

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号  
特許第5840110号  
(P5840110)

(45) 発行日 平成28年1月6日(2016.1.6)

(24) 登録日 平成27年11月20日(2015.11.20)

(51) Int.Cl.

F I

G O 6 F 12/00 (2006.01)

G O 6 F 17/30 (2006.01)

G O 6 F 12/00 5 1 3 D

G O 6 F 17/30 1 8 O D

請求項の数 11 (全 26 頁)

(21) 出願番号	特願2012-243539 (P2012-243539)	(73) 特許権者	000006013
(22) 出願日	平成24年11月5日 (2012.11.5)		三菱電機株式会社
(65) 公開番号	特開2014-92967 (P2014-92967A)		東京都千代田区丸の内二丁目7番3号
(43) 公開日	平成26年5月19日 (2014.5.19)	(74) 代理人	100099461
審査請求日	平成27年1月8日 (2015.1.8)		弁理士 溝井 章司
		(72) 発明者	桑折 朋晃
			東京都千代田区有楽町一丁目7番1号 三
			菱電機ビルテクノサービス株式会社内
		(72) 発明者	藤野 友也
			東京都千代田区丸の内二丁目7番3号 三
			菱電機株式会社内
		審査官	田中 幸雄

最終頁に続く

(54) 【発明の名称】 同一項目検出装置及びプログラム

(57) 【特許請求の範囲】

【請求項1】

複数のデータベースそれぞれに登録されている項目データの項目情報を抽出する項目情報抽出手段と、

前記項目情報抽出手段により抽出された項目情報の中から2つの項目情報を選択することにより、当該2つの項目情報から成る項目情報ペアを少なくとも1つ形成し、その形成した項目情報ペアの中から一致度の算出対象とする項目情報ペアを一致度算出ペアとして特定する一致度算出ペア特定手段と、

前記一致度算出ペア特定手段により特定された一致度算出ペアに含まれる各項目情報に対応する項目データを照合することによって当該項目情報に対応する項目データの一致度を算出する一致度算出手段と、

前記一致度算出手段により一致度が算出された項目情報ペアの中に、一方の項目情報を共通とし、かつ他方の項目情報の間で一致度が算出されていない項目情報ペアの組が存在する場合、前記他方の項目情報の間における一致度の取りうる範囲を算出する一致度範囲算出手段と、

前記一致度範囲算出手段、前記一致度算出ペア特定手段及び前記一致度算出手段における各処理を繰り返し実行することで、前記一致度範囲算出手段により算出された一致度の取りうる範囲の上限が予め設定された一致判定用閾値に達していない項目情報ペアを除き、前記項目情報抽出手段により抽出された項目情報の間における一致度を順次算出させる一致度算出処理制御手段と、

10

20

前記一致判定用閾値以上となる一致度が算出された項目情報ペアに含まれる各項目情報により特定される項目を同一の項目と判定する判定手段と、

を有することを特徴とする同一項目検出装置。

【請求項 2】

項目情報ペア毎に、当該項目情報ペアに含まれる各項目情報に対応する項目データの一致度及び当該一致度の取りうる範囲の各算出値を少なくとも含む項目情報ペア情報を保持する項目情報ペア情報テーブルを記憶する項目情報ペア情報記憶手段と、

前記項目情報抽出手段により抽出された項目情報の中から 2 つの項目情報を選択することにより形成される前記項目情報ペアそれぞれに設定される項目情報ペア情報を保持する前記項目情報ペア情報テーブルを生成する項目情報ペア情報生成手段と、

を有し、

前記一致度算出ペア特定手段及び前記一致度範囲算出手段は、前記項目情報ペア情報テーブルの設定内容を参照してそれぞれ処理を行うことを特徴とする請求項 1 に記載の同一項目検出装置。

【請求項 3】

管理者により一致していると認識されている項目の組から成る項目情報ペアが予め設定される対応関係情報テーブルを記憶する対応関係情報記憶手段を有し、

前記項目情報ペア情報生成手段は、前記対応関係情報テーブルに設定されている項目情報ペアにおける一致度として一致判定用閾値以上の値を前記項目情報ペア情報テーブルに設定することを特徴とする請求項 2 に記載の同一項目検出装置。

【請求項 4】

データベースを関連付けるためのキーとなる項目の項目情報の組から成るキー項目情報ペアが予め設定される対応キー情報テーブルを記憶する対応キー情報記憶手段を有し、

前記一致度算出手段は、前記対応キー情報テーブルに設定されたキー項目情報ペアに基づき前記データベースを関連付けして項目データの一致度を算出することを特徴とする請求項 2 に記載の同一項目検出装置。

【請求項 5】

前記項目情報ペア情報生成手段は、前記項目情報ペア情報テーブルに保持されている各項目情報ペア情報に、当該項目情報ペア情報に含まれている各項目情報が設定されたデータベースに対応する、前記対応キー情報テーブルに設定されているキー項目情報ペアを対応付けして設定することを特徴とする請求項 4 に記載の同一項目検出装置。

【請求項 6】

前記一致度範囲算出手段は、他の項目情報に対応する全ての項目データを用いずに一致度の取りうる範囲を算出することを特徴とする請求項 1 に記載の同一項目検出装置。

【請求項 7】

一致していない 1 組の項目データに含まれている各文字列の間に、予め設定された一致条件を満たす関係がある場合に、一方の項目データは誤入力された項目データと判定する誤入力判定手段を有することを特徴とする請求項 1 に記載の同一項目検出装置。

【請求項 8】

前記誤入力判定手段により誤入力と判定された項目データを自動的に修正する若しくは当該項目データに関する情報を出力する手段を有することを特徴とする請求項 7 に記載の同一項目検出装置。

【請求項 9】

管理者により一致していないと認識されている項目の組から成る項目情報ペアが予め設定される不一致関係情報テーブルを記憶する不一致関係情報記憶手段を有し、

前記項目情報ペア情報生成手段は、前記不一致関係情報テーブルに設定されている項目情報ペアにおける一致度として一致判定用閾値を下回る値を前記項目情報ペア情報テーブルに設定することを特徴とする請求項 2 に記載の同一項目検出装置。

【請求項 10】

前記項目情報ペア情報テーブルに設定された項目情報ペア情報に基づき項目情報の関係

10

20

30

40

50

を視認可能な項目情報対応関係グラフを生成する項目情報対応関係グラフ生成処理手段を有することを特徴とする請求項2に記載の同一項目検出装置。

【請求項11】

コンピュータを、

複数のデータベースそれぞれに登録されている項目データの項目情報を抽出する項目情報抽出手段、

前記項目情報抽出手段により抽出された項目情報の中から2つの項目情報を選択することにより、当該2つの項目情報から成る項目情報ペアを少なくとも1つ形成し、その形成した項目情報ペアの中から一致度の算出対象とする項目情報ペアを一致度算出ペアとして特定する一致度算出ペア特定手段、

10

前記一致度算出ペア特定手段により特定された一致度算出ペアに含まれる各項目情報に対応する項目データを照合することによって当該項目情報に対応する項目データの一致度を算出する一致度算出手段、

前記一致度算出手段により一致度が算出された項目情報ペアの中に、一方の項目情報を共通とし、かつ他方の項目情報の間で一致度が算出されていない項目情報ペアの組が存在する場合、前記他方の項目情報の間における一致度の取りうる範囲を算出する一致度範囲算出手段、

前記一致度範囲算出手段、前記一致度算出ペア特定手段及び前記一致度算出手段における各処理を繰り返し実行することで、前記一致度範囲算出手段により算出された一致度の取りうる範囲の上限が予め設定された一致判定用閾値に達していない項目情報ペアを除き、前記項目情報抽出手段により抽出された項目情報の間における一致度を順次算出させる一致度算出処理制御手段、

20

前記一致判定用閾値以上となる一致度が算出された項目情報ペアに含まれる各項目情報により特定される項目を同一の項目と判定する判定手段、

として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、複数のデータベースそれぞれに登録されている項目の中から同一である項目を検出する技術に関する。

30

【背景技術】

【0002】

従来から、データベース管理者や利用部署の違いなどから同一のシステムにおいて同類のデータベースを複数設けて運用する場合があった。このようなシステムにおいて、データベースを統合したり、各データベースに登録されている各項目を関連付けたりする場合、各データベースそれぞれに登録されている項目の中から同一である項目を検出して対応付ける必要がある。例えば、異なる複数のデータベース中の属性値同士の比較結果の一致度が高いものから属性項目同士を対応付ける技術が提案されている（例えば、特許文献1）。

【先行技術文献】

40

【特許文献】

【0003】

【特許文献1】特開2004-86782号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、従来においては、同じ属性項目同士を対応付ける際に、全てのデータベースに含まれている各項目の属性項目を総当たりで照合し、一致度の高い属性項目同士を対応付けるようにしていたので、データベース間の属性項目の対応付けに多大な処理負荷、時間を要していた。

50

## 【 0 0 0 5 】

本発明は、複数のデータベースそれぞれに設定されている項目の中から同一の項目を対応付ける処理の負荷を従来に比して軽減することを目的とする。

## 【課題を解決するための手段】

## 【 0 0 0 6 】

本発明に係る同一項目検出装置は、複数のデータベースそれぞれに登録されている項目データの項目情報を抽出する項目情報抽出手段と、前記項目情報抽出手段により抽出された項目情報の中から2つの項目情報を選択することにより、当該2つの項目情報から成る項目情報ペアを少なくとも1つ形成し、その形成した項目情報ペアの中から一致度の算出対象とする項目情報ペアを一致度算出ペアとして特定する一致度算出ペア特定手段と、前記一致度算出ペア特定手段により特定された一致度算出ペアに含まれる各項目情報に対応する項目データを照合することによって当該項目情報に対応する項目データの一致度を算出する一致度算出手段と、前記一致度算出手段により一致度が算出された項目情報ペアの中に、一方の項目情報を共通とし、かつ他方の項目情報の間で一致度が算出されていない項目情報ペアの組が存在する場合、前記他方の項目情報の間における一致度の取りうる範囲を算出する一致度範囲算出手段と、前記一致度範囲算出手段、前記一致度算出ペア特定手段及び前記一致度算出手段における各処理を繰り返し実行することで、前記一致度範囲算出手段により算出された一致度の取りうる範囲の上限が予め設定された一致判定用閾値に達していない項目情報ペアを除き、前記項目情報抽出手段により抽出された項目情報の間における一致度を順次算出させる一致度算出処理制御手段と、前記一致判定用閾値以上となる一致度が算出された項目情報ペアに含まれる各項目情報により特定される項目を同一の項目と判定する判定手段と、を有するものである。

## 【 0 0 0 7 】

また、項目情報ペア毎に、当該項目情報ペアに含まれる各項目情報に対応する項目データの一致度及び当該一致度の取りうる範囲の各算出値を少なくとも含む項目情報ペア情報を保持する項目情報ペア情報テーブルを記憶する項目情報ペア情報記憶手段と、前記項目情報抽出手段により抽出された項目情報の中から2つの項目情報を選択することにより形成される前記項目情報ペアそれぞれに設定される項目情報ペア情報を保持する前記項目情報ペア情報テーブルを生成する項目情報ペア情報生成手段と、を有し、前記一致度算出ペア特定手段及び前記一致度範囲算出手段は、前記項目情報ペア情報テーブルの設定内容を参照してそれぞれ処理を行うものである。

## 【 0 0 0 8 】

また、管理者により一致していると認識されている項目の組から成る項目情報ペアが予め設定される対応関係情報テーブルを記憶する対応関係情報記憶手段を有し、前記項目情報ペア情報生成手段は、前記対応関係情報テーブルに設定されている項目情報ペアにおける一致度として一致判定用閾値以上の値を前記項目情報ペア情報テーブルに設定するものである。

## 【 0 0 0 9 】

また、データベースを関連付けるためのキーとなる項目の項目情報の組から成るキー項目情報ペアが予め設定される対応キー情報テーブルを記憶する対応キー情報記憶手段を有し、前記一致度算出手段は、前記対応キー情報テーブルに設定されたキー項目情報ペアに基づき前記データベースを関連付けして項目データの一致度を算出するものである。

## 【 0 0 1 0 】

また、前記項目情報ペア情報生成手段は、前記項目情報ペア情報テーブルに保持されている各項目情報ペア情報に、当該項目情報ペア情報に含まれている各項目情報が設定されたデータベースに対応する、前記対応キー情報テーブルに設定されているキー項目情報ペアを対応付けして設定するものである。

## 【 0 0 1 1 】

また、前記一致度範囲算出手段は、前記他の項目情報に対応する全ての項目データを用いずに一致度の取りうる範囲を算出するものである。

## 【0012】

また、一致していない1組の項目データに含まれている各文字列の間に、予め設定された一致条件を満たす関係がある場合に、一方の項目データは誤入力された項目データと判定する誤入力判定手段を有するものである。

## 【0013】

また、前記誤入力判定手段により誤入力と判定された項目データを自動的に修正する若しくは当該項目データに関する情報を出力する手段を有するものである。

## 【0014】

また、管理者により一致していないと認識されている項目の組から成る項目情報ペアが予め設定される不一致関係情報テーブルを記憶する不一致関係情報記憶手段を有し、前記項目情報ペア情報生成手段は、前記不一致関係情報テーブルに設定されている項目情報ペアにおける一致度として一致判定用閾値を下回る値を前記項目情報ペア情報テーブルに設定するものである。

10

## 【0015】

また、前記項目情報ペア情報テーブルに設定された項目情報ペア情報に基づき項目情報の関係を視認可能な項目情報対応関係グラフを生成する項目情報対応関係グラフ生成処理手段を有するものである。

## 【0016】

本発明に係るプログラムは、コンピュータを、複数のデータベースそれぞれに登録されている項目データの項目情報を抽出する項目情報抽出手段、前記項目情報抽出手段により抽出された項目情報の中から2つの項目情報を選択することにより、当該2つの項目情報から成る項目情報ペアを少なくとも1つ形成し、その形成した項目情報ペアの中から一致度の算出対象とする項目情報ペアを一致度算出ペアとして特定する一致度算出ペア特定手段、前記一致度算出ペア特定手段により特定された一致度算出ペアに含まれる各項目情報に対応する項目データを照合することによって当該項目情報に対応する項目データの一致度を算出する一致度算出手段、前記一致度算出手段により一致度が算出された項目情報ペアの中に、一方の項目情報を共通とし、かつ他方の項目情報の間で一致度が算出されていない項目情報ペアの組が存在する場合、前記他方の項目情報の間における一致度の取りうる範囲を算出する一致度範囲算出手段、前記一致度範囲算出手段、前記一致度算出ペア特定手段及び前記一致度算出手段における各処理を繰り返し実行することで、前記一致度範囲算出手段により算出された一致度の取りうる範囲の上限が予め設定された一致判定用閾値に達していない項目情報ペアを除き、前記項目情報抽出手段により抽出された項目情報の間における一致度を順次算出させる一致度算出処理制御手段、前記一致判定用閾値以上となる一致度が算出された項目情報ペアに含まれる各項目情報により特定される項目を同一の項目と判定する判定手段、として機能させるためのものである。

20

30

## 【発明の効果】

## 【0017】

本発明によれば、複数のデータベースそれぞれに設定されている項目の中から同一の項目に対応付ける処理の負荷を従来に比して軽減することができる。

## 【0018】

また、管理者により一致していると認識されている項目の組から成る項目情報ペアにおける一致度として一致判定用閾値以上の値を計算することなく設定するようにしたので、同一の項目の検出に要する処理負荷を軽減することができる。

40

## 【0019】

また、全ての項目データを用いずに一致度の取りうる範囲を算出するようにしたので、同一の項目の検出に要する処理負荷を軽減することができる。

## 【0020】

また、同一性の判定対象となっている項目の項目データに含まれている誤入力された項目データを検出することができる。

## 【0021】

50

また、誤入力と判定された項目データを自動的に修正したり、当該項目データに関する情報を管理者に通知することができる。

【0022】

また、管理者により一致していないと認識されている項目の組から成る項目情報ペアにおける一致度として一致判定用閾値を下回る値を計算することなく設定するようにしたので、同一の項目の検出に要する処理負荷を軽減することができる。

【0023】

また、項目情報に対応する項目が同一かどうかの判定結果をグラフ形式で視認可能に知らせることができる。

【図面の簡単な説明】

10

【0024】

【図1】本発明に係る同一項目検出装置の一実施の形態を示したブロック構成図である。

【図2】本実施の形態における同一項目検出装置を形成するコンピュータのハードウェア構成図である。

【図3】本実施の形態における対応関係情報記憶部に設定登録された対応関係情報のデータ構成例を示した図である。

【図4】本実施の形態における対応キー情報記憶部に設定登録された対応キー情報のデータ構成例を示した図である。

【図5】本実施の形態における項目情報ペア情報記憶部に設定登録される項目情報ペア情報のデータ構成例を示した図である。

20

【図6】本実施の形態における同一項目検出処理を示したフローチャートである。

【図7】本実施の形態における項目情報対応関係グラフを模式的に示した図である。

【図8】本実施の形態において項目情報ペアの組の対応関係を示す概念図である。

【図9】本実施の形態における一致度算出対象特定部により一致度算出ペアを選出する第1の処理を説明するために用いる、項目情報の対応関係を示す概念図である。

【図10】本実施の形態における一致度算出対象特定部により一致度算出ペアを選出する第2の処理を説明するために用いる、項目情報の対応関係を示す概念図である。

【図11】本実施の形態における一致度算出対象特定部により一致度算出ペアを選出する第3の処理を説明するために用いる、項目情報の対応関係を示す概念図である。

【図12】本実施の形態における一致度算出処理を実施した結果、得られる項目情報対応関係グラフを模式的に示した図である。

30

【図13】本実施の形態において項目の対応付け対象とするデータベースの各データ設定例を示した図である。

【図14】本実施の形態の同一項目検出処理の実行の過程における項目情報ペア情報テーブルのデータ設定例を示した図である。

【図15】図14に示した項目情報ペア情報テーブルのデータ設定の内容に基づき生成された項目情報対応関係グラフを模式的に示した図である。

【図16】図14に示した設定内容から、一致度範囲を算出した後の項目情報ペア情報テーブルのデータ設定例を示した図である。

【図17】図16に示した設定内容から、一致度を算出した後の項目情報ペア情報テーブルのデータ設定例を示した図である。

40

【図18】図17に示した項目情報ペア情報テーブルのデータ設定の内容に基づき生成された項目情報対応関係グラフを模式的に示した図である。

【図19】本実施の形態の同一項目検出処理の実行の過程における項目情報ペア情報テーブルのデータ設定例を示した図である。

【図20】本実施の形態の同一項目検出処理における判定結果を提示するための項目情報対応関係グラフを模式的に示した図である。

【図21】本実施の形態の同一項目検出処理における判定結果を提示するレポートの一例を示した図である。

【発明を実施するための形態】

50

## 【 0 0 2 5 】

以下、図面に基づいて、本発明の好適な実施の形態について説明する。

## 【 0 0 2 6 】

図 1 は、本発明に係る同一項目検出装置の一実施の形態を示したブロック構成図である。本実施の形態における同一項目検出装置 1 0 は、ペア情報設定部 1 1、一致度範囲算出部 1 2、一致度算出ペア特定部 1 3、一致度算出部 1 4、同一項目判定部 1 5、判定結果出力部 1 6、一致度算出処理制御部 1 7、システムデータベース ( D B ) 1、項目情報ペア情報記憶部 2 1、対応関係情報記憶部 2 2、対応キー情報記憶部 2 3 及び一致判定用閾値記憶部 2 4 を有している。

## 【 0 0 2 7 】

10

図 2 は、本実施の形態における同一項目検出装置 1 0 を形成するコンピュータのハードウェア構成図である。本実施の形態における同一項目検出装置 1 0 は、従前から存在する汎用的なコンピュータ、例えばパーソナルコンピュータのハードウェア構成で実現できる。同一項目検出装置 1 0 を実現するコンピュータは、図 2 に示したように C P U 3 1、R O M 3 2、R A M 3 3、ハードディスクドライブ ( H D D ) 3 4 を接続した H D D コントローラ 3 5、入力手段として設けられたマウス 3 6 とキーボード 3 7、及び表示装置として設けられたディスプレイ 3 8 をそれぞれ接続する入出力コントローラ 3 9、通信手段として設けられたネットワークコントローラ 4 0 を内部バス 4 1 に接続して構成される。

## 【 0 0 2 8 】

ここで、本実施の形態において用いる用語について定義しておく。

20

## 【 0 0 2 9 】

まず、「項目」又は「データ項目」というのは、データベースに登録される各レコードを構成するデータの項目 ( アイテム ) のことである。例えば、図 1 3 におけるデータベース 1 a には、建物 I D、建物所属、担当者 I D 及び機種 の 4 項目が設定される。

## 【 0 0 3 0 】

「項目データ」というのは、各項目に設定されるデータ ( 値 ) である。例えば、図 1 3 におけるデータベース 1 a の建物 I D には、“ A A 0 0 0 1 ”、“ A A 0 0 0 2 ” などという項目データが設定される。

## 【 0 0 3 1 】

「項目情報」というのは、各項目を特定するための情報である。本実施の形態では、項目の名称を項目情報として用いる。例えば、図 1 3 におけるデータベース 1 a には、「建物 I D」、「建物所属」、「担当者 I D」及び「機種」が項目名として設定されている。各項目に設定登録される項目データは、異なるデータベース間で同じでも、項目を識別するために付けられた項目名は、例えば、「ユーザ I D」と「ログイン I D」や、「社員証番号」と「社員 I D」などのようにデータベースによって異なっている場合があり得る。

30

## 【 0 0 3 2 】

「項目情報ペア」というのは、2 つの項目情報の組のことをいう。図 1 3 においては、例えば、データベース 1 a の「建物所属」と、データベース 1 b の「契約年数」と、いう 2 つの項目情報によって項目情報ペアは形成される。

## 【 0 0 3 3 】

40

「一致度算出ペア」というのは、一致度の算出対象として選出された項目情報ペアのことをいう。本実施の形態における項目情報ペアは、一致度が一度算出されると、二度と一致度の算出対象として取り扱われない。換言すると、一致度が一度算出された項目情報ペアは、一致度算出ペア特定部 1 3 における一致度の算出対象候補とはならない。一致度算出ペア特定部 1 3 により一致度の算出対象として選出されることで、その選出された項目情報ペアは、一致度算出ペアとなるが、一致度算出部 1 4 によって一致度が算出された後は、一致度算出ペアと呼ばれることはない。

## 【 0 0 3 4 】

図 1 に戻り、同一項目検出装置 1 0 に含まれる各構成要素について説明する。

## 【 0 0 3 5 】

50

ペア情報設定部 1 1 は、項目情報抽出手段として設けられ、複数のデータベース 1 それぞれに登録されている項目データの項目情報を抽出する。ペア情報設定部 1 1 はまた、項目情報ペア情報生成手段としても設けられており、後述する項目情報ペア情報テーブルを生成する。

【 0 0 3 6 】

一致度範囲算出部 1 2 は、一致度範囲算出手段として設けられ、一致度算出部 1 4 により一致度が算出された項目情報ペアの中に、一方の項目情報を共通とし、かつ他方の項目情報の間で一致度が算出されていない項目情報ペアの組が存在する場合、前記他方の項目情報の間における一致度の取りうる範囲を算出する。

【 0 0 3 7 】

一致度算出ペア特定部 1 3 は、一致度算出ペア特定手段として設けられ、ペア情報設定部 1 1 により抽出された項目情報の中から 2 つの項目情報を選択することにより、当該 2 つの項目情報から成る項目情報ペアを少なくとも 1 つ形成し、その形成した項目情報ペアの中から一致度の算出対象とする項目情報ペアを一致度算出ペアとして特定する。

【 0 0 3 8 】

一致度算出部 1 4 は、一致度算出手段として設けられ、一致度算出ペア特定部 1 3 により特定された一致度算出ペアに含まれる各項目情報に対応する項目データを照合することによって当該項目情報に対応する項目データの一致度を算出する。

【 0 0 3 9 】

同一項目判定部 1 5 は、判定手段として設けられ、一致判定用閾値以上となる一致度が算出された項目情報ペアに含まれる各項目情報により特定される項目を同一の項目と判定する。

【 0 0 4 0 】

判定結果出力部 1 6 は、出力手段として設けられ、同一項目判定部 1 5 における判定結果をディスプレイ 3 8 やファイル、あるいは図示しないプリンタ等に出力する。なお、判定結果出力部 1 6 は、同一項目検出装置 1 0 の外部に設けられてもよい。

【 0 0 4 1 】

一致度算出処理制御部 1 7 は、一致度算出処理制御手段として設けられ、一致度範囲算出部 1 2、一致度算出ペア特定部 1 3 及び一致度算出部 1 4 における各処理を繰り返し実行することで、一致度範囲算出部 1 2 により算出された一致度の取りうる範囲の上限が予め設定された一致判定用閾値に達していない項目情報ペアを除き、ペア情報設定部 1 1 により抽出された項目情報の間（すなわち項目情報ペア）における一致度を順次算出させる。一致度算出処理制御部 1 7 は、その他にも同一項目検出装置 1 0 における各構成要素 1 1 ~ 1 6 の全体制御を行う。

【 0 0 4 2 】

データベース（DB）1 は、本実施の形態において項目の対応付け対象とするデータベースである。データベース 1 を構成する各システムのデータベース 1 a, 1 b, 1 c は、同一のシステム内に構築されたデータベースであるものの、データベース管理者やシステム管理者（以下、「管理者」と総称する）の違いなどから別個に構築されたデータベースである。もちろん、全てのデータベース 1 a, 1 b, 1 c は、同一のシステム内に構築されている必要はなく、データベースに登録されている各項目を関連付ける対象として選出されていればよい。本実施の形態では、3 つのデータベース 1 a, 1 b, 1 c を例にして説明するが、このデータベースの数は複数であればよい。なお、上記の通り、各データベース 1 a, 1 b, 1 c をまとめて説明する場合には、「データベース 1」と総称する。

【 0 0 4 3 】

図 3 は、本実施の形態における対応関係情報記憶部 2 2 に設定登録された対応関係情報のデータ構成例を示した図である。対応関係情報記憶部 2 2 は、対応関係情報テーブルを記憶する対応関係情報記憶手段として設けられている。対応関係情報は、図 3 に示した対応関係情報テーブルにて管理される。対応関係情報は、管理者により一致していると認識されている項目の組を少なくとも含む。一致していると言うことは、当該項目情報ペアの

10

20

30

40

50



一致度が一致判定用閾値以上であるということである。

【 0 0 4 4 】

対応関係情報テーブルには、図 3 に例示したように、項目の組を構成する各項目を特定する情報として項目 1 及び項目 2 が設定される。項目 1 には、対応付ける一方の項目に関する情報が設定登録される。具体的には、対応付ける一方のデータベースを特定するデータベース (DB) 1 と当該データベースにおける項目情報 (項目名) 1 と、対応付ける際のキーとなる項目情報 (項目名) を設定するキー 1 と、が対応付けして設定される。項目 2 には、対応付ける他方の項目に関する情報が項目 1 と同様に設定登録される。図 3 には、データベース A における“担当者 ID”とデータベース B における“担当者 ID”との項目の組が管理者により初期設定された例が示されている。ただ、対応関係情報テーブルには、初期設定とする項目情報ペアがされていなくてもよいし、複数の項目情報ペアが初期設定されていてもよい。

10

【 0 0 4 5 】

なお、この例の“担当者 ID”のように、対応付ける項目名は、通常、同じ項目名となっている場合が多いと推測されるが、異なる項目名が設定される可能性もある。

【 0 0 4 6 】

また、本実施の形態では、後述するように、一致度等を一致度算出部 14 に算出させるので対応関係情報に含めていないが、既知である場合には当該項目情報ペアに対応させて予め設定するようにしてもよい。

【 0 0 4 7 】

20

本実施の形態においては、異なるデータベースそれぞれに設定されている 2 つの項目情報 (項目名) が仮に異なっているととしても、同じ項目データが設定されているならば、それら対応する 2 つの項目は同一であると判定し、その判定結果を管理者に提示可能とするが、対応関係情報テーブルには、管理者が同一であると予め認識している項目の組が事前に対応付けされ設定登録される。

【 0 0 4 8 】

図 4 は、本実施の形態における対応キー情報記憶部 23 に設定登録された対応キー情報のデータ構成例を示した図である。対応キー情報記憶部 23 は、データベースを関連付けるためのキーとなる項目の項目情報の組から成る項目情報ペアが予め設定される対応キー情報テーブルを記憶する対応キー情報記憶手段として設けられている。対応キー情報は、図 4 に示したように対応キー情報テーブルにて管理される。本実施の形態においては、異なるデータベースそれぞれに設定されている項目を対応付けるものであるが、対応キー情報には、異なるデータベースの項目を対応付ける際のキーとなる項目の項目情報の組 (キー項目情報ペア) が設定登録される。

30

【 0 0 4 9 】

対応キー情報には、図 4 に例示したように、項目の組を構成する各項目を特定する情報として項目 1 及び項目 2 が設定される。項目 1 には、対応付ける一方の項目に関する情報が設定登録される。具体的には、対応付ける一方のデータベースを特定するデータベース (DB) 1 と、キーとなる項目情報 (項目名) を設定するキー 1 と、が対応付けして設定される。項目 2 には、対応付ける他方の項目に関する情報が項目 1 と同様に設定登録される。

40

【 0 0 5 0 】

対応キー情報テーブルには、各データベースを対応付ける少なくとも 1 組の項目情報が設定されている必要がある。図 4 に示した例によると、データベース A 及びデータベース B は、同じ項目名“建物 ID”の項目によって対応付けられている。また、データベース A 及びデータベース C は、同じ項目名“建物 ID”の項目によって対応付けられている。これにより、データベース C 及びデータベース B も、項目名“建物 ID”の項目によって対応付けられていると言える。

【 0 0 5 1 】

一致判定用閾値記憶部 24 には、算出された一致度と比較される一致判定用閾値が記憶

50

される。

【 0 0 5 2 】

前述した対応関係情報記憶部 2 2、対応キー情報記憶部 2 3 及び一致判定用閾値記憶部 2 4 が、本実施の形態を実施する前に各テーブル内に情報が予め設定されているのに対し、項目情報ペア情報記憶部 2 1 に記憶される項目情報ペア情報テーブルは、本実施の形態の実施開始後に生成され、各種データは、一致度算出処理を繰り返し実行する過程において逐次設定登録される。

【 0 0 5 3 】

図 5 は、本実施の形態における項目情報ペア情報記憶部 2 1 に設定登録される項目情報ペア情報のデータ構成例を示した図である。項目情報ペア情報記憶部 2 1 は、項目情報ペア情報テーブルを記憶する項目情報ペア情報記憶手段として設けられている。項目情報ペア情報は、図 5 に例示した項目情報ペア情報テーブルにて管理される。項目情報ペア情報は、項目情報ペア毎に、項目情報 1、項目情報 2、キー、一致度、全件数、一致件数、欠損値数、一致度上限及び一致度下限の各データを保存する領域が事前に用意され、項目情報ペア情報に含まれるデータの生成に伴い、当該データが所定の保存領域に書き込まれる。「項目情報 1」には、項目情報ペアを形成する一方の項目情報を特定する情報として、データベースと項目情報（項目名）が設定される。「項目情報 2」には、項目情報ペアを形成する他方の項目情報を特定する情報として、データベースと項目情報（項目名）が設定される。「対応キー情報」には、各データベースの各項目情報を対応付ける際に用いるキーとして、各データベース及び当該各データベースにおける項目情報が設定される。「一致度」には、項目情報 1 と項目情報 2 それぞれに対応する項目データが一致する割合が一致度として設定される。「全件数」には、当該項目情報に設定されている項目データの全件数が設定される。「一致件数」には、項目情報 1 に設定されている項目データと、項目情報 2 に設定されている項目データとの双方に設定されている件数が設定される。「欠損値数」には、一方の項目情報のみに設定されている項目データの件数が設定される。一致度、全件数、一致件数及び欠損値数は、一致度算出部 1 4 により算出される。また、本実施の形態では、一致度範囲算出部 1 2 により一致度の取りうる範囲が算出されるが、「一致度上限」にはその範囲の上限値が、「一致度下限」にはその範囲の下限値が、それぞれ設定される。

【 0 0 5 4 】

同一項目検出装置 1 0 における各構成要素 1 1 ~ 1 7 は、同一項目検出装置 1 0 を形成するコンピュータと、コンピュータに搭載された CPU 3 1 で動作するプログラムとの協調動作により実現される。また、各記憶部 2 1 ~ 2 4 は、同一項目検出装置 1 0 に搭載された HDD 3 4 又は RAM 3 3 で実現される。あるいは、外部にある記憶手段をネットワーク経由で利用してもよい。

【 0 0 5 5 】

また、本実施の形態で用いるプログラムは、通信手段により提供することはもちろん、CD-ROM や DVD-ROM 等のコンピュータ読み取り可能な記録媒体に格納して提供することも可能である。通信手段や記録媒体から提供されたプログラムはコンピュータにインストールされ、コンピュータの CPU 3 1 がプログラムを順次実行することで各種処理が実現される。

【 0 0 5 6 】

次に、本実施の形態における動作について説明するが、同一項目検出装置 1 0 に処理を実施させる前に、管理者は、前述したように対応関係情報、対応キー情報及び一致判定用閾値を各記憶部 2 2 ~ 2 4 に予め設定しておく。以下、本実施の形態における同一項目の検出処理を図 6 に示したフローチャートを用いて説明する。

【 0 0 5 7 】

管理者により本処理のプログラムが起動されると、一致度算出処理制御部 1 7 は、まずペア情報設定部 1 1 を起動する。ペア情報設定部 1 1 は、一致度算出処理制御部 1 7 により起動されると、各データベース 1 a , 1 b , 1 c からそれぞれに設定されているデータ

10

20

30

40

50

項目の項目名を抽出する（ステップ１０１）。そして、全ての項目名を組み合わせることで項目情報ペアを形成し、項目情報ペア毎に、前述した項目情報ペア情報を保存するための項目情報ペア情報テーブルを生成する（ステップ１０２）。例えば、図１３に示した例によると、データベース１ａから“建物ＩＤ”、“建物所属”、“担当者ＩＤ”及び“機種”を、データベース１ｂから“建物ＩＤ”、“契約種別”、“契約年数”及び“契約条件”を、データベース１ｃから“建物ＩＤ”、“担当者ＩＤ”、“担当者所属”、“作業種類”及び“作業条件”を、それぞれ抽出し、各項目情報を組み合わせて項目情報ペアを形成し、項目情報ペア情報テーブルに設定登録する。具体的には、ペア情報設定部１１は、各項目情報ペアを図５における項目情報１と項目情報２に登録する。なお、全ての項目名の組み合わせとは言っても、１つのデータベースの中に同じデータ項目が存在することは通常はないので、同じデータベースに含まれる項目情報で項目情報ペアを形成する必要はないかもしれない。ただ、論理的にはあり得るし、また情報として提供するためにも、本実施の形態においては、同じデータベースに含まれる項目情報で項目情報ペアを形成するものとする。もちろん、形成しないようにしてもよい。

10

#### 【００５８】

なお、本実施の形態では、処理対象のデータベース１の複製を同一項目検出装置１０の内部に事前に用意するようにしたが、複製を作成せずに外部にあるデータベース１に直接アクセスするようにしてもよい。

#### 【００５９】

続いて、ペア情報設定部１１は、対応キー情報記憶部２３から対応キー情報として設定されている項目１，２を取り出し、項目情報ペア情報テーブルにおける「対応キー情報」に設定登録する（ステップ１０３）。なお、本実施の形態では、項目情報ペア情報テーブルに対応キー情報を設定登録するようにしたが、設定登録せずに、必要なときにその都度対応キー情報テーブルを参照するようにしてもよい。

20

#### 【００６０】

更に、ペア情報設定部１１は、対応関係情報記憶部２２から対応関係情報を取り出し、この対応関係情報に対応する項目情報ペアに対し、一致度、全件数、一致件数及び欠損損数を算出し、項目情報ペア情報テーブルに設定登録する（ステップ１０４）。この一致度等の算出に関しては、一致度算出部１４に処理を依頼してもよい。一致度の算出の具体的な処理内容については、一致度算出部１４における一致度算出処理の説明と合わせて説明する。

30

#### 【００６１】

ここでの処理のように、一致度の取りうる範囲より先に一致度が算出されると、ペア情報設定部１１は、一致度上限及び一致度下限に、算出された一致度を設定する。なお、対応関係情報テーブルに設定されている項目情報ペアの一致度は、一致判定用閾値以上であることは明らかなので、ペア情報設定部１１は、一致度算出部１４に一致度を算出させることなく、一致判定用閾値以上の値（例えば１．００）を一致度、一致度上限及び一致度下限に設定するようにしてもよい。

#### 【００６２】

以上のように、ステップ１０１～１０４における処理を実行した結果、得られる項目情報ペア情報の例が図５に示されている。なお、ステップ１０３とステップ１０４とは、どちらを先に実行してもよい。

40

#### 【００６３】

ところで、項目情報対応関係グラフ生成処理手段（図示せず）を設けることによって、項目情報ペア情報テーブルに設定された項目情報ペア情報に基づき項目情報対応関係グラフを生成するようにしてもよい。図５に例示した内容の項目情報ペア情報に基づき生成された項目情報対応関係グラフを模式的に示した図を図７に示す。図７において、項目情報グループ１０１ａにはデータベース１ａから抽出した項目情報が含まれる。同様に、項目情報グループ１０１ｂにはデータベース１ｂから抽出した項目情報が、項目情報グループ１０１ｃにはデータベース１ｃから抽出した項目情報が、それぞれ含まれる。そして、対

50

応キー情報に基づきキー項目の対応関係を線 102 で示している。更に、一致度がすでに算出されている項目情報ペアが線 103 で接続されている。

#### 【0064】

以上のようにして、ペア情報設定部 11 により現時点で判明している項目情報ペア情報が項目情報ペア情報テーブルに設定されると、一致度算出処理制御部 17 は、一致度範囲算出部 12 を起動する。

#### 【0065】

一致度範囲算出部 12 は、起動されると、現時点において設定されている項目情報ペア情報を参照し、図 8 に例示したような関係を持つ項目情報ペアの組を探索する。すなわち、一致度範囲算出部 12 は、一致度算出部 14 により一致度が算出された項目情報ペアの中に、一方の項目情報（図 8 における「項目情報 A」）を共通とし、かつ他方の項目情報（図 8 における「項目情報 B」及び「項目情報 C」）の間で一致度が現時点で算出されていない項目情報ペア（図 8 における「ペア P1」及び「ペア P2」）の組が存在する場合、その項目情報ペアの組の全てに対し、他方の項目情報の間における一致度の取りうる範囲を算出する。つまり、図 8 に示した例では、項目情報ペア P1 は一致度が、項目情報ペア P2 は一致度が、それぞれ算出されており、項目情報ペア P1 と項目情報ペア P2 は、それぞれに共通する項目情報（項目情報 A）を含んでおり、各項目情報ペアの他方の項目情報（「項目情報 B」及び「項目情報 C」）における一致度は算出されていないことが確認できた場合、項目情報 B と項目情報 C との間における一致度の取りうる範囲を算出する。一致度の取りうる範囲は、以下の式を用いて計算する。

$$\max \{ \quad + \quad - 1, 0 \} \quad | \quad - \quad | \quad \cdots (1)$$

#### 【0066】

但し、式（1）において、 $\alpha$  は項目情報ペア P1 の一致度、 $\beta$  は項目情報ペア P2 の一致度、 $\gamma$  は項目情報ペア P1 及び項目情報ペア P2 に含まれる項目情報の組であって算出されていない一致度、 $\max$  関数は、 $(\alpha + \beta - 1)$  と 0 とのうち大きい値とする。本実施の形態では、このように既知の項目情報ペアにおける一致度を用い、当該項目情報ペアを構成する項目の項目データを用いない方法にて一致度の取りうる範囲を算出する。

#### 【0067】

以上のようにして一致度の取りうる範囲を算出すると、一致度範囲算出部 12 は、その算出結果を、項目情報ペア情報における当該項目情報ペア（上記例では、項目情報 B と項目情報 C との組）に対応する一致度上限及び一致度下限に設定登録する（ステップ 105）。

#### 【0068】

なお、一致度範囲算出部 12 における最初の処理実行時点では、対応関係情報テーブルに設定されている項目情報ペアの間の一一致度しか求められていない。本実施の形態の場合、図 3 に例示したように、データベース 1a とデータベース 1b の各担当者 ID の組による項目情報ペア P1 のみが存在する状況であるため、図 8 に例示した関係が 1 組も形成されない。このような場合には、一致度範囲算出部 12 において一致度の取りうる範囲は算出されないことになる。

#### 【0069】

一致度範囲算出部 12 における処理が終了すると、一致度算出処理制御部 17 は、一致度算出ペア特定部 13 を起動する。一致度算出ペア特定部 13 は、起動されると、項目情報ペア情報テーブルを参照し、次の手順にて項目情報ペアの中から一致度の算出対象とする一致度算出ペアを特定する（ステップ 106）。本実施の形態では、一致度の算出対象候補となる項目情報ペアがなくなるまで、常に 1 つの項目情報ペアを一致度算出ペアとして選出することになっている。本実施の形態における一致度算出ペア特定部 13 は、4 つの処理手順を設定して、各処理手順に設定された優先順位（本実施の形態では第 1 の処理から第 4 の処理まで順番）に従って各処理を順次実施するようにしている。

#### 【0070】

まず、第1の処理について図9を用いて説明する。図9において英字はそれぞれ項目情報名とする。図9(a)では、項目情報Aと各項目情報B, C, Dとの間でのみ一致度が算出されているものとする。図9(b)においては項目情報Fと各項目情報G, Hとの間でのみ一致度が算出されているものとする。

【0071】

ここで、項目情報Aと項目情報Eとから成る項目情報ペアを一致度算出ペアとして選出したと仮定する。つまり、ステップ107において、項目情報Aと項目情報Eとの間の一致度を算出したと仮定する。そうすると、図9(a)の関係を図8に当てはめると、次に実施されるステップ105においては、項目情報Eと項目情報B, C, Dとの間で一致度の取りうる範囲が計算可能になる。つまり、新規に3カ所の一致度の取りうる範囲(以下、「一致度範囲」ともいう)を算出対象とすることができる。

10

【0072】

一方、図9(b)に例示したように、項目情報Fと項目情報Iとから成る項目情報ペアを一致度算出ペアとして選出したと仮定する。つまり、ステップ107において、項目情報Fと項目情報Iとの間の一致度を算出したと仮定する。そうすると、図9(b)の関係を図8に当てはめると、次に実施されるステップ105においては、項目情報Iと項目情報G, Hとの間で一致度の取りうる範囲が計算可能になる。つまり、新規に2カ所の一致度範囲を算出対象とすることができる。

【0073】

第1の処理では、次に実施されるステップ105において一致度範囲の算出対象となる項目情報ペアの数を求め、その項目情報ペアの数が最大となる項目情報ペアを一致度算出ペアとして選出する。図9に示した例では、項目情報Aと項目情報Eとから成る項目情報ペアは、項目情報Fと項目情報Iとから成る項目情報ペアより多く、最大の項目情報ペアを一致度範囲の算出対象とすることができるので、一致度算出ペアとして選出されることになる。

20

【0074】

第1の処理では、一致度の取りうる範囲が計算もされていない項目情報ペアの数を減らすことを目的としている。後述するように、一致度の上限が一致判定用閾値に達していないということは、当該項目情報ペアに含まれる1組の項目情報は、同じ項目でないと判定されることになる。ステップ107においては、一致度を1つの項目情報ペアずつ算出していくので、1つの項目情報ペアのみが、項目が一致しているか否かの判定が可能になる。これに対し、ステップ105においては、図8に示すような関係に該当する項目情報ペア全てに対し一致度範囲を算出している。つまり、ステップ105における一度の処理で複数の一致度範囲が算出可能であり、その算出により得られた一致度上限が一致判定用閾値に達していなければ、その達していない項目情報ペアに含まれる項目は同一でないと判定できる。つまり、ステップ107では常に1組の項目情報のみを同一か否かを判定しているのに対し、ステップ105では複数組の項目情報を同一か否かを判定することが可能になる。処理効率化の観点から第1の処理を一致度算出ペアを選出する最初の処理手順としている。

30

【0075】

この第1の処理でただ1つの項目情報ペアが一致度算出ペアとして選出できれば、以降の第2乃至第4の処理は実行せずに終了する。但し、次に実施されるステップ105において一致度範囲の算出対象となる項目情報ペアの数が最大となる項目情報ペアが複数存在する場合、ただ1つの項目情報ペアを特定できないので、この場合は第2の処理に移行する。以下、第2の処理について図10を用いて説明する。

40

【0076】

第2の処理の対象となるのは、第1の処理において選出された複数の項目情報ペアである。そして、この中で一致度の最も高い項目情報ペアを特定し、その項目情報ペアに含まれる一方の項目情報との間でまだ一致度が算出されていない項目情報とから成る項目情報ペアを一致度算出ペアとして選出する。図10に示した例によると、一致度が一致度

50

より大きく最大と判定されたので、図 10 ( a ) に示した項目情報ペアに含まれる項目情報 K と、項目情報 K との間で一致度が算出されていない項目情報 L とから成る項目情報ペアを一致度算出ペアとして選出する。図 10 ( a ) を図 9 ( a ) に当てはめると、項目情報 J は項目情報 B , C , D に対応し、項目情報 K が項目情報 A に対応し、項目情報 L が項目情報 E に対応する。なお、図 9 ( a ) を参照すると、項目情報 A と項目情報 E から成る項目情報ペアが一致度算出ペアとして常に選出されるように見える。ただ、図 9 ( a ) に例示したように、新規 3 カ所の一致度範囲が算出可能な項目情報ペアは複数存在するので、複数の項目情報ペアが一致度の算出候補として存在している。

【 0 0 7 7 】

第 2 の処理では、項目情報 J と項目情報 K との間における一致度が高いので、項目情報 K と項目情報ペアを成す項目情報は、他と比較して一致度が高い可能性があるとも考えられる。従って、第 2 の処理では、一致度が高いであろう項目情報ペアを優先して一致度を算出するようにしている。

【 0 0 7 8 】

この第 2 の処理でただ 1 つの項目情報ペアが一致度算出ペアとして選出できれば、以降の第 3 及び第 4 の各処理は実行せずに終了する。ここで、一致度 が最大であるのにもかかわらず、一致度 の項目情報ペアが複数存在することにより、ただ 1 つの項目情報ペアを特定できない場合には、第 3 の処理に移行する。以下、第 3 の処理について図 1 1 を用いて説明する。

【 0 0 7 9 】

第 3 の処理対象となるのは、第 2 の処理において選出された複数の項目情報ペアである。そして、選出された複数の項目情報ペアの中で、一致度範囲が算出されており、その算出された一致度範囲の幅が最大となる項目情報ペアを一致度算出ペアとして選出する。つまり、図 1 1 ( a ) に示したように、項目情報ペア P 3 における一致度上限  $\max 1$  と一致度下限  $\min 1$  の差分である一致度範囲値 (  $\max 1 - \min 1$  ) を求める。図 1 1 ( b ) においても同様に一致度範囲値 (  $\max 2 - \min 2$  ) を求める。そして、一致度範囲値を求めた項目情報ペアの中で一致度範囲値が最大値となる項目情報ペアを一致度算出ペアとして選出する。仮に、複数の項目情報ペアが該当する場合は、いずれか任意の一方を選出すればよい。どちらを選ぼうと同じだからである。

【 0 0 8 0 】

第 3 の処理は、一致度範囲が広すぎるため一致度を確定させたい場合に有効である。特に、一致度上限が一致判定用閾値を超え、一致度下限が一致判定用閾値を下回る場合は一致度を明確にしたいので、このような場合を選出条件としてもよい。なお、一致度上限が一致判定用閾値を下回っている項目情報ペアは、一致度算出ペアの選出対象とする必要はない。

【 0 0 8 1 】

この第 3 の処理でただ 1 つの項目情報ペアが一致度算出ペアとして選出できれば、第 4 の各処理は実行せずに終了する。該当する項目情報ペアが複数存在することにより、ただ 1 つの項目情報ペアが選出できない場合には、第 4 の処理に移行する。

【 0 0 8 2 】

第 4 の処理は、第 3 の処理において選出された複数の項目情報ペアの中から任意の 1 つの項目情報ペアを選出する。あるいは、第 3 の処理を通過していない項目情報ペアを含めた全体の中から選出してもよい。

【 0 0 8 3 】

一致度算出ペア特定部 1 3 は、前述した処理手順に従ってただ 1 つの一致度算出ペアを特定する。この一致度算出ペアを特定する処理においては、一致度 の上限が一致判定用閾値に達していない項目情報ペアは一致度の算出対象候補から外すようにしている。これは、第 1 の処理においても説明したように、一致度 の上限が一致判定用閾値に達していないということは、当該項目情報ペアに含まれる 1 組の項目情報は、同じ項目でないことは明らかだからである。従って、一致度を算出する必要はない。本実施の形態では、この

10

20

30

40

50

ように一致度 の上限が一致判定用閾値に達していない項目情報ペアに対しては一致度を算出しないようにしたので、項目の同一性を判断する処理を効率化できる。

#### 【 0 0 8 4 】

一致度算出ペア特定部 1 3 により一致度算出ペアが特定されると、一致度算出処理制御部 1 7 は、一致度算出部 1 4 を起動する。一致度算出部 1 4 は、起動されると、次の一致度算出処理を実行する（ステップ 1 0 7）。

#### 【 0 0 8 5 】

一致度算出部 1 4 は、特定された一致度算出ペアにおける対応キー情報を項目情報ペア情報テーブルから取り出す。また、一致度算出部 1 4 は、特定された一致度算出ペアに含まれる項目情報に対応する項目データを該当するデータベース 1 から取り出す。そして、  
10  
対応キー情報に含まれている項目情報をキーとして各項目データを結合する。結合すると、結合したテーブルにおいて、全件数、一致件数及び欠損値数を求め、以下の式（ 2 ）に代入することで一致度を算出する。

$$\text{一致度} = \text{一致件数} \div (\text{全件数} - \text{欠損値数}) \quad \cdots (2)$$

#### 【 0 0 8 6 】

一致度算出部 1 4 は、算出した一致度並びに算出の過程において算出した全件数、一致件数及び欠損値数を、項目情報ペア情報テーブルにおいて一致度算出ペアに対応する保存領域に設定登録する。

#### 【 0 0 8 7 】

ここで、未処理の項目情報ペア情報が存在する場合（ステップ 1 0 8 で Y）、前述したステップ 1 0 5 ~ 1 0 7 を繰り返し実行する。ここで、未処理の項目情報ペア情報というのは、算出された一致度範囲が一致判定用閾値に達していない項目情報ペア情報を除き、一致度が算出されていない項目情報ペア情報がこれに該当する。算出された一致度範囲が一致判定用閾値に達していない項目情報ペア情報を除くのは、前述したように一致度を算出する必要がないからである。  
20

#### 【 0 0 8 8 】

前述したステップ 1 0 5 ~ 1 0 7 を繰り返し実行した結果、未処理の項目情報ペア情報が存在しなくなった場合（ステップ 1 0 8 で N）、一致度算出処理制御部 1 7 は、同一項目判定部 1 5 を起動する。同一項目判定部 1 5 は、起動されると、項目情報ペア情報テ  
30  
ブルを参照し、一致度が一致判定用閾値以上の項目情報ペア情報を抽出する。同一項目判定部 1 5 は、この抽出した項目情報ペアに含まれる各項目情報に対応する項目を同一と判定する（ステップ 1 0 9）。

#### 【 0 0 8 9 】

同一項目判定部 1 5 による判定処理が終了すると、一致度算出処理制御部 1 7 は、判定結果出力部 1 6 を起動する。判定結果出力部 1 6 は、起動されると、同一項目判定部 1 5 により同一と判定された項目情報ペア、すなわち項目情報の組を出力する（ステップ 1 1 0）。出力先は、ディスプレイ 3 8、図示しないプリンタ、あるいはファイル等でもよい。また、判定結果に基づき項目情報対応関係グラフ生成処理手段（図示せず）により生成された項目情報対応関係グラフを模式的に示した図を図 1 2 に示す。  
40

#### 【 0 0 9 0 】

管理者は、出力された項目情報対応関係グラフを参照することによって事前に同一と判明していたデータベース A とデータベース C の各「担当者 ID」の組の他に、データベース A における「建物所属」とデータベース C における「担当者所属」との組も一致判定用閾値を超えているため同一の項目であると認識することができる。

#### 【 0 0 9 1 】

図 1 3 は、本実施の形態において項目の対応付け対象とするデータベース 1 a , 1 b , 1 c の各データ設定例を示した図である。この項目データに基づき具体的な同一項目検出処理の内容について説明する。

#### 【 0 0 9 2 】

10

20

30

40

50

ペア情報設定部 11 は、一致度算出処理制御部 17 により起動されると、各データベース 1a, 1b, 1c それぞれに設定されているデータ項目の項目名を抽出する（ステップ 101）。図 13 に示した例によると、データベース 1a から“建物 ID”、“建物所属”、“担当者 ID”及び“機種”を、データベース 1b から“建物 ID”、“契約種別”、“契約年数”及び“契約条件”を、データベース 1c から“建物 ID”、“担当者 ID”、“担当者所属”、“作業種類”及び“作業条件”を、それぞれ抽出する。続いて、ペア情報設定部 11 は、抽出した項目情報から項目情報ペアを形成し、項目情報ペア情報テーブルを生成する（ステップ 102）。そして、ペア情報設定部 11 は、対応キー情報記憶部 23 から図 4 に例示した対応キー情報を取り出し、項目情報ペア情報テーブルに設定登録する（ステップ 103）。更に、ペア情報設定部 11 は、対応関係情報記憶部 22 から図 3 に例示した対応関係情報を取り出すと、このデータベース A の「担当者 ID」とデータベース C の「担当者 ID」との関係における一致度、全件数、一致件数及び欠損値数の算出を一致度算出部 14 に依頼する。

10

#### 【0093】

図 13 に示した項目データの例によると、データベース A の「担当者 ID」とデータベース C の「担当者 ID」の各項目データの全件数は 29 件、一致する項目データ数（一致件数）は 19 件、一方、一致していない項目データ数（欠損値数）は 10 件、であるので、一致度は、 $19 \div (29 - 10) = 1.00$  と算出される。ペア情報設定部 11 は、一致度算出部 14 から算出結果を受け取ると、項目情報ペア情報テーブルに設定登録する（ステップ 104）。なお、一致度範囲より先に一致度が算出されると、ペア情報設定部 11 は、一致度上限及び一致度下限の双方に一致度を設定する。

20

#### 【0094】

この結果、項目情報ペア情報テーブルに設定された項目情報ペア情報は、図 5 に示したようになり、また、この項目情報ペア情報に基づき描画される項目情報対応関係グラフは、図 7 に示すようになる。

#### 【0095】

続いて、一致度範囲算出部 12 は、項目情報ペア情報テーブルを参照することによって図 8 に示したような対応関係を持つ項目情報ペアの組を探す。この時点では対応関係情報に基づく 1 つの項目情報ペアしか登録されていないので、一致度範囲を算出できずに処理を終える（ステップ 105）。

30

#### 【0096】

続いて、一致度算出ペア特定部 13 は、一致度算出ペアを特定しようとするが、この時点では、データベース A 及びデータベース C の各「担当者 ID」のいずれかと項目情報ペアを形成しても、第 1 の処理の結果は最大値は 1 であり、かつ複数存在するので、第 2 の処理に移行する。第 2 の処理でも同様に、1 つの項目情報ペアしか存在しないこの時点では、ただ 1 つの項目情報を特定できないため、第 3 の処理に移行する。この時点では、一致度範囲は、対応関係情報に予め登録されていた、データベース A の「担当者 ID」とデータベース C の「担当者 ID」との組しか存在されず、かつこの組は一致度がすでに算出されているので、第 4 の処理に移行する。第 4 の処理では、任意の 1 つの項目情報ペアを一致度算出ペアとして特定する。ここでは、一致度算出ペア特定部 13 により無作為に選出された結果、データベース A の「担当者 ID」とデータベース C の「担当者所属」とから成る項目情報ペアが一致度算出ペアとして特定された場合の例を示す。

40

#### 【0097】

図 13 における項目データの設定例によると、全件数は 29 件、一致する項目データ数（一致件数）は 0 件、一方、一致していない項目データ数（欠損値数）は 10 件、であるので、一致度は、 $0 \div (29 - 10) = 0.00$  と算出される（ステップ 107）。一致度算出部 14 は、この算出結果を項目情報ペア情報テーブルに設定登録する。なお、一致度範囲より先に一致度が算出されると、一致度算出部 14 は、一致度上限及び一致度下限の双方に一致度を設定する。このときの項目情報ペア情報テーブルのデータ設定例を図 14 に、この項目情報ペア情報テーブルのデータ設定の内容に基づき生成された項目情報対

50



応関係グラフを模式的に示した図を図 15 に、それぞれ示す。

【 0 0 9 8 】

ここで、未処理の項目情報ペアは存在するので（ステップ 108 で N）、ステップ 105 に移行する。一致度範囲算出部 12 は、項目情報ペア情報テーブルを参照することによって図 8 に示したような対応関係を持つ項目情報ペアの組を探す。図 15 によると、この時点では、データベース A の「担当者 ID」を共通とし、他方の項目情報（共にデータベース C における「担当者 ID」と「担当者所属」）の間で一致度範囲が算出されていない項目情報ペアの組が存在する。従って、一致度範囲算出部 12 は、上記式（1）を用いてデータベース C における「担当者 ID」と「担当者所属」との間の一致度の取りうる範囲を算出する。具体的には、図 15 に示した対応関係を図 8 に当てはめると、一致度は 1.00、一致度は 0.00 なので、0 と算出される。一致度範囲算出部 12 は、このようにして求めた一致度範囲を、項目情報ペア情報テーブルにおいて対応するデータベース C における「担当者 ID」と「担当者所属」との間の一致度上限及び一致度下限に設定する（ステップ 105）。このときの項目情報ペア情報テーブルのデータ設定例を図 16 に示す。

10

【 0 0 9 9 】

続いて、一致度算出ペア特定部 13 は、前述した処理手順に従って項目情報ペアの中から一致度の算出対象とする一致度算出ペアを特定する（ステップ 106）。ここでの処理は、前回と異なり、データベース A における「担当者 ID」とデータベース C における「担当者 ID」及び「担当者所属」から図 9 に示した対応関係となる項目情報ペアの組が存在するが、項目 E、I の位置にどの項目情報を設定しても、一致度範囲算出対象の最大値は 1 となる。つまり、複数の項目情報ペアが一致度算出ペアの候補となるので、第 2 の処理に移行する。第 2 の処理においてもデータベース A における「担当者 ID」とデータベース C における「担当者 ID」との間における一致度 1.00 が最大なので、データベース A における「担当者 ID」と項目情報ペアを形成する項目情報が複数存在するので、第 2 の処理ではただ 1 つの一致度算出ペアを特定できず、よって第 3 の処理に移行する。第 3 の処理においても一致度範囲の最大となる項目情報ペアは存在しないので、ここでも第 4 の処理に移行し、第 4 の処理によって任意の 1 つの項目情報ペアが一致度算出ペアとして特定される。ここでは、一致度算出ペア特定部 13 により無作為に選出された結果、データベース A の「担当者 ID」とデータベース C の「作業種類」とから成る項目情報ペアが一致度算出ペアとして特定されたものとして説明を続ける。

20

30

【 0 1 0 0 】

なお、図 16 を参照すると、この時点で一致度上限が一致判定用閾値以上であって一致度がまだ算出されていない項目情報ペアは存在しないので、第 4 の処理によって一致度算出ペアが選出されるのは明らかである。ちなみに、今までのステップ 105 ~ 107 の処理によって、データベース A の「担当者 ID」とデータベース C の「担当者所属」との組、よってデータベース C の「担当者 ID」とデータベース C の「担当者所属」との組は、それぞれ一致度上限が一致判定用閾値を下回ることは判明しているので、この 2 組の項目情報ペアは、一致度算出ペアを特定する処理において一致度算出対象候補から事前に外されている。

40

【 0 1 0 1 】

図 13 における項目データの設定例によると、データベース A の「担当者 ID」とデータベース C の「作業種類」との関係において、全件数は 29 件、一致する項目データ数（一致件数）は 0 件、一方、一致していない項目データ数（欠損値数）は 10 件、であるので、一致度は、 $0 \div (29 - 10) = 0.00$  と算出される（ステップ 107）。一致度算出部 14 は、この算出結果を項目情報ペア情報テーブルに設定登録する。なお、一致度範囲より先に一致度が算出されると、一致度算出部 14 は、一致度上限及び一致度下限の双方に一致度を設定する。このときの項目情報ペア情報テーブルのデータ設定例を図 17 に、この項目情報ペア情報テーブルのデータ設定の内容に基づき生成された項目情報対応関係グラフを模式的に示した図を図 18 に、それぞれ示す。

50

## 【 0 1 0 2 】

このとき、未処理の項目情報ペアは存在するので（ステップ 1 0 8 で N）、ステップ 1 0 5 に移行するが、未処理の項目情報ペアが存在しなくなるまでステップ 1 0 5 ~ 1 0 7 は、以上のように繰り返し実行される。

## 【 0 1 0 3 】

ここで、ステップ 1 0 6 における一致度算出ペア特定処理において、データベース A の「建物所属」とデータベース C の「担当者所属」とから成る項目情報ペアが一致度算出ペアとして特定された場合における一致度算出部 1 4 における処理について説明する。

## 【 0 1 0 4 】

図 1 3 における項目データの設定例によると、データベース A の「建物所属」とデータベース C の「担当者所属」との関係において、全件数は 2 9 件、一致する項目データ数（一致件数）は 1 7 件、一方、一致していない項目データ数（欠損値数）は 1 0 件、であるので、一致度は、 $17 \div (29 - 10) = 0.89$  と算出される（ステップ 1 0 7）。一致度算出部 1 4 は、この算出結果を項目情報ペア情報テーブルに設定登録する。なお、一致度範囲より先に一致度が算出されたので、一致度算出部 1 4 は、一致度上限及び一致度下限の双方に一致度を設定する。このときの項目情報ペア情報テーブルのデータ設定例の一部を図 1 9 に示す。

## 【 0 1 0 5 】

未処理の項目情報ペアは存在する場合（ステップ 1 0 8 で N）、前述したようにステップ 1 0 5 ~ 1 0 7 の処理を繰り返すことになるが、この処理を繰り返すことによって、一致度が算出された項目情報ペア、及び一致度範囲が算出されかつ一致度上限が一致判定用閾値を超えていない項目情報ペアが増え、いずれ未処理の項目情報ペア、すなわち、一致度範囲が設定されていない項目情報ペア、又は一致度が算出されてなくかつ設定された一致度上限が一致判定用閾値以上の項目情報ペアは存在しなくなる。本実施の形態では、一致度が算出された項目情報ペアが増えると、ステップ 1 0 5 における 1 度の一致度範囲算出処理において複数の項目情報ペアに対して一致度範囲が算出されてくるようになる。一致度上限が一致判定用閾値を超えていない項目情報ペアは、ステップ 1 0 6 における処理の対象外とすることができるので、換言すると一致度を算出する必要がないので、同一である項目の検出に要する処理負荷を軽減することができるようになる。

## 【 0 1 0 6 】

未処理の項目情報ペアは存在しなくなると（ステップ 1 0 8 で N）、一致度算出処理制御部 1 7 は、同一項目判定部 1 5 を起動する。

## 【 0 1 0 7 】

同一項目判定部 1 5 は、起動されると、項目情報ペア情報テーブルを参照し、一致度が一致判定用閾値以上となる項目情報ペアを抽出し、当該項目情報ペアに含まれている各項目情報に対応する項目は同一であると判定する（ステップ 1 0 9）。同一項目判定部 1 5 が項目の同一性を判定すると、同一項目判定部 1 5 は、判定結果を HDD 3 4 又は RAM 3 3 などの記憶手段に記憶させる。

## 【 0 1 0 8 】

その後、例えば、ユーザは、マウス 3 6 等の入力手段を用いて所定の操作にすることで判定結果の出力を要求する。一致度算出処理制御部 1 7 は、ユーザからの判定結果出力要求に応じて判定結果出力部 1 6 を起動する。判定結果出力部 1 6 は、起動されると、記憶手段から同一項目判定部 1 5 による判定結果を読み出し、その判定結果を所定の形式にてディスプレイ 3 8 などユーザにより指定された出力手段に出力させる（ステップ 1 1 0）。例えば、判定結果出力部 1 6 は、項目情報対応関係グラフ生成処理手段（図示せず）により判定結果を表示する項目情報対応関係グラフを生成させ、それを表示するようにしてもよい。図 2 0 は、この項目情報対応関係グラフの表示例を示した図である。図 1 9 に示したように、項目情報対応関係グラフには、キー項目の対応関係を示す線 1 0 4 と、一致度が一致判定用閾値 0.8 以上となった項目情報ペアを表す線 1 0 5 と、が含まれており、線 1 0 5 には、一致度が対応付けて表示される。

## 【 0 1 0 9 】

なお、本実施の形態では、一致度が一致判定用閾値 0.8 以上となった項目情報ペアのみを表示するようにしたが、一致判定用閾値を下回る情報を表示するようにしてもよい。

## 【 0 1 1 0 】

管理者は、この表示内容を参照して、同一と判定された項目の項目データを確認する。また、データベースを結合するなどの作業を行ってもよい。

## 【 0 1 1 1 】

図 2 1 は、同一項目判定部 1 5 による判定結果をレポート形式にて出力した場合のレポートの一例を示した図である。図 2 1 のように、不一致の項目データを明示して判定結果を出力するようにしてもよい。

10

## 【 0 1 1 2 】

本実施の形態においては、以上説明した処理によって同一である項目を検出し、その結果を管理者に提示するようにしたが、これに限定することなく種々の変形例を適用してもよい。

## 【 0 1 1 3 】

例えば、一致度範囲算出部 1 2 は、式 ( 1 ) を用いて一致度の取りうる範囲を算出したが、以下のようにして算出してもよい。

## 【 0 1 1 4 】

一致度範囲を算出する項目情報ペアに含まれる各項目の項目データからランダムに選択した  $N'$  件の範囲での一致度を  $\gamma'$  としたとき、全体の一致度を信頼率 95% で区間推定すると、一致度は、

20

## 【 数 1 】

$$\gamma' - 1.96 \sqrt{\frac{\gamma'(1-\gamma')}{N'}} \leq \gamma \leq \gamma' + 1.96 \sqrt{\frac{\gamma'(1-\gamma')}{N'}}$$

という式にて算出できる。この計算式の詳細については、「基礎統計学Ⅰ 統計学入門」東京大学教養学部統計学教室編、東京大学出版会、1991年を参照できる。この統計的方式を用いることで、項目情報ペアに含まれる各項目の全ての項目データを用いずに一致度の取りうる範囲を算出できる。

30

## 【 0 1 1 5 】

ところで、項目データが文字列の場合、人的な入力ミス等により本来一致するはずの項目データが異なるデータと判断されてしまう可能性がある。例えば、“営業部設備課”と入力するべきところを“営業部設営課”と誤入力してしまうと、一致度算出部 1 4 により一致度を算出する過程において一致しない項目データとみなされ欠損値数として加算されてしまうことになる。従って、このような誤入力を検出する機能を設けるようにしてもよい。

## 【 0 1 1 6 】

本実施の形態では、異なる値の項目データの組が実際には一致していると判定するために、一致条件としてそれぞれに閾値 ( 閾値 1 , 閾値 2 ) を利用した 2 つの条件を予め設定している。第 1 の条件は、照合する一方の項目データを形成する文字列 1 と他方の項目データを形成する文字列 2 とで相違する文字数が閾値 1 以下である、ということである。第 2 の条件は、各項目データの当該データベースに含まれる件数 ( 全レコード数 ) に占める割合をそれぞれ算出し、算出した割合の多い方から少ない割合で除算した数が閾値 2 以上である、ということである。ここでは、閾値 1 = 1、閾値 2 = 100 として具体的な例を挙げて説明する。

40

## 【 0 1 1 7 】

例えば、照合する項目データが“営業部設備課”及び“営業部設営課”という文字列であったとする。各項目データにおいて相違する文字は、“備”と“営”であり、文字数は

50

1である。従って、第1の条件として設定された閾値  $1 = 1$  以下であるため、この第1の条件は満たすことになる。

【0118】

続いて、当該データベースにおける全項目データに含まれる“営業部設備課”の数が7%、当該データベースにおける全項目データに含まれる“営業部設営課”の数が0.01%、であったとする。この場合、大きい数を小さい数で除算するので、 $7\% / 0.01\% = 700$ となり、第2の条件として設定された閾値  $2 = 100$  以上であるため、この第2の条件も満たすことになる。

【0119】

このように、一致していない1組の項目データに含まれている各文字列の間に、予め設定された一致条件を満たす関係がある場合に、一方の項目データは誤入力された項目データと判定する誤入力判定手段（図示せず）を設けて、比較する各項目データが一致しない場合でも、一致条件を満たしたことで当該各項目データは一致していると判定する。一方、いずれかの1つの条件でも満たさない場合は当該各項目データは一致していないと判定する。

10

【0120】

この誤入力判定手段は、例えば一致度算出部14の一機能としてもよい。この場合、一致度算出部14は、照合により一致していないと判定した場合でも、前述した一致条件を満たした場合には、照合した1組項目データは一致していると判定する。これにより、この項目データの組を欠損値数に加算すべきところを一致件数として加算する。また、前述した第2の条件の結果を参照し、データベースに含まれる件数に占める割合の少ない方の項目データが誤入力されていると判断して、当該項目データを自動的に修正してもよい。

20

【0121】

この場合、修正した項目データに関する情報は、修正した旨、修正した項目データを特定する所定の項目データ変更情報を出力するなどして管理者に通知する。なお、この通知機能は、一致度算出部14に持たせてもよいし、判定結果出力部16に持たせてもよい。あるいは、別の手段として設けてもよい。

【0122】

また、誤入力判定手段は、例えば同一項目判定部15又は判定結果出力部16の一機能としてもよい。あるいは、別の手段として設けてもよい。この場合、一致度算出部14は、該当する項目データの組は欠損値数として加算して一致度を算出することになる。ただ、誤入力判定手段は、判定結果、誤入力有りと判定した項目データの組を特定する情報、また、例えば、“営業部設営課”が“営業部設備課”の誤入力である可能性があります。」などのように正しい入力を示唆する情報を含む誤入力通知情報を出力する機能を持つ必要がある。誤入力通知情報は、単独で出力してもよいし、図6のステップ110において判定結果と共に、あるいは判定結果に含めて出力してもよい。

30

【0123】

なお、前述した変形例では、一致条件として2つの条件を設定したが、設定する条件の数や内容はこれに限定する必要はない。また、条件として設定する閾値もこれに限定する必要はない。

40

【0124】

ところで、本実施の形態においては、一致していると管理者が認識している項目情報ペアを対応関係情報テーブルに予め設定した。ただ、これとは逆の関係についての情報も事前に設定登録してもよい。すなわち、管理者により一致していないと認識されている項目情報ペアを特定する情報を予め設定しておき、この項目情報ペアに関しては一致度の算出対象にしないようにする。これにより、一致していないことが明らかな項目情報ペアに対しては、一致度算出ペアとして選出させない、つまり一致度を算出対象の候補から除外するようにしたので、同一項目の検出処理を効率的に実行させることができる。

【0125】

本実施の形態では、不一致関係情報テーブルを記憶する不一致関係情報記憶手段を設け

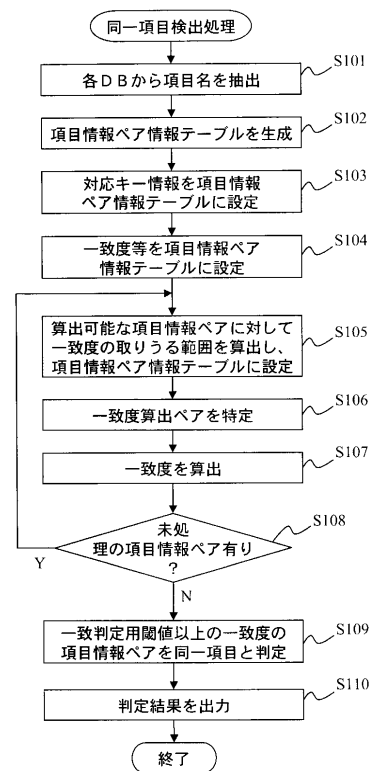
50



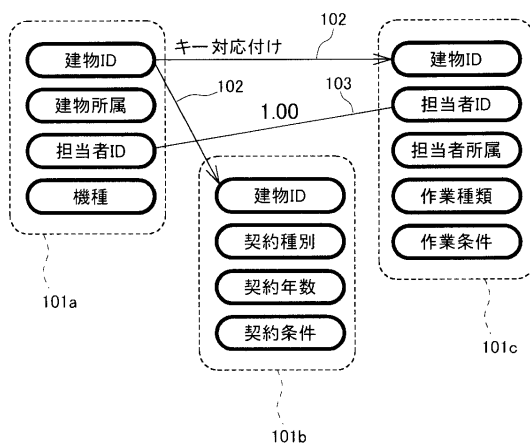
【図 5】

項目情報ペア情報テーブル				一致度下限	一致度上限	欠損値数	一致件数	全件数	一致度	対応キー情報	項目情報2	項目情報1
A	建物ID	A, 建物ID									A	建物ID
A	建物所属	A, 建物ID									A	建物ID
A	担当者ID	A, 建物ID									A	担当者ID
.	.	.									.	.
A	建物ID	A, 建物ID B, 建物ID									A	建物ID
.	.	.									.	.
A	担当者ID	A, 建物ID C, 建物ID				10	19	29	1.00		A	担当者ID
.	.	.									.	.
B	建物ID	B, 建物ID C, 建物ID									B	建物ID
.	.	.									.	.
C	作業種類	C, 建物ID C, 建物ID									C	作業種類

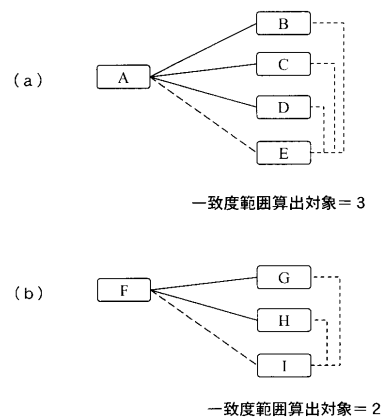
【図 6】



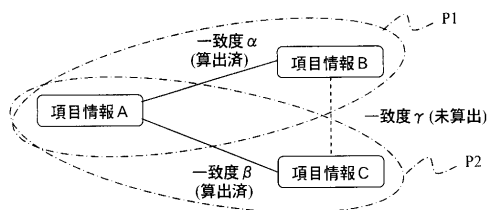
【図 7】



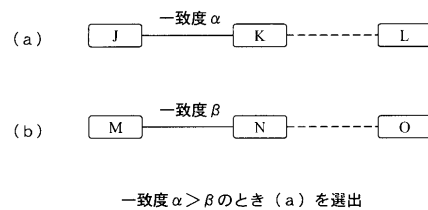
【図 9】



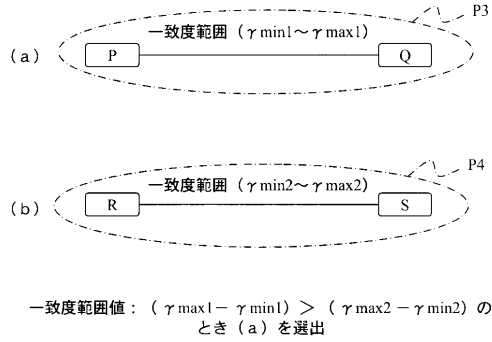
【図 8】



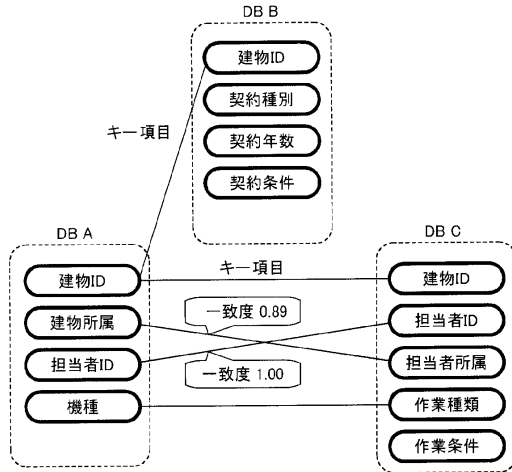
【図 10】



【図 1 1】



【図 1 2】



【図 1 4】

項目情報1	項目情報2	対応キー情報	一致度	全件数	一致件数	欠損値数	一致度上限	一致度下限
.	.	.	.	.	.	.	.	.
A 担当者ID	C 担当者ID	A, 建物ID C, 建物ID	1.00	29	19	10	1.00	1.00
A 担当者ID	C 担当者所属	A, 建物ID C, 建物ID	0.00	29	0	10	0.00	0.00
.	.	.	.	.	.	.	.	.

項目情報ベア情報テーブル

【図 1 3】

データベースC

建物ID	担当者ID	担当者所属	作業種別	作業条件
AA0001	aa101	A部1課	C-B1-a	なし
AA0001	aa101	A部1課	C-A3-a	立会
AA0001	aa101	A部1課	C-B1-b	なし
AA0002	aa101	A部1課	C-A1-a	夜間
AA0003	aa102	A部1課	C-A2-a	夜間
AA0003	aa102	A部1課	C-A2-b	なし
AA0003	aa102	A部1課	C-A2-c	夜間
AA0003	aa102	A部1課	C-A2-d	夜間
AA0005	aa201	A部2課	C-A2-b	立会
AA0005	aa201	A部2課	C-B1-a	立会
AA0008	aa302	A部3課	C-B2-a	なし
BB0001	bb101	B部1課	C-B2-a	なし
BB0001	bb101	B部1課	C-B2-b	夜間
BB0001	bb101	B部1課	C-B2-c	なし
BB0002	bb102	B部2課	C-A3-a	無停止
BB0002	bb102	B部2課	C-B2-a	なし
BB0005	bb201	B部2課	C-B1-a	なし
BB0005	bb201	B部2課	C-B1-a	なし
BB0006	bb201	B部2課	C-A3-a	立会

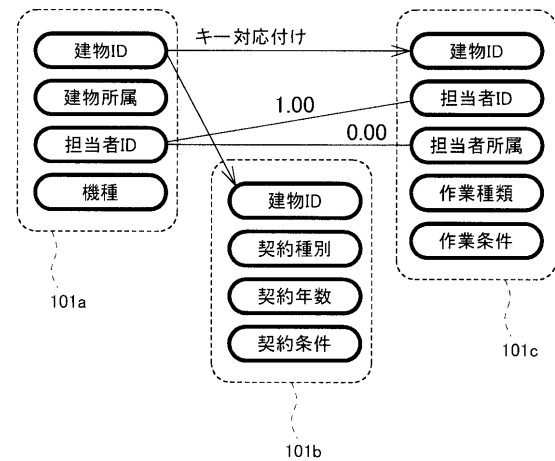
データベースB

建物ID	契約種別	契約年数	契約条件
AA0001	C-B1	3	なし
AA0001	C-A3	5	なし
AA0002	C-A1	9	夜間
AA0003	C-A2	12	夜間
AA0004	C-A3	8	無停止
AA0004	C-B2	11	無停止
AA0005	C-A2	3	立会
AA0006	C-A2	12	立会
AA0007	C-A2	8	無停止
AA0008	C-B1	3	なし
BB0001	C-B2	10	なし
BB0002	C-A3	2	無停止
BB0002	C-B2	11	なし
BB0003	C-A3	3	なし
BB0004	C-B1	6	無停止
BB0005	C-B1	2	なし
BB0006	C-A3	12	立会
BB0007	C-A1	8	無停止

データベースA

建物ID	建物所属	担当者ID	機種
AA0001	A部1課	aa101	XX2
AA0002	A部1課	aa101	YY1
AA0003	A部1課	aa102	YY2
AA0004	A部2課	aa201	YY1
AA0005	A部2課	aa201	XX2
AA0006	A部3課	aa301	ZZ1
AA0007	A部3課	aa302	YY2
AA0008	A部3課	bb101	YY2
BB0001	B部1課	bb102	YY2
BB0002	B部1課	bb102	YY1
BB0003	B部2課	bb201	ZZ1
BB0004	B部2課	bb201	XX1
BB0005	B部2課	bb201	YY2
BB0006	B部2課	cc101	YY2
BB0007	B部2課	cc101	ZZ1
CC0001	C部1課	cc101	YY1
CC0002	C部1課	cc102	XX2
CC0003	C部1課	cc102	YY2
CC0004	C部1課	cc103	YY2

【図 1 5】



【図 16】

項目情報ベース情報テーブル

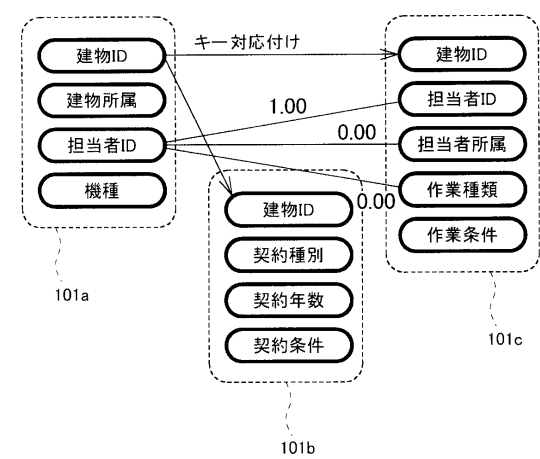
項目情報1	項目情報2	対応キー情報	一致度	全件数	一致件数	欠損値数	一致度上限	一致度下限
.	.							
A 担当者ID	C 担当者ID	A, 建物ID C, 建物ID	1.00	29	19	10	1.00	1.00
A 担当者ID	C 担当者所属	A, 建物ID C, 建物ID	0.00	29	0	10	0.00	0.00
.	.							
C 担当者ID	C 担当者所属	C, 建物ID C, 建物ID					0.00	0.00
.	.							

【図 17】

項目情報ベース情報テーブル

項目情報1	項目情報2	対応キー情報	一致度	全件数	一致件数	欠損値数	一致度上限	一致度下限
.	.							
A 担当者ID	C 担当者ID	A, 建物ID C, 建物ID	1.00	29	19	10	1.00	1.00
A 担当者ID	C 担当者所属	A, 建物ID C, 建物ID	0.00	29	0	10	0.00	0.00
A 担当者ID	C 作業種類	A, 建物ID C, 建物ID	0.00	29	0	10	0.00	0.00
.	.							
C 担当者ID	C 担当者所属	C, 建物ID C, 建物ID					0.00	0.00
.	.							

【図 18】



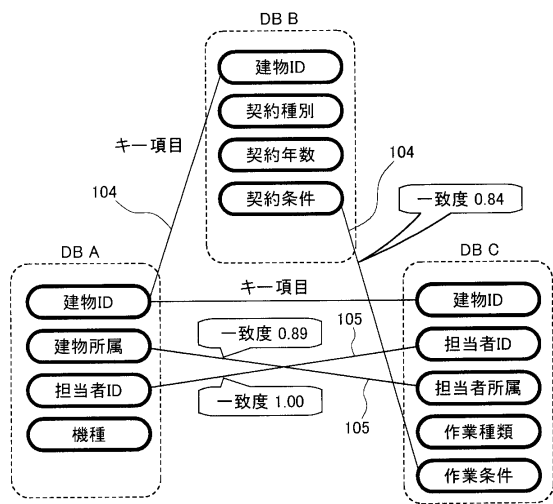
【図 19】

項目情報ベース情報テーブル

項目情報1	項目情報2	対応キー情報	一致度	全件数	一致件数	欠損値数	一致度上限	一致度下限
.	.							
A 担当者ID	C 担当者ID	A, 建物ID C, 建物ID	1.00	29	19	10	1.00	1.00
A 担当者ID	C 担当者所属	A, 建物ID C, 建物ID	0.00	29	0	10	0.00	0.00
A 担当者ID	C 作業種類	A, 建物ID C, 建物ID	0.00	29	0	10	0.00	0.00
.	.							
A 建物所属	C 担当者所属	A, 建物ID C, 建物ID	0.89	29	17	10	0.89	0.89
.	.							
C 担当者ID	C 担当者所属	C, 建物ID C, 建物ID					0.00	0.00
.	.							



【図 2 0】



【図 2 1】

項目同一性判定結果

データベースAの項目「建物所属」とデータベースCの項目「担当者所属」は、「建物ID」をキーとして対応付けを行った場合、89%が一致しています。

以下の不一致と判定された11%に該当する項目データを修正することにより、データの適正化を図ることができます。

DB A	DB C	キー（「建物ID」）
A部B課	A部R課	A1111
C部D課	F部D課	B1111
E部F課	T部Q課	B2222

---

フロントページの続き

(56)参考文献 国際公開第2008/011424(WO,A1)

特開2012-128760(JP,A)

特開2008-250861(JP,A)

(58)調査した分野(Int.Cl.,DB名)

G06F 12/00

G06F 17/30