类型;其中,所述数据来源类型包括生产设备对应的第一类型和业务系统对应的第二类型;

按照所述数据来源类型对应的匹配策略,将规范化处理后的待映射数据与所述多个目标标准数据进行匹配,得到匹配结果;其中,所述第一类型对应基于统计特征的匹配策略;所述第二类型对应基于文本相似度的匹配策略;

基于所述匹配结果构建从所述第一服务器到所述数据中台系统的数据接口,以使所述第一服务器与所述数据中台系统进行数据交互。

[0006] 根据本公开实施例的第二方面,提供一种数据中台系统的接口数据自动映射装置,其特征在于,包括:

处理单元,用于接收第一服务器发送的待映射数据,并对所述待映射数据进行规范化处理;所述第一服务器为数据接口服务的发起方服务器,或者数据接口服务的接收方服务器;

筛选单元,用于利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配,从所述预设的标准数据中筛选得到多个目标标准数据;

确定单元,用于确定所述待映射数据的数据来源类型;其中,所述数据来源类型包

括生产设备对应的第一类型和业务系统对应的第二类型；

匹配单元,用于按照所述数据来源类型对应的匹配策略,将规范化处理后的待映射数据与所述多个目标标准数据进行匹配,得到匹配结果;其中,所述第一类型对应基于统计特征的匹配策略;所述第二类型对应基于文本相似度的匹配策略;

构建单元,用于基于所述匹配结果构建从所述第一服务器到所述数据中台系统的数据接口,以使所述第一服务器与所述数据中台系统进行数据交互。

[0007] 根据本公开实施例的第三方面,一种电子设备,包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,所述处理器执行所述计算机程序时,实现如第一方面中任一项所述的方法。

[0008] 根据本公开实施例的第四方面,提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现如第一方面中任一项所述的方法。

[0009] 根据本公开实施例的第五方面,提供一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现如第一方面中任一项所述的方法。

[0010] 本公开的实施例提供的技术方案可以包括以下有益效果:通过接收第一服务器发送的待映射数据,并对所述待映射数据进行规范化处理;利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配,从所述预设的标准数据中筛选得到多个目标标准数据;确定所述待映射数据的数据来源类型;按照所述数据来源类型对应的匹配策略,将规范化处理后的待映射数据与所述多个目标标准数据进行匹配,得到匹配结果;基于所述匹配结果构建从所述第一服务器到所述数据中台系统的数据接口,以使所述第一服务器与所述数据中台系统进行数据交互。智能识别数据来源于生产设备或业务系统,并据此分别采用融合统计特征与领域知识的策略或综合文本相似度的策略进行精准匹配,最终依据匹配结果自动化构建或装配数据接口,从而在无需人工一对一编码的情况下,有效解决了多源异构系统间因数据标准不一导致的数据对接难题,实现了数据的高效整合与接口的快速生成。

[0011] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0012] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本发明的实施例,并与说明书一起用于解释本发明的原理。

[0013] 图1是根据一示例性实施例示出的一种数据中台系统的接口数据自动映射方法的流程图。

[0014] 图2是根据一示例性实施例示出的一种数据中台系统的接口数据自动映射装置的框图。

[0015] 图3是根据一示例性实施例示出的一种用于数据中台系统的接口数据自动映射方法的装置的框图。

具体实施方式

[0016] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及

附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

[0017] 在本公开实施例使用的术语是仅仅出于描述特定实施例的目的,而非旨在限制本公开实施例。在本公开实施例和所附权利要求书中所使用的单数形式的“一种”和“该”也旨在包括多数形式,除非上下文清楚地表示其他含义。

[0018] 应当理解,尽管在本公开实施例可能采用术语第一、第二、第三等来描述各种信息,但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如,在不脱离本公开实施例范围的情况下,第一信息也可以被称为第二信息,类似地,第二信息也可以被称为第一信息。取决于语境,如在此所使用的词语“如果”及“若”可以被解释成为“在……时”或“当……时”或“响应于确定”。

[0019] 此外,可以使用本公开实施例所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发申请中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0020] 需要说明的是,本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。

[0021] 相关技术中,数据中台是一种数据管理架构,提供数据采集、处理、存储和服务等功能,支持跨业务系统的数据共享和协作。目前,煤炭开采领域涉及大量生产时序数据的采集和设备直接数据的对接工作,这些生产设备包括采煤机、掘进机、液压履带钻机、运输机、转载机、连采集、输水车等生产设备,这些生产设备在生产过程中产生大量实时数据,数据存在数量大、信息密度低、实时性高、不同设备存在一数多源、数据分散重复、各个设备及相关系统没有统一的数据标准,各自系统相同数据格式存在一定的相似度,但由于设备五花八门,品牌标准不同,无法做到一对一精准匹配对接,无法满足生产流程协同、各项生产数据同步的需求,以及随着信息化建设的迅速发展及开采业务发展,矿上生产数据会逐渐增多。迫切需要建立各个生产设备及系统的统一数据对接、打破信息壁垒、贯通数据共享、提升数据质量及数据应用价值。依据煤炭开采领域的的数据特点,在无法保证一对一精准接口对接的情况下,借助数据中台系统,构建统一的数据中间层,通过数据清洗、映射、归一化处理,实现多源异构数据的标准化整合。

[0022] 根据中煤科工开采研究院数字化转型规划,通过大数据中台系统建设,稳步推进开采院历史资料数据化、业务管理数字化和数据采集的自动化,以提高数据资源的完整性和一致性。通过整合现有业务系统,优化管理流程,完善数据内容,确保所需数据的完整性以满足数据应用的需要。同时为中煤科工的开采研究和未来管理提供数据支撑,并促进数字化与业务提升的转型,统一行业数据标准,助力行业发展。

[0023] 为了解决上述问题,本公开提供了一种数据中台系统的接口数据自动映射方法及装置,通过接收第一服务器发送的待映射数据,并对所述待映射数据进行规范化处理;利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配,从所述预设的标准数据中筛选得到多个目标标准数据;确定所述待映射数据的数据来源类型;按照所述数据来源类型对应的匹配策略,将规范化处理后的待映射数据与所述多个目标标准数据进行匹配,得到匹配结果;基于所述匹配结果构建从所述第一服

务器到所述数据中台系统的数据接口,以使所述第一服务器与所述数据中台系统进行数据交互。智能识别数据来源于生产设备或业务系统,并据此分别采用融合统计特征与领域知识的策略或综合文本相似度的策略进行精准匹配,最终依据匹配结果自动化构建或装配数据接口,从而在无需人工一对一编码的情况下,有效解决了多源异构系统间因数据标准不一导致的数据对接难题,实现了数据的高效整合与接口的快速生成。

[0024] 图1是根据一示例性实施例示出的一种数据中台系统的接口数据自动映射方法的流程图,如图1所示,需要说明的是,本公开实施例的数据中台系统的接口数据自动映射方法应用于数据中台系统的接口数据自动映射装置中。如图1所示,该方法可以包括以下步骤:

步骤101,接收第一服务器发送的待映射数据,并对待映射数据进行规范化处理。

[0025] 其中,第一服务器为数据接口服务的发起方服务器,或者数据接口服务的接收方服务器。

[0026] 在本申请实施例中,第一服务器是相对于数据中台而言的外部系统实体,可以是需要与数据中台系统进行数据交换的源系统或目标系统。

[0027] 具体而言,当第一服务器作为数据接口服务的发起方时,是源系统服务器,负责向数据中台提供待整合的原始数据;当第一服务器作为数据接口服务的接收方时,是目标系统服务器,向数据中台请求消费标准化后的数据。无论作为提供方还是消费方,第一服务器均通过本公开实现的自动映射接口与数据中台进行连接,从而将自身从传统的、复杂的点对点系统对接中解耦出来,简化为统一与数据中台进行标准化数据交互的角色。

[0028] 在本申请实施例中,待映射数据可以包括数据名称、数据值及系统标识等信息,是数据中台进行自动化识别与匹配的输入对象。无论是作为提供方的源系统上传的生产或业务数据,还是作为消费方的目标系统发起的数据请求,均可以待映射数据的形式进入数据中台处理流程。

[0029] 在一个实施例中,数据中台系统可以对待映射数据进行规范化处理,例如名称去除*, -, ”等非规范符号,统一用正则表达式规范规则规范,数值将非结构化数据值转换为地址,地址指向实际非结构化数据的值,系统标识f_sys使用统一的命名规则,能够和各个系统一一映射。

[0030] 步骤102,利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配,从预设的标准数据中筛选得到多个目标标准数据。

[0031] 在一个实施例中,可以对已完成规范化处理的待映射数据进行初步匹配,调用预设相似度匹配算法,将待映射数据中的关键标识(如数据名称)与数据中台系统中预存的所有标准数据进行快速比对。该过程旨在从全量标准数据中高效筛选出与待映射数据具有一定相似性的多个候选数据项,也即目标标准数据,大幅缩减了后续精确匹配的计算范围,提升了整体处理效率。

[0032] 作为一种可能实施方式的示例,预设相似度匹配算法可以是莱文斯坦距离算法,还可以是杰卡德相似度算法。

[0033] 步骤103,确定待映射数据的数据来源类型。

[0034] 其中,数据来源类型包括生产设备对应的第一类型和业务系统对应的第二类型。

[0035] 在一个实施例中,可以根据待映射数据的元信息(例如来源系统标识f_sys)及其

数据特征(如字段命名规律、数值格式、更新频率等),通过规则匹配或预训练的分类模型,将其区分为“生产设备对应的第一类型”或“业务系统对应的第二类型”。

[0036] 需要说明的是,生产设备数据通常具有高实时性、时序性强、信息密度低但数据规模大的特点;而业务系统数据则更偏向结构化、语义明确、文本信息丰富。这一分类结果将直接决定后续匹配阶段所采用的策略与算法组合,是系统实现按照数据来源进行匹配、提升映射精度的前提。

[0037] 步骤104,按照数据来源类型对应的匹配策略,将规范化处理后的待映射数据与多个目标标准数据进行匹配,得到匹配结果。

[0038] 其中,第一类型对应基于统计特征的匹配策略;第二类型对应基于文本相似度的匹配策略。

[0039] 在本申请一个实施例中,针对第一类型(生产设备数据),可以采用基于统计特征与领域知识价值融合的匹配策略。该策略不仅计算字段名称的文本相似度,更重点考量数据背后反映的客观统计特征,例如该设备产生的数据总量占比以及数据存储规模占比,并将其与从煤炭开采技术领域专业词汇库中提取的语义价值相结合进行加权计算。从而能够应对生产设备数据“信息密度低但规模显著、专业术语性强”的特点,使匹配结果更贴合工业场景下的数据重要性分布与实际煤炭开采业务含义。

[0040] 在本申请另一个实施例中,对于第二类型(业务系统数据),可以采用基于文本相似度与领域知识价值融合的匹配策略。该策略充分利用业务数据结构化程度高、文本信息丰富的特点,综合运用莱文斯坦距离、杰卡德相似度等算法评估名称相似性,并运用余弦相似度与TF-IDF相似度对数据值进行深层次语义比对。所有这些文本相似度指标均与煤炭开采领域知识价值协同计算,确保匹配过程既能捕捉细微的文字差异,又能理解字段在专业语境下的真实意图,从而显著提升对煤炭开采领域相关的财务、科研、管理等业务系统复杂数据的映射准确性。

[0041] 在本公开一些实施例中,待映射数据包括待映射数据名和待映射数据值,目标标准数据包括标准数据名和标准数据值,步骤104具体可以包括以下步骤:

步骤a1,针对每个目标标准数据,按照数据来源类型对应的匹配策略,分别确定目标标准数据的标准数据名与待映射数据名之间的名称匹配度,以及,分别确定目标标准数据的标准数据值与待映射数据值之间的数值匹配度。

[0042] 步骤a2,采用以下公式计算待映射数据名与数据中台中全部标准数据名之间的名称综合匹配度Match(name):

$$\text{Match}(\text{name}) = \frac{\sum_{k=1}^n \frac{x(k)}{k}}{\ln(n)}$$

[0043] 其中,x(k)为第k个目标标准数据的标准数据名与待映射数据名之间的名称匹配度,n为数据中台中预设的标准数据总数量。

[0044] 在本申请实施例中,综合考虑了所有目标标准数据的匹配贡献,避免了单一最优匹配可能存在的偶然性,提高了评估的稳健性,此外,通过对数函数ln(n)对候选数量n进行缓变归一化,使得匹配度结果在标准数据规模大幅增长时仍能保持数值稳定,不会因候选集扩大而被过度稀释,从而确保了匹配度评估在不同数据规模下的可比性与系统扩展适应

性,为后续接口自动构建提供了可靠、一致的决策依据。

[0045] 步骤a3,采用以下公式计算待映射数据值与数据中台中全部标准数据值之间的数值综合匹配度Match(value):

$$\text{Match (value)} = \frac{\sum_{k=1}^n \frac{y(k)}{n+k}}{\ln 2}$$

[0046] 其中,y(k)为第k个目标标准数据的目标标准数据值与待映射数据值之间的数值匹配度。

[0047] 在本申请实施例中,通过将各个候选标准数据的数值匹配度y(k)以特定权重进行聚合,计算出数值综合匹配度,有效融合了不同候选数据在数值层面的相似性贡献,尤其通过设计合理的加权与归一化机制(如利用n+k调节权重分布,并通过ln函数稳定输出范围),使得评估结果既能反映最相似标准数据的主导作用,又能平滑处理候选集大小变动带来的波动,从而提供稳定、可比的数值相似性度量,与名称匹配度相辅相成,共同构成全面、可靠的自动映射决策基础。

[0048] 步骤a4,采用以下公式计算待映射数据匹配值,基于待映射数据匹配值确定匹配结果:

$$\text{Match} = \sqrt{c1 * \text{Match}(\text{name})^2 + c2 * \text{Match}(\text{value})^2}$$

[0049] 其中,Match为待映射数据匹配值,c1为名称综合匹配度的调节系数,c2为数值综合匹配度的调节系数。

[0050] 在本申请一些实施例中,c1和c2可以根据实际情况进行预先设定和调整。

[0051] 在本公开一些实施例中,步骤a1具体可以包括以下步骤:

步骤a111,在数据来源类型为第一类型的情况下,采用以下公式计算目标标准数据的标准数据名与待映射数据名之间的名称匹配度:

$$x(k) = \frac{1}{2} (a1 * v(p_name) + a2 * \frac{m}{p})$$

[0052] 其中,x(k)为第k个目标标准数据的标准数据名与待映射数据名之间的名称匹配度,v(p_name)为待映射数据的数据价值,是基于预设的针对煤炭开采领域的专业词汇数据库确定的,m为预先存储的待映射数据所属生产设备的数据名称总量;p为预先存储的全部生产设备的数据名称总量,a1为数据价值的调节系数,a2为第一总数据量在第二总数据量中占比值的调节系数。

[0053] 在本申请实施例中,通过将专业领域知识价值v(p_name)与设备数据规模占比m/p相结合来计算名称匹配度,针对生产设备数据专业性强、信息密度低的特点,在评估名称相似性时不仅考虑字面文本特征,还引入了反映术语重要性的领域知识价值以及体现设备数据贡献度的统计权重,从而有效避免了单纯依赖文本相似度可能导致的专业术语误匹配,确保了关键生产设备及其数据在映射过程中获得与其业务重要性相符的优先级和准确性,显著提升了在煤炭开采等专业领域中设备数据映射的可靠性和业务贴合度。

[0054] 步骤a112,采用以下公式计算目标标准数据的目标标准数据值与待映射数据值之间的数值匹配度:

$$y(k) = \frac{1}{2} (b1 * \text{Cosine}(f_value, m_value(k)) + b2 * \frac{r}{s})$$

[0055] 其中, $\text{Cosine}(f_value, m_value(k))$ 为用于计算待映射数据值 f_value 与第 k 个标准数据值 $m_value(k)$ 之间相似度的余弦相似度函数, r 为预先存储的待映射数据所属生产设备的数据值总量; s 为预先存储的全部生产设备的数据值总量, $b1$ 为余弦相似度函数的调节系数, $b2$ 为 r/s 的调节系数。

[0056] 通过融合语义相似性评估与设备数据规模权重来计算数值匹配度, 针对生产设备数据值往往具有高维、时序性且难以直接比对的特性, 创新性地将设备名称的语义相关性 (通过余弦相似度衡量) 作为数据值相似性的间接有效指示, 同时引入设备数据量占比 (r/s) 这一关键统计特征来反映该设备在整体数据生态中的重要性权重, 从而在缺乏显式数值相似度量度的场景下, 构建出一个既贴合业务语义又符合数据分布规律的实用化匹配指标, 显著增强了生产设备类数值数据在自动映射过程中的可评估性与匹配合理性。

[0057] 在本公开另一些实施例中, 步骤a1具体可以包括以下步骤:

步骤a121, 在数据来源类型为第二类型的情况下, 采用以下公式计算目标标准数据的目标标准数据名与待映射数据名之间的名称匹配度:

$$x(k) = \frac{1}{3} (a1 * v(p_name) + a2 * \text{Levenshtein}(f_name, m_name(k)) + a3 * \text{Jaccard}(f_name, m_name(k)))$$

其中, $x(k)$ 为第 k 个目标标准数据的标准数据名与待映射数据名之间的名称匹配度, $v(p_name)$ 为待映射数据的数据价值, 是基于预设的针对煤炭开采领域的专业词汇数据库确定的, $\text{Levenshtein}(f_name, m_name(k))$ 为用于计算待映射数据名 f_name 与第 k 个标准数据名 $m_name(k)$ 之间相似度的莱文斯坦距离函数, Jaccard 为杰卡德相似度函数, $a1$ 为数据价值的调节系数, $a2$ 为杰卡德相似度函数的调节系数, $a3$ 为杰卡德相似度函数的调节系数。

[0058] 在本申请实施例中, 通过将煤炭开采领域知识价值、莱文斯坦距离和杰卡德相似度三类指标进行融合加权, 实现了对业务系统数据名称的多维度综合评估: 既通过莱文斯坦距离捕捉拼写差异和形态变异, 又通过杰卡德相似度识别共有词汇与结构共性, 同时借助领域知识库赋予专业术语更高的权重, 从而显著提升了在命名规范多样、专业术语密集的业务系统场景下名称匹配的准确性与鲁棒性, 避免了单一算法导致的误判或漏判。

[0059] 步骤a122, 采用以下公式计算目标标准数据的标准数据值与待映射数据值之间的数值匹配度:

$$y(k) = \frac{1}{2} (b1 * \text{Cosine}(f_value, m_value(k)) + b2 * \text{TF-IDF}(f_value, m_value(k)))$$

[0060] 其中, $y(k)$ 为第 k 个目标标准数据的标准数据值与待映射数据值之间的数值匹配度, Cosine 为余弦相似度函数, TF-IDF 为词频-逆文档频率相似度函数, $b1$ 为余弦相似度函数的调节系数, $b2$ 为 TF-IDF 函数的调节系数。

[0061] 在本申请实施例中, 通过结合余弦相似度与 TF-IDF 相似度对业务系统数据值进行匹配度评估, 针对业务数据值常为文本描述、文档内容等非结构化信息的特点, 利用余弦相似度从向量空间角度衡量整体语义方向的接近程度, 同时借助 TF-IDF 相似度聚焦于关键词的权重分布差异, 从而实现对长文本数据值既宏观又精细的双层相似性度量, 这种融合策略有效克服了单一方法在应对文本长度变化、词频分布不均时的局限性, 显著提升了业务系统复杂文本数据在自动映射过程中的语义理解深度与匹配准确性。

[0062] 在本公开一些实施例中, $v(p_name)$ 采用以下公式计算得到:

$$v(p_name) = \frac{\sum_{q=1}^Q TF-IDF(t(q)) * (Num(t(q)) + v_eva)}{Q}$$

[0063] 其中, p_name 为专业词汇数据库中存储的目标字段, 是通过将待映射数据名与专业词汇库中的全部字段进行相似度匹配后得到的, Q 为专业词汇数据库中与 p_name 关联字段的总数量, $TF-IDF(t(q))$ 为计算 p_name 与第 q 个关联字段之间相似度的 $TF-IDF$ 函数, $Num(tn)$ 为 p_name 出现在专业词汇库内行业知识文本中的次数, v_eva 为 p_name 的预设估计价值。

[0064] 需要说明的是, v_eva 是预设的先验价值估计, 以基于领域专家的历史经验、行业知识沉淀或对历史数据价值的挖掘与分析而预先设定, 确保数据匹配模型在量化术语重要性时, 不仅能依据客观的统计特征 (如词频、 $TF-IDF$), 还能融入人类专家的经验判断, 使评估体系更加全面、灵活, 并具备适应实际业务场景动态调整的能力。

[0065] 在本申请实施例中, 通过系统性地融合术语的统计显著性 ($TF-IDF$)、语料出现频次 (Num) 及专家经验预设价值 (v_eva), 为煤炭开采领域的专业词汇提供了量化的价值评估方法, 将传统依赖主观经验的术语重要性判断转化为客观、可计算的度量指标, 使得领域知识能够以结构化、可调参的方式融入自动匹配流程, 从而显著提升了数据映射过程中对专业术语的识别精度与语义理解深度, 确保了后续匹配策略在行业特定语境下的准确性与适应性。

[0066] 步骤105, 基于匹配结果构建从第一服务器到数据中台系统的数据接口, 以使第一服务器与数据中台系统进行数据交互。

[0067] 在一个实施例中, 可以依据匹配结果 (如整体匹配值是否达到阈值) 自动决策: 若匹配成功, 则在数据服务层自动装配或调用对应的标准化数据服务接口, 实现数据的无缝传输; 若匹配未成功, 则触发手动配置流程并生成提示。该步骤实质上是将数据层面的语义匹配, 落地为服务层面的物理连接, 最终实现第一服务器 (作为数据提供方的源系统, 或者作为消费方的目标系统) 与数据中台之间标准化、自动化、可管理的数据交互通道的快速建立与动态维护。

[0068] 在本公开一些实施例中, 步骤105具体可以包括以下步骤:

在待映射数据匹配值 $Match$ 大于或等于预设匹配值的情况下, 确定第一服务器与数据中台系统匹配成功, 构建从第一服务器到数据中台系统的数据接口;

在待映射数据匹配值 $Match$ 小于预设匹配值的情况下, 确定待映射数据与数据中台系统匹配失败, 输出匹配失败信息, 以使用户手动匹配待映射数据的数据接口。

[0069] 在本公开一些实施例中, 方法还可以包括:

在第一服务器为发起方服务器的情况下, 如果匹配结果为匹配成功, 则将每个目标标准数据的更新度加一, 如果匹配结果为匹配失败, 则将规范化处理后的待映射数据存储为新的标准数据, 将新的标准数据的初始更新度设置为0;

在第一服务器为接收方服务器的情况下, 如果匹配结果为匹配成功, 则将每个目标标准数据的匹配度加一, 如果匹配结果为匹配失败, 则将规范化处理后的待映射数据存储为新的标准数据, 将新的标准数据的初始匹配度设置为0;

在从预设的标准数据中筛选得到目标标准数据之后,方法还包括:

计算每个目标标准数据的更新度与匹配度之和,得到每个目标标准数据的第一和值;

在存在大于预设和值的目标第一和值的情况下,确定匹配结果为匹配成功,执行基于匹配结果构建从第一服务器到数据中台系统的数据接口的步骤;

在不存在大于预设和值的目标第一和值的情况下,执行确定待映射数据的数据来源类型的步骤。

[0070] 在本申请实施例中,可以为数据中台内的每个标准数据项维护两个动态置信度指标:更新度(update_number)与匹配度(match_number)。当待映射数据来自发起方服务器(源系统)并匹配成功时,对应标准数据的update_number加一,此操作记录该数据被源头系统“确认”的次数;当请求来自接收方服务器(目标系统)并匹配成功时,对应标准数据的match_number加一,此操作记录该数据接收端成功调用的次数。这两个指标分别从“数据供给”和“数据消费”两个维度,量化了该标准数据项的可靠性与流行度。

[0071] 在此基础上,引入了快速匹配通道以提升效率。在初步筛选出目标标准数据后,可以优先计算每个目标标准数据的置信度总和(也即第一和值update_number+match_number)。若其中任一目标标准数据的第一和值大于或等于预设的快速匹配阈值(即预设和值),则直接判定匹配成功,并跳过后续所有复杂的相似度模型计算,立即执行接口构建步骤。这一机制确保了对那些经过长期实践验证、高度可靠的数据映射关系,能够实现毫秒级的瞬时匹配。

[0072] 此外,这一设计实现了效率与可靠性的平衡,能够将有限的计算资源集中于处理新的、不确定的匹配请求,而对成熟的、高频的映射则利用“记忆”高效完成。同时,通过置信度的持续累积,使得数据中台系统具备了自我学习与演进的能力,映射关系越用越可靠,越可靠则响应越快,形成了一个越用越智能的良性循环,极大地提升了数据中台在长期运行中的整体性能和适应性。

[0073] 在本公开一些实施例中,方法还可以包括:获取对数据接口中交互数据的人工干预修改次数,在人工干预修改次数大于预设次数的情况下,停用数据接口。

[0074] 在本申请实施例中,可以为每个自动装配的接口维护一个人工干预修改次数的计数器(如adjust_number),每当接口输出的数据映射关系因不准确而被管理员手动修正时,该计数器便累加一次。当累计次数超过预设阈值时,判定此接口的自动化映射逻辑存在持续性问题,可靠性不足,进而主动停用该接口。这一机制不仅及时阻止了错误数据的持续扩散,保障了数据服务的质量,还将人工运维反馈转化为了明确的系统信号,为后续的根本分析、模型参数调优(如提高该字段匹配算法的严格性)或接口重构提供了直接的决策依据。

[0075] 在本公开一些实施例中,方法还可以包括:

获取数据接口的正确对接数据的正确数据量,以及数据接口的数据传输总量;

在正确数据量在数据传输总量中的占比值小于或者等于预设占比值的情况下,停用数据接口;

在停用数据接口预设时长之后,若数据接口的停用次数小于预设次数,重新启用数据接口。

[0076] 在本申请实施例中,可以持续监控每个数据接口的传输质量,通过统计正确对接

数据量与数据传输总量并计算其正确率占比；当该占比低于或等于预设 占比值时，表明接口映射可能已出现偏差或源数据特征已发生漂移，可以自动停用该接口以防止错误传播。与此同时，机制并非采取永久封禁策略，而是在停用预设时长后，若该接口的累计停用次数未超过系统允许的上限，则自动重新启用该接口，既给予了接口在数据环境可能恢复稳定后的再次服务机会，也通过停用次数上限防止了低质接口的无限循环重启，实现了对数据接口运行状态的智能化、闭环化管控，确保了数据服务层在长期运行中的动态可靠性与健壮性。

[0077] 根据本公开实施例提出的数据中台系统的接口数据自动映射方法，通过接收第一服务器发送的待映射数据，并对待映射数据进行规范化处理；利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配，从预设的标准数据中筛选得到多个目标标准数据；确定待映射数据的数据来源类型；按照数据来源类型对应的匹配策略，将规范化处理后的待映射数据与多个目标标准数据进行匹配，得到匹配结果；基于匹配结果构建从第一服务器到数据中台系统的数据接口，以使第一服务器与数据中台系统进行数据交互。智能识别数据来源于生产设备或业务系统，并据此分别采用融合统计特征与领域知识的策略或综合文本相似度的策略进行精准匹配，最终依据匹配结果自动化构建或装配数据接口，从而在无需人工一对一编码的情况下，有效解决了多源异构系统间因数据标准不一导致的数据对接难题，实现了数据的高效整合与接口的快速生成。本公开采用元数据管理与动态适配技术，兼容不同设备品牌与通信协议，建立灵活可扩展的数据接入机制。同时强化实时数据采集能力，确保数据时序性与完整性，为上层业务系统提供稳定、高质量的数据支撑，最终实现生产全流程的数据互通与协同管控，通过构建接口自动相似度匹配机制，实现多源异构数据的智能识别与动态映射。

[0078] 图2是根据一示例性实施例示出的一种数据中台系统的接口数据自动映射装置框图。参照图2，该装置包括处理单元201，筛选单元202，确定单元203，匹配单元204和构建单元205。

[0079] 其中，处理单元201，用于接收第一服务器发送的待映射数据，并对待映射数据进行规范化处理；第一服务器为数据接口服务的发起方服务器，或者数据接口服务的接收方服务器；

筛选单元202，用于利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配，从预设的标准数据中筛选得到多个目标标准数据；

确定单元203，用于确定待映射数据的数据来源类型；其中，数据来源类型包括生产设备对应的第一类型和业务系统对应的第二类型；

匹配单元204，用于按照数据来源类型对应的匹配策略，将规范化处理后的待映射数据与多个目标标准数据进行匹配，得到匹配结果；其中，第一类型对应基于统计特征的匹配策略；第二类型对应基于文本相似度的匹配策略；

构建单元205，用于基于匹配结果构建从第一服务器到数据中台系统的数据接口，以使第一服务器与数据中台系统进行数据交互。

[0080] 在本申请一些实施例中，待映射数据包括待映射数据名和待映射数据值；目标标准数据包括标准数据名和标准数据值，匹配单元204具体可以用于：

针对每个目标标准数据,按照数据来源类型对应的匹配策略,分别确定目标标准数据的标准数据名与待映射数据名之间的名称匹配度,以及,分别确定目标标准数据的标准数据值与待映射数据值之间的数值匹配度;

采用以下公式计算待映射数据名与数据中台中全部标准数据名之间的名称综合匹配度Match(name):

$$\text{Match}(\text{name}) = \frac{\sum_{k=1}^n \frac{x(k)}{k}}{\ln(n)}$$

[0081] 其中,x(k)为第k个目标标准数据的标准数据名与待映射数据名之间的名称匹配度,n为数据中台中预设的标准数据总数量;

采用以下公式计算待映射数据值与数据中台中全部标准数据值之间的数值综合匹配度Match(value):

$$\text{Match}(\text{value}) = \frac{\sum_{k=1}^n \frac{y(k)}{n+k}}{\ln 2}$$

[0082] 其中,y(k)为第k个目标标准数据的目标标准数据值与待映射数据值之间的数值匹配度;

采用以下公式计算待映射数据匹配值,基于待映射数据匹配值确定匹配结果:

$$\text{Match} = \sqrt{c1 * \text{Match}(\text{name})^2 + c2 * \text{Match}(\text{value})^2}$$

[0083] 其中,Match为待映射数据匹配值,c1为名称综合匹配度的调节系数,c2为数值综合匹配度的调节系数。

[0084] 在本申请一些实施例,匹配单元204具体可以用于:

在数据来源类型为第一类型的情况下,采用以下公式计算目标标准数据的标准数据名与待映射数据名之间的名称匹配度:

$$x(k) = \frac{1}{2} (a1 * v(p_name) + a2 * \frac{m}{p})$$

[0085] 其中,x(k)为第k个目标标准数据的标准数据名与待映射数据名之间的名称匹配度,v(p_name)为待映射数据的数据价值,是基于预设的针对煤炭开采领域的专业词汇数据库确定的,m为预先存储的待映射数据所属生产设备的数据名称总量;p为预先存储的全部生产设备的数据名称总量,a1为数据价值的调节系数,a2为第一总数据量在第二总数据量中占比值的调节系数;

采用以下公式计算目标标准数据的目标标准数据值与待映射数据值之间的数值匹配度:

$$y(k) = \frac{1}{2} (b1 * \text{Cosine}(f_value, m_value(k)) + b2 * \frac{r}{s})$$

[0086] 其中,Cosine(f_value, m_value(k))为用于计算待映射数据值f_value与第k个标准数据值m_value(k)之间相似度的余弦相似度函数,r为预先存储的所述待映射数据所属生产设备的数据值总量;s为预先存储的全部生产设备的数据值总量,b1为余弦相似度函数的调节系数,b2为r/s的调节系数。

[0087] 在本申请一些实施例中,匹配单元204具体可以用于:

在数据来源类型为第二类型的情况下,采用以下公式计算目标标准数据的目标标准数据名与待映射数据名之间的名称匹配度:

$$x(k) = \frac{1}{3}(a1 * v(p_name) + a2 * \text{Levenshtein}(f_name, m_name(k)) + a3 * \text{Jaccard}(f_name, m_name(k)))$$

其中, $x(k)$ 为第 k 个目标标准数据的标准数据名与待映射数据名之间的名称匹配度, $v(p_name)$ 为待映射数据的数据价值,是基于预设的针对煤炭开采领域的专业词汇数据库确定的, $\text{Levenshtein}(f_name, m_name(k))$ 为用于计算待映射数据名 f_name 与第 k 个标准数据名 $m_name(k)$ 之间相似度的莱文斯坦距离函数, Jaccard 为杰卡德相似度函数, $a1$ 为数据价值的调节系数, $a2$ 为杰卡德相似度函数的调节系数, $a3$ 为杰卡德相似度函数的调节系数;

采用以下公式计算目标标准数据的标准数据值与待映射数据值之间的数值匹配度:

$$y(k) = \frac{1}{2}(b1 * \text{Cosine}(f_value, m_value(k)) + b2 * \text{TF-IDF}(f_value, m_value(k)))$$

[0088] 其中, $y(k)$ 为第 k 个目标标准数据的标准数据值与待映射数据值之间的数值匹配度, Cosine 为余弦相似度函数, TF-IDF 为词频-逆文档频率相似度函数, $b1$ 为余弦相似度函数的调节系数, $b2$ 为 TF-IDF 函数的调节系数。

[0089] 在本申请一些实施例中, $v(p_name)$ 采用以下公式计算得到:

$$v(p_name) = \frac{\sum_{q=1}^Q \text{TF-IDF}(t(q)) * (\text{Num}(t(q)) + v_eva)}{Q}$$

[0090] 其中, p_name 为专业词汇数据库中存储的目标字段,是通过将待映射数据名与专业词汇库中的全部字段进行相似度匹配后得到的, Q 为专业词汇数据库中与 p_name 关联字段的总数量, $\text{TF-IDF}(t(q))$ 为计算 p_name 与第 q 个关联字段之间相似度的 TF-IDF 函数, $\text{Num}(tn)$ 为 p_name 出现在专业词汇库内行业知识文本中的次数, v_eva 为 p_name 的预设估计价值。

[0091] 在本申请一些实施例中,构建单元205具体可以用于:

在待映射数据匹配值 Match 大于或等于预设匹配值的情况下,确定第一服务器与数据中台系统匹配成功,构建从第一服务器到数据中台系统的数据接口;

在待映射数据匹配值 Match 小于预设匹配值的情况下,确定待映射数据与数据中台系统匹配失败,输出匹配失败信息,以使用户手动匹配待映射数据的数据接口。

[0092] 在本申请一些实施例中,装置还可以包括:

更新度确定单元,用于在第一服务器为发起方服务器的情况下,如果匹配结果为匹配成功,则将每个目标标准数据的更新度加一,如果匹配结果为匹配失败,则将规范化处理后的待映射数据存储为新的标准数据,将新的标准数据的初始更新度设置为0;

匹配度确定单元,用于在第一服务器为接收方服务器的情况下,如果匹配结果为匹配成功,则将每个目标标准数据的匹配度加一,如果匹配结果为匹配失败,则将规范化处理后的待映射数据存储为新的标准数据,将新的标准数据的初始匹配度设置为0;

在从预设的标准数据中筛选得到目标标准数据之后,方法还包括:

求和单元,用于计算每个目标标准数据的更新度与匹配度之和,得到每个目标标

准数据的第一和值；

第一执行单元,用于在存在大于预设和值的目标第一和值的情况下,确定匹配结果为匹配成功,执行基于匹配结果构建从第一服务器到数据中台系统的数据接口的步骤；

第二执行单元,用于在不存在大于预设和值的目标第一和值的情况下,执行确定待映射数据的数据来源类型的步骤。

[0093] 在本申请一些实施例中,装置还可以包括：

停用单元,用于获取对数据接口中交互数据的人工干预修改次数,在人工干预修改次数大于预设次数的情况下,停用数据接口。

[0094] 在本申请一些实施例中,装置还可以包括：

获取单元,用于获取数据接口的正确对接数据的正确数据量,以及数据接口的数据传输总量；

停用单元,还用于在正确数据量在数据传输总量中的占比值小于或者等于预设占比值的情况下,停用数据接口；

重启单元,用于在停用数据接口预设时长之后,若数据接口的停用次数小于预设次数,重新启用数据接口。

[0095] 关于上述实施例中的装置,其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述,此处将不做详细阐述说明。

[0096] 根据本公开实施例提出的数据中台系统的接口数据自动映射装置,通过接收第一服务器发送的待映射数据,并对待映射数据进行规范化处理；利用预设相似度匹配算法将规范化处理后的待映射数据与数据中台系统中预设的标准数据进行初步匹配,从预设的标准数据中筛选得到多个目标标准数据；确定待映射数据的数据来源类型；按照数据来源类型对应的匹配策略,将规范化处理后的待映射数据与多个目标标准数据进行匹配,得到匹配结果；基于匹配结果构建从第一服务器到数据中台系统的数据接口,以使第一服务器与数据中台系统进行数据交互。智能识别数据来源于生产设备或业务系统,并据此分别采用融合统计特征与领域知识的策略或综合文本相似度的策略进行精准匹配,最终依据匹配结果自动化构建或装配数据接口,从而在无需人工一对一编码的情况下,有效解决了多源异构系统间因数据标准不一导致的数据对接难题,实现了数据的高效整合与接口的快速生成。

[0097] 图3是根据一示例性实施例示出的一种用于数据中台系统的接口数据自动映射方法的装置的框图。例如,装置300可以是电子设备,例如可以是移动电话,计算机,数字广播终端,消息收发设备,游戏控制台,平板设备,医疗设备,健身设备,个人数字助理等。

[0098] 参照图3,装置300可以包括以下一个或多个组件:处理组件302,存储器304,电力组件306,多媒体组件308,音频组件310,输入/输出(I/O)的接口312,传感器组件314,以及通信组件316。

[0099] 处理组件302通常控制装置300的整体操作,诸如与显示,电话呼叫,数据通信,相机操作和记录操作相关联的操作。处理组件302可以包括一个或多个处理器320来执行指令,以完成上述的方法的全部或部分步骤。此外,处理组件302可以包括一个或多个模块,便于处理组件302和其他组件之间的交互。例如,处理组件302可以包括多媒体模块,以方便多媒体组件308和处理组件302之间的交互。

[0100] 存储器304被配置为存储各种类型的数据以支持在设备300的操作。这些数据的示例包括用于在装置300上操作的任何应用程序或方法的指令,联系人数据,电话簿数据,消息,图片,视频等。存储器304可以由任何类型的易失性或非易失性存储设备或者它们的组合实现,如静态随机存取存储器(SRAM),电可擦除可编程只读存储器(EEPROM),可擦除可编程只读存储器(EPROM),可编程只读存储器(PROM),只读存储器(ROM),磁存储器,快闪存储器,磁盘或光盘。

[0101] 电力组件306为装置300的各种组件提供电力。电力组件306可以包括电源管理系统,一个或多个电源,及其他与为装置300生成、管理和分配电力相关联的组件。

[0102] 多媒体组件308包括在所述装置300和用户之间的提供一个输出接口的屏幕。在一些实施例中,屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板,屏幕可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中,多媒体组件308包括一个前置摄像头和/或后置摄像头。当设备300处于操作模式,如拍摄模式或视频模式时,前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

[0103] 音频组件310被配置为输出和/或输入音频信号。例如,音频组件310包括一个麦克风(MIC),当装置300处于操作模式,如呼叫模式、记录模式和语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器304或经由通信组件316发送。在一些实施例中,音频组件310还包括一个扬声器,用于输出音频信号。

[0104] I/O接口312为处理组件302和外围接口模块之间提供接口,上述外围接口模块可以是键盘,点击轮,按钮等。这些按钮可包括但不限于:主页按钮、音量按钮、启动按钮和锁定按钮。

[0105] 传感器组件314包括一个或多个传感器,用于为装置300提供各个方面的状态评估。例如,传感器组件314可以检测到设备300的打开/关闭状态,组件的相对定位,例如所述组件为装置300的显示器和小键盘,传感器组件314还可以检测装置300或装置300一个组件的位置改变,用户与装置300接触的存在或不存在,装置300方位或加速/减速和装置300的温度变化。传感器组件314可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件314还可以包括光传感器,如CMOS或CCD图像传感器,用于在成像应用中使用。在一些实施例中,该传感器组件314还可以包括加速度传感器,陀螺仪传感器,磁传感器,压力传感器或温度传感器。

[0106] 通信组件316被配置为便于装置300和其他设备之间有线或无线方式的通信。装置300可以接入基于通信标准的无线网络,如WiFi,2G或3G,或它们的组合。在一个示例性实施例中,通信组件316经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中,所述通信组件316还包括近场通信(NFC)模块,以促进短程通信。例如,在NFC模块可基于射频识别(RFID)技术,红外数据协会(IrDA)技术,超宽带(UWB)技术,蓝牙(BT)技术和其他技术来实现。

[0107] 在示例性实施例中,装置300可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列

(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,用于执行上述方法。

[0108] 在示例性实施例中,还提供了一种包括指令的非临时性计算机可读存储介质,例如包括指令的存储器304,上述指令可由装置300的处理器320执行以完成上述方法。例如,所述非临时性计算机可读存储介质可以是ROM、随机存取存储器(RAM)、CD-ROM、磁带、软盘和光数据存储设备等。

[0109] 在示例性实施例中,还提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被装置300的处理器320执行时实现上述方法。

[0110] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本发明的其它实施方案。本公开旨在涵盖本发明的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本发明的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本发明的真正范围和精神由下面的权利要求指出。

[0111] 应当理解的是,本发明并不局限于上面已经描述并在附图中示出的精确结构,并且可以在不脱离其范围进行各种修改和改变。本发明的范围仅由所附的权利要求来限制。

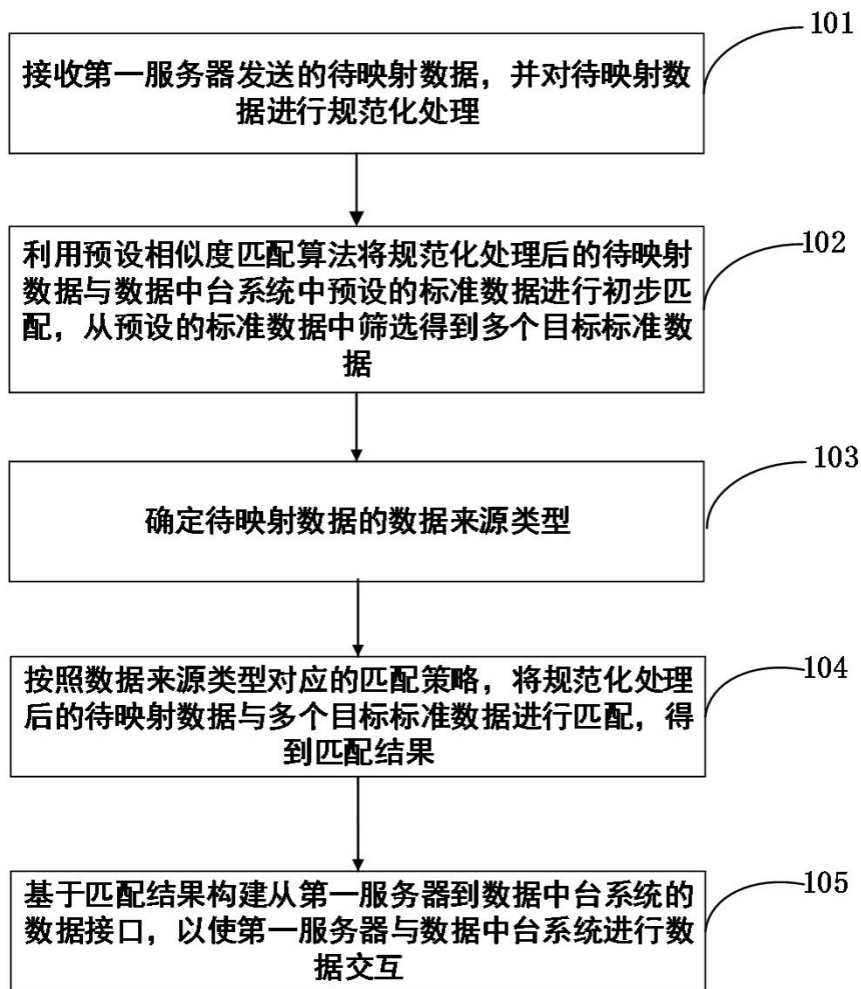


图1

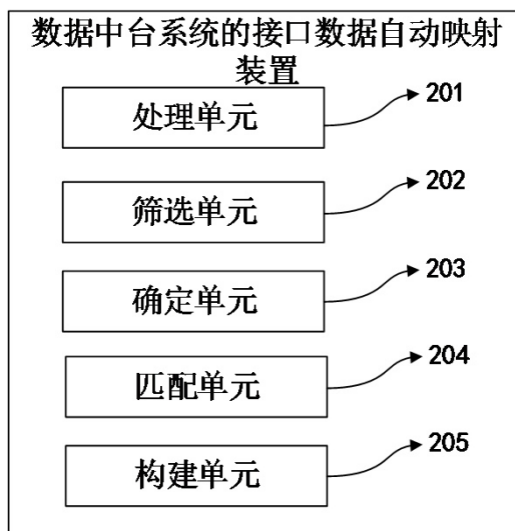


图2

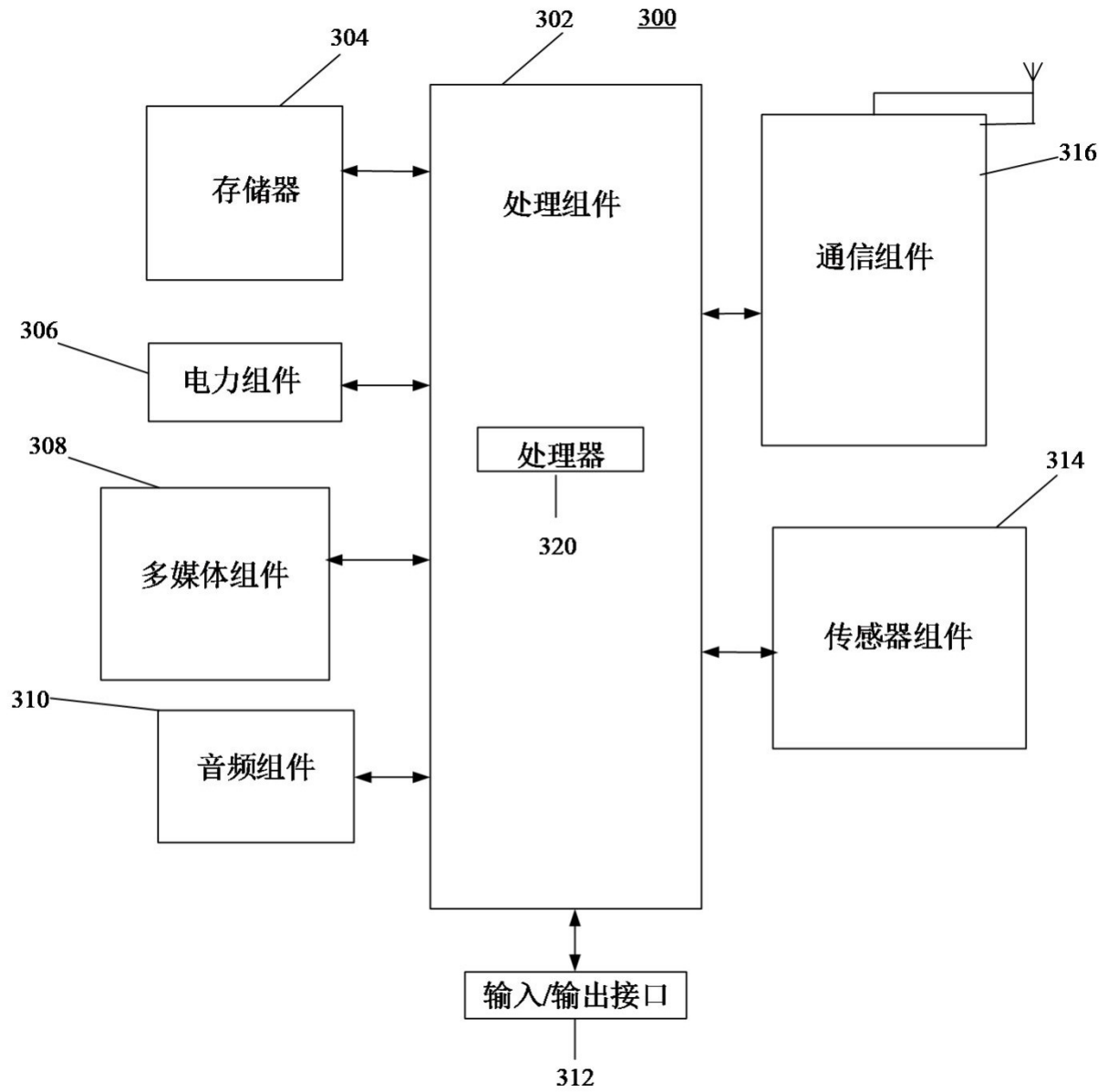


图3