

明 細 書

発明の名称：データ連携支援装置及びデータ連携支援方法

技術分野

[0001] 本発明は、データ連携支援装置及びデータ連携支援方法に関するものである。

背景技術

[0002] システム間の連携、あるいは、システム同士の統合に伴って、システム別のデータベース間のデータ連携、あるいは、同じデータベース内の別々のテーブル間のデータ連携が必要となる。データ連携では、連携先となるテーブル（即ち、連携先テーブル）のカラムに格納するデータを、どのテーブル（即ち、連携元テーブル）のカラムから抽出すればよいかというデータ連携関係を定義する必要がある。

[0003] データ連携関係を定義するためには、テーブル間のカラムの対応関係を何らかの手段で検出する必要がある。さらに、連携元テーブルから連携先テーブルのカラムの組み合わせであるテーブルレイアウトがテーブル操作によって作成できる必要がある。このようなデータ構造の変換が実施できないと、連携先テーブルにデータを格納することができない。連携先テーブルとレイアウトが一致する連携元テーブルがなく、2つ以上の連携元テーブルからデータを集めて連携先テーブルを作成する場合には、テーブル間のカラムの類似性によって対応関係を判断するだけでは不十分である。実際にそれら2つ以上の連携元テーブルに対して結合（例えば、JOIN）操作を実施できることが必要である。

[0004] 上記のように、データ連携では、連携元テーブルを結合（例えば、JOIN）して連携先テーブルを生成する必要がある。よって、データ連携関係を定義するためには、事前にテーブル間の参照関係を特定する必要がある。しかし、一般的なデータベースでは、主キー制約は付与されているものの、外部参照制約が付与されておらず、テーブル間の参照関係が明らかでないこと

が多い。そのため、従来は、テーブル間の参照関係を人手で抽出していた。その作業は、データベース技術と業務内容を理解した高度な技術者が労働集約的に実施する必要があった。

[0005] テーブル間の参照関係を自動抽出する技術として、データベースのテーブルのレコード（即ち、実データ）を利用する技術が存在する（例えば、非特許文献1 参照）。

[0006] 類似するカラムの対応関係を自動抽出する技術として、いわゆるスキーママッチング技術が存在する（例えば、非特許文献2 参照）。

先行技術文献

非特許文献

[0007] 非特許文献1: Ling Ling Yan, Renee J. Miller, Laura M. Haas, and Ronald Fagin, “Data-Driven Understanding and Refinement of Schema Mappings”, SIGMOD '01 Proceedings of the 2001 ACM SIGMOD international conference on Management of data, 2001

非特許文献2: Ronald Fagin, Laura M. Haas, Mauricio Hernandez, Renee J. Miller, Lucian Popa, and Yannis Velegrakis, “Clio: Schema Mapping Creation and Data Exchange”, Lecture Notes in Computer Science Volume 5600, 2009

発明の概要

発明が解決しようとする課題

[0008] 非特許文献1 では、データベースのテーブルのレコードを読み込み、レコ

ード間の関係性及びテーブル間の結合関係を推測して、テーブル間の参照関係を自動抽出する。参照関係の抽出精度を高めるためには、全てのレコードを利用することが必要である。その場合、大量のレコードに対する処理が必要となる。そのような処理を大規模なシステムに対して実行した場合、処理時間が実用的な時間に収まらない可能性が高い。必然的に処理の対象とするシステムを絞り込まざるを得ない。

[0009] 利用するレコードの数を減らすためにサンプリング技術を活用することも可能である。しかし、テーブル間の参照関係を考慮しながらサンプリングを行うことは困難である。サンプリングの結果、本来は参照関係のあるテーブルが、参照関係のないように見えてしまう可能性がある。

[0010] 以上のように、非特許文献1のようなレコードの利用が必須となる技術は、レコードの量によって適用できるシステムの範囲が制限されるという課題がある。

[0011] 非特許文献2のようなスキーママッチング技術では、テーブル定義情報のようなメタデータのみから、類似するカラムの対応関係を抽出できる。そのため、レコードを利用する技術と比較すると、扱うデータ量が少なく済むという利点がある。大規模なシステム全体を処理の対象とすることができる。

[0012] しかし、スキーママッチング技術は、類似するカラムを見つけ出す技術である。そのため、テーブル間の参照関係を自動抽出するために、そのまま利用することはできない。特に、データベース内に同義のカラムが多数存在する場合、スキーママッチング技術の観点からは正しい対応関係であっても、参照関係ではない対応関係が多数検出されてしまう。

[0013] また、スキーママッチング技術では、データ連携を行うために連携元テーブルから連携先テーブルのレイアウトをテーブル操作によって作成する必要がある点について考慮されていないという課題がある。

[0014] 本発明は、例えば、データ連携において、連携先となるテーブルと、どのテーブルとの間に参照関係があるかを、効率よく検出することを目的とする

。

課題を解決するための手段

- [0015] 本発明の一の態様に係るデータ連携支援装置は、
複数のテーブルを有するデータベースのデータ連携を支援する装置であり、
、
前記複数のテーブルのそれぞれに含まれる複数のカラムのそれぞれが主キーになっているかどうかを示す主キー情報と、前記複数のテーブルのそれぞれに含まれる複数のカラムのそれぞれに格納されるデータの属性を示す属性情報とを記憶する情報記憶部と、
前記複数のテーブルのうち、データ連携の連携先となる1つのテーブルを選択するテーブル選択部と、
前記複数のテーブルのうち前記テーブル選択部により選択されたテーブル以外のテーブルのそれぞれについて、前記情報記憶部に記憶された主キー情報に基づき、該当するテーブルに含まれる複数のカラムから、主キーになっている第1カラムと、主キーになっていない第2カラムとを特定するカラム特定部と、
前記複数のテーブルのうち前記テーブル選択部により選択されたテーブル以外のテーブルから、前記情報記憶部に記憶された属性情報に基づき、前記テーブル選択部により選択されたテーブルに、前記カラム特定部により特定された第1カラムと類似する第3カラムが含まれ、かつ、前記テーブル選択部により選択されたテーブルに、前記カラム特定部により特定された第2カラムと類似する第4カラムが含まれるテーブルを抽出するテーブル抽出部と、
、
前記テーブル選択部により選択されたテーブルと前記テーブル抽出部により抽出されたテーブルとの組み合わせをデータ連携に利用することを推奨する推奨情報を出力する情報出力部とを備える。

発明の効果

- [0016] 本発明の一の態様によれば、データ連携支援装置が、データ連携において

、連携先となるテーブルと、どのテーブルとの間に参照関係があるかを、効率よく検出することができる。

図面の簡単な説明

- [0017] [図1]実施の形態1に係るデータ連携支援装置の構成を示すブロック図。
- [図2]実施の形態1に係る連携元情報の例を示す表。
- [図3]実施の形態1に係るデータ連携支援装置の動作の例を示すフローチャート。
- [図4]実施の形態1に係る参照元情報及び主キー定義情報の例を示す表。
- [図5]実施の形態1に係るスキーママッチングの結果の例を示す表。
- [図6]実施の形態1に係る参照関係情報の例を示す表。
- [図7]実施の形態1に係るデータ連携支援装置の動作の例を示すフローチャート。
- [図8]実施の形態1に係る非外部キー定義情報及び非主キー定義情報の例を示す表。
- [図9]実施の形態1に係るスキーママッチングの結果の例を示す表。
- [図10]実施の形態1に係る連携関係情報の例を示す表。
- [図11]実施の形態1に係るデータ連携支援装置の動作の例を示すフローチャート。
- [図12]実施の形態1に係る抽出結果の例を示す表。
- [図13]実施の形態1に係るデータ連携支援装置の動作の例を示すフローチャート。
- [図14]実施の形態1に係る入力情報及び推奨情報の例を示す表。
- [図15]実施の形態1に係るデータ連携支援装置のハードウェア構成の例を示す図。

発明を実施するための形態

[0018] 以下、本発明の実施の形態について、図を用いて説明する。

[0019] 実施の形態1.

図1は、本実施の形態に係るデータ連携支援装置100の構成を示すブロ

ック図である。

[0020] 図1において、データ連携支援装置100は、複数のテーブルを有するデータベースのデータ連携を支援する装置である。データ連携支援装置100により支援されるデータ連携は、データベースの異なるインスタンス間のデータ連携であってもよいし、同じインスタンスの異なるスキーマ間のデータ連携であってもよいし、同じスキーマの異なるテーブル間のデータ連携であってもよい。

[0021] データ連携支援装置100は、情報記憶部101、テーブル選択部102、カラム特定部103、テーブル抽出部104、情報蓄積部105、情報出力部106を備える。

[0022] 図1には示していないが、データ連携支援装置100は、処理装置、記憶装置、入力装置、出力装置等のハードウェアを備える。ハードウェアはデータ連携支援装置100の各部によって利用される。例えば、処理装置は、データ連携支援装置100の各部でデータ又は情報の演算、加工、読み取り、書き込み等を行うために利用される。記憶装置は、そのデータ又は情報を記憶するために利用される。入力装置は、そのデータ又は情報を入力するために利用される。出力装置は、そのデータ又は情報を出力するために利用される。

[0023] 情報記憶部101は、連携元情報111、連携先情報112を、記憶装置を利用して記憶する。

[0024] 連携元情報111は、データ連携の連携元となるデータベース（即ち、連携元データベース）の複数のテーブル（即ち、連携元テーブル）を定義する情報である。連携元情報111には、連携元データベースの主キー情報及び属性情報が含まれる。連携元データベースの主キー情報とは、複数の連携元テーブルのそれぞれに含まれる複数のカラムのそれぞれが主キーになっているかどうかを示す情報である。連携元データベースの属性情報とは、複数の連携元テーブルのそれぞれに含まれる複数のカラムのそれぞれに格納されるデータの属性（例えば、カラム名、データ型）を示す情報である。

- [0025] 連携先情報 1 1 2 は、データ連携の連携先となるデータベース（即ち、連携先データベース）の複数のテーブル（即ち、連携先テーブル）を定義する情報である。連携先情報 1 1 2 には、連携先データベースの主キー情報及び属性情報が含まれる。連携先データベースの主キー情報とは、複数の連携先テーブルのそれぞれに含まれる複数のカラムのそれぞれが主キーになっているかどうかを示す情報である。連携先データベースの属性情報とは、複数の連携先テーブルのそれぞれに含まれる複数のカラムのそれぞれに格納されるデータの属性（例えば、カラム名、データ型）を示す情報である。
- [0026] 連携元情報 1 1 1 及び連携先情報 1 1 2 は、入力装置を介して入力され、情報記憶部 1 0 1 に記憶される。
- [0027] テーブル選択部 1 0 2 は、複数の連携先テーブルのうち、1 つの連携先テーブルを選択する。
- [0028] 例えば、テーブル選択部 1 0 2 は、情報記憶部 1 0 1 に記憶された連携先情報 1 1 2 を読み取る。テーブル選択部 1 0 2 は、読み取った連携先情報 1 1 2 により定義された連携先テーブルから、1 つの連携先テーブルを選択する。
- [0029] カラム特定部 1 0 3 は、連携元データベース及び連携先データベースの複数のテーブルのうちテーブル選択部 1 0 2 により選択された連携先テーブル以外のテーブルのそれぞれについて、前述した主キー情報に基づき、該当するテーブルに含まれる複数のカラムから、主キーになっている第 1 カラムと、主キーになっていない第 2 カラムとを特定する。
- [0030] 例えば、カラム特定部 1 0 3 は、情報記憶部 1 0 1 に記憶された連携元情報 1 1 1 及び連携先情報 1 1 2 を読み取る。カラム特定部 1 0 3 は、読み取った連携元情報 1 1 1 により定義された連携元テーブルのそれぞれについて、読み取った連携元情報 1 1 1 に含まれる主キー情報に基づき、第 1 カラムと第 2 カラムとを特定する。また、カラム特定部 1 0 3 は、読み取った連携先情報 1 1 2 により定義された連携先テーブルのうちテーブル選択部 1 0 2 により選択された連携先テーブル以外の連携先テーブルのそれぞれについて

、読み取った連携先情報 112 に含まれる主キー情報に基づき、第 1 カラムと第 2 カラムとを特定する。

[0031] テーブル抽出部 104 は、連携元データベース及び連携先データベースの複数のテーブルのうちテーブル選択部 102 により選択された連携先テーブル以外のテーブルのそれぞれについて、前述した属性情報に基づき、以下の（１）及び（２）の条件を満たすテーブルを抽出し、抽出結果 141 を出力する。

（１）テーブル選択部 102 により選択された連携先テーブルに、カラム特定部 103 により特定された第 1 カラムと類似する第 3 カラムが含まれる。

（２）テーブル選択部 102 により選択された連携先テーブルに、カラム特定部 103 により特定された第 2 カラムと類似する第 4 カラムが含まれる。

[0032] 本実施の形態において、テーブル抽出部 104 は、カラム特定部 103 により 2 つ以上の第 1 カラムが特定されたテーブルについては、テーブル選択部 102 により選択された連携先テーブルに、全ての第 1 カラムに対して個別に第 3 カラムが存在するテーブルを抽出する。例えば、ある連携元テーブルに 2 個の第 1 カラムが存在していたとする。テーブル選択部 102 により選択された連携先テーブルに、当該 2 個の第 1 カラムの一方と類似する第 3 カラムと、当該 2 個の第 1 カラムの他方と類似する別の第 3 カラムとが存在していれば、その連携元テーブルは上記（１）の条件を満たしていることになる。なお、連携元テーブルだけでなく、テーブル選択部 102 により選択された連携先テーブルとは別の連携先テーブルが上記（１）の条件を満たすこともあり得る。

[0033] 本実施の形態において、テーブル抽出部 104 は、カラム特定部 103 により 2 つ以上の第 2 カラムが特定されたテーブルについては、テーブル選択部 102 により選択された連携先テーブルに、最も多くの第 2 カラムに対して個別に第 4 カラムが存在するテーブルのみを抽出結果 141 に含める。例えば、ある連携元テーブルに 5 個の第 2 カラムが存在していたとする。テーブル選択部 102 により選択された連携先テーブルに、当該 5 個の第 2 カラ

ムのうち１個と類似する第４カラムと、当該５個の第２カラムの他方と類似する別の第４カラムとが存在していたとする。他の連携元テーブルについて、２個より多くの第４カラムが存在せず、また、テーブル選択部１０２により選択された連携先テーブルとは別の連携先テーブルについても、２個より多くの第４カラムが存在しなければ、その連携元テーブルは上記（２）の条件を満たしていることになる。なお、連携元テーブルだけでなく、テーブル選択部１０２により選択された連携先テーブルとは別の連携先テーブルが上記（２）の条件を満たすこともあり得る。

[0034] 本実施の形態において、テーブル抽出部１０４は、テーブル選択部１０２により選択された連携先テーブルと同じスキーマに属するテーブルと、テーブル選択部１０２により選択された連携先テーブルと同じインスタンスの異なるスキーマに属するテーブルと、テーブル選択部１０２により選択された連携先テーブルと異なるインスタンスのスキーマに属するテーブルとの３種類のテーブルを抽出結果１４１に含めることができる。

[0035] 情報蓄積部１０５は、参照関係情報１５１、連携関係情報１５２を、記憶装置を利用して記憶する。

[0036] 参照関係情報１５１は、テーブル選択部１０２により選択された連携先テーブルとテーブル抽出部１０４により抽出されたテーブルとの間で、参照関係をもたらすカラムの組を示す情報である。即ち、参照関係情報１５１は、第１カラムと第３カラムとの組を示す情報である。

[0037] 連携関係情報１５２は、テーブル選択部１０２により選択された連携先テーブルとテーブル抽出部１０４により抽出されたテーブルとの間で、参照関係をもたらすカラム以外に、対応するカラムの組を示す情報である。即ち、参照関係情報１５１は、第２カラムと第４カラムとの組を示す情報である。

[0038] 参照関係情報１５１及び連携関係情報１５２は、テーブル抽出部１０４により作成されて情報蓄積部１０５に記憶される。参照関係情報１５１及び連携関係情報１５２は、テーブル抽出部１０４により最終的な抽出結果１４１として出力されてもよいし、一時的に情報蓄積部１０５に保存されて最終的

な抽出結果 141 を決定するために利用されてもよい。

[0039] 情報出力部 106 は、テーブル抽出部 104 から出力された抽出結果 141 に基づき、テーブル選択部 102 により選択された連携先テーブルとテーブル抽出部 104 により抽出されたテーブルとの組み合わせをデータ連携に利用することを推奨する推奨情報 161 を生成する。情報出力部 106 は、生成した推奨情報 161 を、出力装置を介して出力する。

[0040] 情報出力部 106 は、テーブル抽出部 104 から出力された抽出結果 141 に、前述した 3 種類のテーブルのうち少なくとも 2 種類のテーブルが含まれている場合、テーブルの種類ごとに、テーブル選択部 102 により選択された連携先テーブルとテーブル抽出部 104 により抽出されたテーブルとの組み合わせをデータ連携に利用することを推奨する情報を推奨情報 161 として出力する。

[0041] 出力装置は、推奨情報 161 を画面に表示する表示装置であってもよいし、推奨情報 161 を利用して何らかの処理を実行するコンピュータであってもよい。

[0042] 以下では、データ連携支援装置 100 の動作（即ち、本実施の形態に係るデータ連携支援方法）の例を説明する。

[0043] 図 2 は、連携元情報 111 の例を示す表である。

[0044] 図 2 において、連携元情報 111 は、前述したように、連携元テーブルを定義する情報である。連携元情報 111 は、DDL (Data Definition Language) 又は XML (eXtensible Markup Language) で記述されたファイルとして予め情報記憶部 101 に記憶される。あるいは、連携元情報 111 は、SQL 文の発行等によって連携元データベースから取得されて情報記憶部 101 に記憶される。連携元情報 111 には、テーブル所在情報 201、テーブル定義情報 202 が含まれる。

[0045] テーブル所在情報 201 には、インスタンス名、スキーマ名、テーブル名の情報が含まれる。例えば、テーブル所在情報 201 では、「発注システム

」というインスタンスの「SVR001」というスキーマに「店舗」テーブルがあることが定義されている。

[0046] テーブル定義情報202は、テーブル所在情報201で定義されている連携元テーブルごとに記憶される。テーブル定義情報202には、カラム名、データ型、主キーの情報（即ち、属性情報及び主キー情報）が含まれる。例えば、「発注システム」の「SVR001」にある「店舗」テーブルのテーブル定義情報202では、その「店舗」テーブルが「店舗ID」、「店舗名」、「住所」、「店長社員ID」というカラムを含み、「店舗ID」が単独で主キーになっていることが定義されている。

[0047] 図示していないが、連携先情報112も、図2に示した連携元情報111と同様の形式の情報である。

[0048] 図3は、データ連携支援装置100の動作の例を示すフローチャートである。図4は、後述する参照元情報211及び主キー定義情報212の例を示す表である。図5は、スキーママッチングの結果221の例を示す表である。図6は、参照関係情報151の例を示す表である。

[0049] ステップS11において、テーブル選択部102は、情報記憶部101から連携先情報112を読み取る。テーブル選択部102は、連携先情報112により定義された連携先テーブルの一覧を、出力装置を介して画面に表示する。テーブル選択部102は、この一覧を見たユーザの操作を、入力装置を介して受け付ける。テーブル選択部102は、ユーザの操作に従って、1つの連携先テーブルを選択する。テーブル選択部102は、テーブル定義情報202と同様の形式で、選択した連携先テーブルを定義する参照元情報211を生成する。

[0050] 例えば、テーブル選択部102は、「発注システム」の「SVR001」にある「注文明細テーブル」について、図4に示すような参照元情報211を生成する。この参照元情報211では、「発注システム」の「SVR001」にある「注文明細テーブル」が「注文詳細ID」、「注文ID」、「店舗ID」、「注文数」、「店舗名」というカラムを含み、「注文詳細ID」

及び「注文ID」が主キーになっていることが定義されている。

[0051] テーブル選択部102は、参照元情報211をテーブル抽出部104に引き渡す。

[0052] ストップS12において、コラム特定部103は、情報記憶部101から連携元情報111を読み取る。コラム特定部103は、連携元情報111に含まれるテーブル所在情報201で定義されている連携元テーブルごとに、

連携元情報111に含まれるテーブル定義情報202を参照する。コラム特定部103は、各連携元テーブルのテーブル定義情報202から、各連携元テーブルの主キーになっているテーブル(即ち、第1コラム)を特定する。コラム特定部103は、テーブル定義情報202を、特定したコラムのみに絞り込んだ主キー定義情報212を生成する。

[0053] 例えば、コラム特定部103は、「発注システム」の「SVR001」にある「店舗」テーブルについて、図4に示すような主キー定義情報212を生成する。この主キー定義情報212では、「発注システム」の「SVR001」にある「店舗」テーブルが主キーとして「店舗ID」というコラムを含むことが定義されている。

[0054] コラム特定部103は、各連携元テーブルについて、主キー定義情報212をテーブル抽出部104に引き渡す。

[0055] コラム特定部103は、連携先情報112についても同様の処理を行って、連携先テーブルごとに主キー定義情報212を生成する。ただし、コラム特定部103は、ストップS11で選択された連携先テーブルについては、主キー定義情報212を生成しない。

[0056] コラム特定部103は、各連携先テーブルについて、主キー定義情報212をテーブル抽出部104に引き渡す。

[0057] ストップS13において、テーブル抽出部104は、各連携元テーブルについて、テーブル選択部102から引き渡された参照元情報211と、コラム特定部103から引き渡された主キー定義情報212との組を設定する。

テーブル抽出部104は、各連携先テーブルについても、テーブル選択部1

０２から引き渡された参照元情報２１１と、カラム特定部１０３から引き渡された主キ一定義情報２１２との組を設定する。

[0058] ステップＳ１４において、テーブル抽出部１０４は、ステップＳ１３で設定した組のうち、ステップＳ１４～Ｓ１７の処理を未だ実行していない組を選択する。テーブル抽出部１０４は、選択した組の参照元情報２１１と主キ一定義情報２１２とのスキーママッチングを実行する。例えば、テーブル抽出部１０４は、図５に示すようなスキーママッチングの結果２２１を得る。図５では、スキーママッチングの結果２２１が、主キ一定義情報２１２のカラムと参照元情報２１１のカラムとの類似度を０～１．０の間で示した表の形式になっている。図５の例において、主キ一定義情報２１２の「店舗ＩＤ」カラムと参照元情報２１１の「注文数」カラムとの類似度は０である。一方、主キ一定義情報２１２の「店舗ＩＤ」カラムと参照元情報２１１の「店舗ＩＤ」カラムとの類似度は１．０である。

[0059] ステップＳ１５において、テーブル抽出部１０４は、ステップＳ１４で得たスキーママッチングの結果２２１から、類似度が予め定められた閾値より高いカラムの組（即ち、第１カラムと第３カラムとの組）を抽出する。主キ一定義情報２１２に複数のカラムがある場合、テーブル抽出部１０４は、主キ一定義情報２１２のカラムの全てについて、参照元情報２１１のカラムに重複がないように、カラムの組を抽出しなければならない。主キ一定義情報２１２の同じカラムに対して、類似度が閾値より高いカラムの組が２つ以上ある場合、テーブル抽出部１０４は、類似度が最も高いカラムの組を選択する。類似度が閾値より高いカラムの組が１つもない場合、テーブル抽出部１０４は、カラムの組を抽出しないでステップＳ１８に進む。図５の例において、閾値を０．８とすると、テーブル抽出部１０４は、主キ一定義情報２１２の「店舗ＩＤ」カラムと参照元情報２１１の「店舗ＩＤ」カラムとの組を抽出することになる。

[0060] ステップＳ１５で抽出されたカラムの組がある場合、ステップＳ１６において、テーブル抽出部１０４は、抽出したカラムの組を示す参照関係情報１

51を生成する。テーブル抽出部104は、参照関係情報151を情報蓄積部105に保存する。

[0061] 例えば、テーブル抽出部104は、図6に示すような参照関係情報151

を生成する。この参照関係情報151には、参照先インスタンス名、参照先
 スキーマ名、参照先テーブル名、参照先カラム名、参照元インスタンス名、
 参照元スキーマ名、参照元テーブル名、参照元カラム名の情報が含まれる。

この参照関係情報151では、「発注システム」の「SVR001」にある「店舗」テーブルと「発注システム」の「SVR001」にある「注文明細」テーブルとの間に参照関係があることが定義されている。また、「店舗」テーブルの「店舗ID」が「注文明細」テーブルの「注文明細ID」が定義されている（即ち、外部キー）であることが定義されている。

[0062] ストップS17において、データ連携支援装置100は、図7に示す動作を行う。

[0063] ステップS18において、テーブル抽出部104は、ステップS13で設定した組のうち、ステップS14～S17の処理を未だ実行していない組があるかどうかを判定する。そのような組がある場合、テーブル抽出部104は、ステップS14に戻る。そのような組がない場合、テーブル抽出部104は、処理を終了する。

[0064] 図7は、スレッドS17におけるデータ連携支援装置100の動作の例を示すフローチャートである。図8は、後述する非外部キ一定義情報231及び非主キ一定義情報232の例を示す表である。図9は、スキーマツチングの結果241の例を示す表である。図10は、連携関係情報152の例を示す表である。

[0065] ステップS21において、テーブル選択部102は、情報蓄積部105から参照関係情報151を読み取る。テーブル選択部102は、ステップS11で生成した参照元情報211を、参照関係情報151で定義されている参照元カラム以外のカラムのみに絞り込んだ非外部キ一定義情報231を生成

する。

[0066] 例えば、テーブル選択部１０２は、「発注システム」の「ＳＶＲ００１」にある「注文明細テーブル」について、図８に示すような非外部キー定義情報２３１を生成する。この非外部キー定義情報２３１では、「発注システム」の「ＳＶＲ００１」にある「注文明細テーブル」が非外部キーとして「注文詳細ＩＤ」、「注文ＩＤ」、「注文数」、「店舗名」というカラムを含み、「注文詳細ＩＤ」及び「注文ＩＤ」が主キーになっていることが定義されている。

[0067] テーブル選択部１０２は、非外部キー定義情報２３１をテーブル抽出部１０４に引き渡す。

[0068] ステップＳ２２において、カラム特定部１０３は、情報蓄積部１０５から参照関係情報１５１を読み取る。カラム特定部１０３は、参照関係情報１５１で定義されている参照先テーブルについて、ステップＳ１２で読み取った連携元情報１１１又は連携先情報１１２に含まれるテーブル定義情報２０２を参照する。カラム特定部１０３は、参照先テーブルのテーブル定義情報２０２から、参照先テーブルの主キーになっていないカラム（即ち、第２カラム）を特定する。カラム特定部１０３は、テーブル定義情報２０２を、特定したカラムのみに絞り込んだ非主キー定義情報２３２を生成する。

[0069] 例えば、カラム特定部１０３は、「発注システム」の「ＳＶＲ００１」にある「店舗」テーブルについて、図８に示すような非主キー定義情報２３２を生成する。この非主キー定義情報２３２では、「発注システム」の「ＳＶＲ００１」にある「店舗」テーブルが非主キーとして「店舗名」、「住所」、「店長社員ＩＤ」というカラムを含むことが定義されている。

[0070] カラム特定部１０３は、非主キー定義情報２３２をテーブル抽出部１０４に引き渡す。

[0071] ステップＳ２３において、テーブル抽出部１０４は、テーブル選択部１０２から引き渡された非外部キー定義情報２３１と、カラム特定部１０３から引き渡された非主キー定義情報２３２との組を設定する。

- [0072] ステップS 2 4において、テーブル抽出部1 0 4は、ステップS 2 3で設定した組の非外部キー定義情報2 3 1と非主キー定義情報2 3 2とのスキーママッチングを実行する。例えば、テーブル抽出部1 0 4は、図9に示すようなスキーママッチングの結果2 4 1を得る。図9では、スキーママッチングの結果2 4 1が、図5の例と同様に、非主キー定義情報2 3 2のカラムと非外部キー定義情報2 3 1のカラムとの類似度を0～1. 0の間で示した表の形式になっている。図9の例において、非主キー定義情報2 3 2の「店舗名」カラムと非外部キー定義情報2 3 1の「注文詳細ID」カラムとの類似度は0である。一方、非主キー定義情報2 3 2の「店舗名」カラムと非外部キー定義情報2 3 1の「店舗名」カラムとの類似度は1. 0である。
- [0073] ステップS 2 5において、テーブル抽出部1 0 4は、ステップS 2 4で得たスキーママッチングの結果2 4 1から、類似度が予め定められた閾値より高いカラムの組（即ち、第2カラムと第4カラムとの組）を抽出する。非主キー定義情報2 3 2に複数のカラムがある場合、テーブル抽出部1 0 4は、非外部キー定義情報2 3 1のカラムに重複がないように、カラムの組を抽出しなければならない。非主キー定義情報2 3 2の同じカラムに対して、類似度が閾値より高いカラムの組が2つ以上ある場合、テーブル抽出部1 0 4は、類似度が最も高いカラムの組を選択する。類似度が閾値より高いカラムの組が1つもない場合、テーブル抽出部1 0 4は、カラムの組を抽出しないで処理を終了する。図9の例において、閾値を0. 8とすると、テーブル抽出部1 0 4は、非主キー定義情報2 3 2の「店舗名」カラムと非外部キー定義情報2 3 1の「店舗名」カラムとの組を抽出することになる。テーブル抽出部1 0 4は、非主キー定義情報2 3 2の「住所」カラム及び「店長社員ID」カラムについては、カラムの組を抽出しない。
- [0074] ステップS 2 5で抽出されたカラムの組がある場合、ステップS 2 6において、テーブル抽出部1 0 4は、抽出したカラムの組を示す連携関係情報1 5 2を生成する。テーブル抽出部1 0 4は、連携関係情報1 5 2を情報蓄積部1 0 5に保存する。

[0075] 例えば、テーブル抽出部 104 は、図 10 に示すような連携関係情報 152 を生成する。この連携関係情報 152 には、連携元カラム名、連携先カラム名の情報が含まれる。連携関係情報 152 は、図 6 に示した参照関係情報 151 と合成される。即ち、参照関係情報 151 及び連携関係情報 152 が合成された情報には、参照先インスタンス名、参照先スキーマ名、参照先テーブル名、参照先カラム名、連携元カラム名、参照元インスタンス名、参照元スキーマ名、参照元テーブル名、参照元カラム名、連携先カラム名の情報が含まれる。この合成された情報では、「発注システム」の「SVR001」にある「店舗」テーブルの「店舗名」カラムが連携元であり、「発注システム」の「SVR001」にある「注文明細」テーブルの「店舗名」カラムが連携先であることが定義されている。連携元と連携先との組が 2 つ以上ある場合、連携元カラム名、連携先カラム名が同じ順番で記憶される。例えば、「店舗」テーブルの「店舗名」カラムと「注文明細」テーブルの「店舗名」カラムとが組であり、「店舗」テーブルの「住所」カラムと「注文明細」テーブルの「住所」カラムとが組であるとする。連携元カラム名が「店舗名」、「住所」の順番で記憶されていれば、連携先カラム名も「店舗名」、「住所」の順番で記憶される。

[0076] 図 11 は、図 3 の動作の後におけるデータ連携支援装置 100 の動作の例を示すフローチャートである。図 12 は、抽出結果 141 の例を示す表である。

[0077] ステップ S31 において、テーブル抽出部 104 は、情報蓄積部 105 から参照関係情報 151 及び連携関係情報 152 を読み取る。テーブル抽出部 104 は、参照関係情報 151 及び連携関係情報 152 により示された参照先カラムと参照元カラムとの組（即ち、第 1 カラムと第 3 カラムとの組）を、参照先テーブルと参照元テーブルとがどのスキーマに属するかによってグループ分けする。具体的には、テーブル抽出部 104 は、参照先カラムと参照元カラムとの組を、参照先テーブルと参照元テーブルとが同じスキーマに属するグループと、参照先テーブルと参照元テーブルとが同じインスタンス

の異なるスキーマに属するグループと、参照先テーブルと参照元テーブルとが異なるインスタンスのスキーマに属するグループとの3つのグループに分類する。

[0078] ステップS 3 2において、テーブル抽出部1 0 4は、前述した3つのグループのうち、ステップS 3 2及びS 3 3の処理を未だ実行していないグループを選択する。テーブル抽出部1 0 4は、ステップS 3 1で、選択したグループに分類した組のうち、対応する連携元カラムと連携先カラムとの組の個数が最も多いものを抽出する。

[0079] ステップS 3 3において、テーブル抽出部1 0 4は、参照関係情報1 5 1及び連携関係情報1 5 2を合成した情報と同様の形式で、ステップS 3 2で抽出した参照先カラムと参照元カラムとの組を示す抽出結果1 4 1を出力する。

[0080] 例えば、テーブル抽出部1 0 4は、図1 2に示すような抽出結果1 4 1を出力する。この抽出結果1 4 1では、「発注システム」の「SVR 0 0 1」にある「店舗」テーブルと「発注システム」の「SVR 0 0 1」にある「注文明細」テーブルとの間に参照関係があることが示されている。また、「店舗」テーブルの「店舗ID」カラムが参照先、「店舗名」が連携元であり、「注文明細」テーブルの「店舗ID」カラムが参照元（即ち、外部キー）、「店舗名」が連携先であることが示されている。

[0081] ステップS 3 4において、テーブル抽出部1 0 4は、前述した3つのグループのうち、ステップS 3 2及びS 3 3の処理を未だ実行していないグループがあるかどうかを判定する。そのようなグループがある場合、テーブル抽出部1 0 4は、ステップS 3 2に戻る。そのようなグループがない場合、テーブル抽出部1 0 4は、処理を終了する。

[0082] 図1 3は、図1 1の動作の後におけるデータ連携支援装置1 0 0の動作の例を示すフローチャートである。図1 4は、後述する入力情報2 5 1及び推奨情報1 6 1の例を示す表である。

[0083] ステップS 4 1において、情報出力部1 0 6は、参照先を検索するために

ユーザから入力される入力情報 251 を、入力装置を介して取得する。

[0084] 例えば、情報出力部 106 は、図 14 に示すような入力情報 251 を取得する。この入力情報 251 では、「発注システム」の「SVR001」にある「注文明細」テーブルの「店舗ID」カラムに対して、ユーザが参照先を検索したいことが示されている。カラム名が省略されている場合、情報出力部 106 は、「注文明細」テーブルの全てのカラムに対して、ユーザが参照先を検索したいことが示されていると解釈してもよい。

[0085] ステップ S42 において、情報出力部 106 は、ステップ S33 でテーブル抽出部 104 から出力された抽出結果 141 を参照する。情報出力部 106 は、抽出結果 141 から、ステップ S41 で取得した入力情報 251 に一致する情報を抽出する。具体的には、情報出力部 106 は、抽出結果 141 で示されている参照元インスタンス、参照元スキーマ、参照元テーブル、参照元カラムの組み合わせのうち、入力情報 251 で示されているインスタンス、スキーマ、テーブル、カラムの組み合わせに一致するものを抽出する。

[0086] ステップ S43 において、情報出力部 106 は、ステップ S42 で抽出した情報があるかどうかを判定する。抽出した情報がある場合、情報出力部 106 は、ステップ S44 に進む。抽出した情報がない場合、情報出力部 106 は、ステップ S45 に進む。

[0087] ステップ S44 において、情報出力部 106 は、ステップ S42 で抽出した情報に基づき、前述した 3 つのグループのそれぞれについて、入力情報 251 に対応する参照先を推奨する推奨情報 161 を生成する。情報出力部 106 は、推奨情報 161 を、出力装置を介して出力し、処理を終了する。

[0088] 例えば、情報出力部 106 は、図 14 に示すような推奨情報 161 を画面に表示する。この推奨情報 161 では、参照先テーブルと参照元テーブルとが同じスキーマに属するグループ（即ち、「同一スキーマ」）について、「発注システム」の「SVR001」にある「店舗」テーブルの「店舗ID」カラムを参照先として推奨することが示されている。また、入力情報 251 で示されている「注文明細」テーブルの「店舗名」カラムに格納するデータ

を、「発注システム」の「SVR001」にある「店舗」テーブルの「店舗名」カラムから抽出すればよいことが示されている。

[0089] ステップS45において、情報出力部106は、入力情報251に対応する参照先がないことを通知する情報を、出力装置を介して出力し、処理を終了する。

[0090] 本実施の形態では、以上説明した動作により、データベースの定義情報のみから、テーブル間の参照関係及び外部キーとなるカラムを特定することができる。さらに、参照関係のあるテーブル間で共通するカラムを特定し、データ連携に利用するのに最適なテーブル（即ち、データ連携に利用できるカラムの数が最も多いテーブル）を抽出することができる。つまり、本実施の形態によれば、データ連携において、連携先となるテーブルと、どのテーブルとの間に参照関係があるかを、効率よく検出することができる。さらに、連携先となるテーブルのカラムに格納するデータを、どのテーブルのカラムから抽出すればよいかを、効率よく検出することができる。

[0091] 前述したように、本実施の形態では、テーブル抽出部104が、カラム特定部103により2つ以上の第2カラムが特定されたテーブルについては、テーブル選択部102により選択された連携先テーブルに、最も多くの第2カラムに対して個別に第4カラムが存在するテーブルのみを抽出結果141に含める。本実施の形態の変形例において、テーブル抽出部104は、カラム特定部103により2つ以上の第2カラムが特定されたテーブルについて、テーブル選択部102により選択された連携先テーブルに、2番目に多くの第2カラムに対して個別に第4カラムが存在するテーブルも抽出結果141に含めることができる。あるいは、「2番目」だけでなく、「上位n番目」まで含めることができる。「n」には、2より大きい任意の数を設定可能である。

[0092] 図15は、実施の形態1に係るデータ連携支援装置100のハードウェア構成の例を示す図である。

[0093] 図15において、データ連携支援装置100は、コンピュータであり、L

CD901 (Liquid・Crystal・Display)、キーボード902 (K/B)、マウス903、FDD904 (Flexible・Disk・Drive)、CDD905 (Compact・Disc・Drive)、プリンタ906といったハードウェアデバイスを備えている。これらのハードウェアデバイスはケーブルや信号線で接続されている。LCD901の代わりに、CRT (Cathode・Ray・Tube)、あるいは、その他の表示装置が用いられてもよい。マウス903の代わりに、タッチパネル、タッチパッド、トラックボール、ペンタブレット、あるいは、その他のポインティングデバイスが用いられてもよい。

[0094] データ連携支援装置100は、プログラムを実行するCPU911 (Central・Processing・Unit) を備えている。CPU911は、処理装置の例である。CPU911は、バス912を介してROM913 (Read・Only・Memory)、RAM914 (Random・Access・Memory)、通信ボード915、LCD901、キーボード902、マウス903、FDD904、CDD905、プリンタ906、HDD920 (Hard・Disk・Drive) と接続され、これらのハードウェアデバイスを制御する。HDD920の代わりに、SSD (Solid・State・Drive)、フラッシュメモリ、光ディスク装置、メモリカードリーダーライター、あるいは、その他の記録媒体が用いられてもよい。

[0095] RAM914は、揮発性メモリの一例である。ROM913、FDD904、CDD905、HDD920は、不揮発性メモリの一例である。これらは、記憶装置の例である。通信ボード915、キーボード902、マウス903、FDD904、CDD905は、入力装置の一例である。また、通信ボード915、LCD901、プリンタ906は、出力装置の一例である。

[0096] 通信ボード915は、LAN (Local・Area・Network) 等に接続されている。通信ボード915は、LANに限らず、IP-VPN (Internet・Protocol・Virtual・Private

・Network)、広域LAN、ATM (Asynchronous・Transfer・Mode) ネットワークといったWAN (Wide・Area・Network)、あるいは、インターネットに接続されていても構わない。LAN、WAN、インターネットは、ネットワークの例である。

[0097] HDD 920には、オペレーティングシステム921 (OS)、ウィンドウシステム922、プログラム群923、ファイル群924が記憶されている。プログラム群923のプログラムは、CPU 911、オペレーティングシステム921、ウィンドウシステム922により実行される。プログラム群923には、本発明の実施の形態の説明において「～部」として説明する機能を実行するプログラムが含まれている。プログラムは、CPU 911により読み出され実行される。ファイル群924には、本発明の実施の形態の説明において、「～データ」、「～情報」、「～ID (識別子)」、「～フラグ」、「～結果」として説明するデータや情報や信号値や変数値やパラメータが、「～ファイル」や「～データベース」や「～テーブル」の各項目として含まれている。「～ファイル」や「～データベース」や「～テーブル」は、RAM 914やHDD 920等の記録媒体に記憶される。RAM 914やHDD 920等の記録媒体に記憶されたデータや情報や信号値や変数値やパラメータは、読み書き回路を介してCPU 911によりメインメモリやキャッシュメモリに読み出され、抽出、検索、参照、比較、演算、計算、制御、出力、印刷、表示といったCPU 911の処理 (動作) に用いられる。抽出、検索、参照、比較、演算、計算、制御、出力、印刷、表示といったCPU 911の処理中、データや情報や信号値や変数値やパラメータは、メインメモリやキャッシュメモリやバッファメモリに一時的に記憶される。

[0098] 本発明の実施の形態の説明において用いるブロック図やフローチャートの矢印の部分は主としてデータや信号の入出力を示す。データや信号は、RAM 914等のメモリ、FDD 904のフレキシブルディスク (FD)、CD 905のコンパクトディスク (CD)、HDD 920の磁気ディスク、光ディスク、DVD (Digital・Versatile・Disc)、あ

るいは、その他の記録媒体に記録される。また、データや信号は、バス 9 1 2、信号線、ケーブル、あるいは、その他の伝送媒体により伝送される。

[0099] 本発明の実施の形態の説明において「～部」として説明するものは、「～回路」、「～装置」、「～機器」であってもよく、また、「～ステップ」、「～工程」、「～手順」、「～処理」であってもよい。即ち、「～部」として説明するものは、ROM 9 1 3 に記憶されたファームウェアで実現されていても構わない。あるいは、「～部」として説明するものは、ソフトウェアのみ、あるいは、素子、デバイス、基板、配線といったハードウェアのみで実現されていても構わない。あるいは、「～部」として説明するものは、ソフトウェアとハードウェアとの組み合わせ、あるいは、ソフトウェアとハードウェアとファームウェアとの組み合わせで実現されていても構わない。ファームウェアとソフトウェアは、プログラムとして、フレキシブルディスク、コンパクトディスク、磁気ディスク、光ディスク、DVD等の記録媒体に記憶される。プログラムはCPU 9 1 1 により読み出され、CPU 9 1 1 により実行される。即ち、プログラムは、本発明の実施の形態の説明で述べる「～部」としてコンピュータを機能させるものである。あるいは、プログラムは、本発明の実施の形態の説明で述べる「～部」の手順や方法をコンピュータに実行させるものである。

[0100] 以上、本発明の実施の形態について説明したが、本発明は、この実施の形態に限定されるものではなく、必要に応じて種々の変更が可能である。

符号の説明

[0101] 1 0 0 データ連携支援装置、1 0 1 情報記憶部、1 0 2 テーブル選択部、1 0 3 カラム特定部、1 0 4 テーブル抽出部、1 0 5 情報蓄積部、1 0 6 情報出力部、1 1 1 連携元情報、1 1 2 連携先情報、1 4 1 抽出結果、1 5 1 参照関係情報、1 5 2 連携関係情報、1 6 1 推奨情報、2 0 1 テーブル所在情報、2 0 2 テーブル定義情報、2 1 1 参照元情報、2 1 2 主キー定義情報、2 2 1 結果、2 3 1 非外部キー定義情報、2 3 2 非主キー定義情報、2 4 1 結果、2 5 1 入力情報、

901 LCD、902 キーボード、903 マウス、904 FDD、
905 CDD、906 プリンタ、911 CPU、912 バス、91
3 ROM、914 RAM、915 通信ボード、920 HDD、92
1 オペレーティングシステム、922 ウィンドウシステム、923 プ
ログラム群、924 ファイル群。

請求の範囲

[請求項1] 複数のテーブルを有するデータベースのデータ連携を支援するデータ連携支援装置において、

前記複数のテーブルのそれぞれに含まれる複数のカラムのそれぞれが主キーになっているかどうかを示す主キー情報と、前記複数のテーブルのそれぞれに含まれる複数のカラムのそれぞれに格納されるデータの属性を示す属性情報とを記憶する情報記憶部と、

前記複数のテーブルのうち、データ連携の連携先となる1つのテーブルを選択するテーブル選択部と、

前記複数のテーブルのうち前記テーブル選択部により選択されたテーブル以外のテーブルのそれぞれについて、前記情報記憶部に記憶された主キー情報に基づき、該当するテーブルに含まれる複数のカラムから、主キーになっている第1カラムと、主キーになっていない第2カラムとを特定するカラム特定部と、

前記複数のテーブルのうち前記テーブル選択部により選択されたテーブル以外のテーブルから、前記情報記憶部に記憶された属性情報に基づき、前記テーブル選択部により選択されたテーブルに、前記カラム特定部により特定された第1カラムと類似する第3カラムが含まれ、かつ、前記テーブル選択部により選択されたテーブルに、前記カラム特定部により特定された第2カラムと類似する第4カラムが含まれるテーブルを抽出するテーブル抽出部と、

前記テーブル選択部により選択されたテーブルと前記テーブル抽出部により抽出されたテーブルとの組み合わせをデータ連携に利用することを推奨する推奨情報を出力する情報出力部とを備えることを特徴とするデータ連携支援装置。

[請求項2] 前記テーブル抽出部は、前記カラム特定部により2つ以上の第1カラムが特定されたテーブルについては、前記テーブル選択部により選択されたテーブルに、全ての第1カラムに対して個別に前記第3カラム

ムが存在するテーブルを抽出することを特徴とする請求項1のデータ連携支援装置。

[請求項3] 前記テーブル抽出部は、前記カラム特定部により2つ以上の第2カラムが特定されたテーブルについては、前記テーブル選択部により選択されたテーブルに、最も多くの第2カラムに対して個別に前記第4カラムが存在するテーブルを抽出することを特徴とする請求項1のデータ連携支援装置。

[請求項4] 前記テーブル抽出部は、前記テーブル選択部により選択されたテーブルと同じスキーマに属するテーブルと、前記テーブル選択部により選択されたテーブルと同じインスタンスの異なるスキーマに属するテーブルと、前記テーブル選択部により選択されたテーブルと異なるインスタンスのスキーマに属するテーブルとの3種類のテーブルのうち少なくとも2種類のテーブルを抽出し、

前記情報出力部は、前記テーブル抽出部により抽出されたテーブルの種類ごとに、前記テーブル選択部により選択されたテーブルと前記テーブル抽出部により抽出されたテーブルとの組み合わせをデータ連携に利用することを推奨する情報を前記推奨情報として出力することを特徴とする請求項1のデータ連携支援装置。

[請求項5] 複数のテーブルを有するデータベースのデータ連携を支援するデータ連携支援方法において、

前記複数のテーブルのそれぞれに含まれる複数のカラムのそれぞれが主キーになっているかどうかを示す主キー情報と、前記複数のテーブルのそれぞれに含まれる複数のカラムのそれぞれに格納されるデータの属性を示す属性情報とを記憶する情報記憶部を備えるコンピュータを用いて、

前記複数のテーブルのうち、データ連携の連携先となる1つのテーブルを選択し、

前記複数のテーブルのうち前記選択により選択されたテーブル以外

のテーブルのそれぞれについて、前記情報記憶部に記憶された主キー情報に基づき、該当するテーブルに含まれる複数のカラムから、主キーになっている第1カラムと、主キーになっていない第2カラムとを特定し、

前記複数のテーブルのうち前記選択により選択されたテーブル以外のテーブルから、前記情報記憶部に記憶された属性情報に基づき、前記選択により選択されたテーブルに、前記特定により特定された第1カラムと類似する第3カラムが含まれ、かつ、前記選択により選択されたテーブルに、前記特定により特定された第2カラムと類似する第4カラムが含まれるテーブルを抽出し、

前記選択により選択されたテーブルと前記抽出により抽出されたテーブルとの組み合わせをデータ連携に利用することを推奨する推奨情報を出力することを特徴とするデータ連携支援方法。