

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2012年3月22日(22.03.2012)

PCT

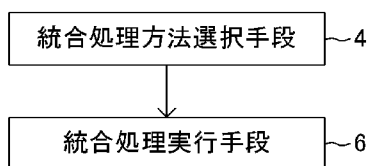
(10) 国際公開番号  
WO 2012/035754 A1

- (51) 国際特許分類:  
G06F 17/30 (2006.01)
  - (21) 国際出願番号: PCT/JP2011/005129
  - (22) 国際出願日: 2011年9月13日(13.09.2011)
  - (25) 国際出願の言語: 日本語
  - (26) 国際公開の言語: 日本語
  - (30) 優先権データ:  
特願 2010-204210 2010年9月13日(13.09.2010) JP
  - (71) 出願人 (米国を除く全ての指定国について): 日本電気株式会社 (NEC Corporation) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 Tokyo (JP).
  - (72) 発明者: および
  - (75) 発明者/出願人 (米国についてのみ): 有熊 威 (ARIKUMA, Takeshi) [JP/JP]; 〒1088001 東京都港区芝五丁目7番1号 日本電気株式会社内 Tokyo (JP).
  - (74) 代理人: 馬場 資博, 外 (BABA, Motohiro et al.); 〒3500046 埼玉県川越市菅原町25番地1 石井ビル2階 Saitama (JP).
  - (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
  - (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- 添付公開書類:  
— 国際調査報告 (条約第21条(3))

(54) Title: DATA INTEGRATION PROCESSING DEVICE, SYSTEM, METHOD AND PROGRAM

(54) 発明の名称: データ統合処理装置、システム、方法及びプログラム

[図16]



4... INTEGRATION PROCESSING METHOD SELECTING MEANS

6... INTEGRATION PROCESSING EXECUTING MEANS

(57) Abstract: A data integration processing device for integrating a plurality of graphs comprises an integration processing method selecting means (4) for selecting an integration processing method used for the integration of a group of inputted graphs; and an integration processing executing means (6) which has a plurality of integration processing methods and integrates the plurality of graphs by executing integration processing in accordance with the integration processing method the integration processing method selecting means (4) selected from the plurality of integration processing methods; wherein the integration processing method selecting means (4) selects an integration processing method used for integrating lower-level nodes depending on the frequency of matching of a lower-level node upon matching an upper-level node with respect to each node of the inputted graphs.

(57) 要約: 複数のグラフを統合するデータ統合処理装置であって、入力されたグラフ群の統合に用いる統合処理方法を選択する統合処理方法選択手段4と、統合処理方法を複数有し、複数の統合処理方法のうち統合処理方法選択手段4が選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合する統合処理実行手段6とを備え、統合処理方法選択手段4は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、下位のノードを統合するために用いる統合処理方法を選択する。



WO 2012/035754 A1

## 明 細 書

発明の名称：

データ統合処理装置、システム、方法及びプログラム

### 技術分野

[0001] 本発明は、複数のグラフを統合するデータ統合処理装置、データ統合処理システム、データ統合処理方法及びデータ統合処理プログラムに関する。

### 背景技術

[0002] データ統合技術として、特許文献1や非特許文献1に記載されたシステムが知られている。

[0003] まず、特許文献1に記載された有向グラフの単一化装置は、図15に示すように、表現手段13と、併合手段14と、タグチェック手段15とを備えている。そのような構成の有向グラフの単一化装置は、次のようにして有向グラフを統合する。

[0004] まず、表現手段13は、入力された有向グラフを、タグとそれに対応する部分有向グラフとのペアのリストとして表現する。このリストのことをタグリストと呼ぶ。次に、併合手段14は、2つの有向グラフに対応するタグリストを併合する。

[0005] 次に、タグチェック手段15は、併合されたタグリストにおける同じ名のタグに対応する部分有向グラフが同一であることをチェックする。

[0006] また、非特許文献1に記載されたグラフ統合アルゴリズムでは、マップ関数として外部で定義された2つのグラフ間での同一ノード判定ルールを用いて、下記のアルゴリズムでグラフの一致を判定しながら統合を実施する。

[0007] ・マップMはブランクノードをブランクノードへ写像する。

[0008] ・グラフGのノードであるすべての具体的な値（RDF リテラル） $l_i t$ に対して $M(l_i t) = l_i t$ とする。

[0009] ・Gのノードであるすべての中間ノード（RDF URI参照） $u r i$ に対して $M(u r i) = u r i$ とする。

[0010] ・ノードとノードとがエッジで繋がれている事を示すトリプル (s, p, o) がG中であるならば、そのときに限ってかならずトリプル (M (s), p, M (o)) はG' 中にある。

[0011] 上記アルゴリズムによりトリプル単位での一致・不一致を基にしてグラフの統合処理を行う。

## 先行技術文献

## 特許文献

[0012] 特許文献1：特開平5－204647号公報

## 非特許文献

[0013] 非特許文献1：Graham Klyne et, al., "RDF Concepts and Abstract Syntax", W3C Recommendation, <http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>, (2004)

## 発明の概要

## 発明が解決しようとする課題

[0014] 上述した技術における問題点は、統合機能とスループットとが、トレードオフの関係にあることである。既存の統合手段では、特許文献1に記載された方法のように提供する統合機能を単純化し、[グラフ中のノード数]<sup>2</sup>で処理を実現している。

[0015] 一方、非特許文献1に記載された方法のようにノードの周辺情報やマップに登録された同義語情報を元にトリプルの同一性を判定するような高度な統合機能を提供する手法の場合には、処理速度が（[グラフ中のノード数] × [ノードあたりの平均プロパティ数]）<sup>2</sup>の処理時間が必要となる。このように、既存の技術では、統合処理とスループットとを両立できないという問題がある。

[0016] そこで、本発明は、提供する統合機能への制限を低く抑えながら、高いスループットを実現することができるデータ統合処理装置、データ統合処理システム、データ統合処理方法及びデータ統合処理プログラムを提供すること

を目的とする。

### 課題を解決するための手段

- [0017] 本発明によるデータ統合処理装置は、複数のグラフを統合するデータ統合処理装置であって、入力されたグラフ群の統合に用いる統合処理方法を選択する統合処理方法選択手段と、統合処理方法を複数有し、複数の統合処理方法のうちの統合処理方法選択手段が選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合する統合処理実行手段とを備え、統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、下位のノードを統合するために用いる統合処理方法を選択することを特徴とする。
- [0018] 本発明によるデータ統合処理システムは、複数のグラフを統合するデータ統合処理システムであって、入力されたグラフ群の統合に用いる統合処理方法を選択する統合処理方法選択手段と、統合処理方法を複数有し、複数の統合処理方法のうちの統合処理方法選択手段が選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合する統合処理実行手段とを含み、統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、下位のノードを統合するために用いる統合処理方法を選択することを特徴とする。
- [0019] 本発明によるデータ統合処理方法は、複数のグラフを統合するデータ統合処理方法であって、入力されたグラフ群の統合に用いる統合処理方法を選択し、統合処理方法を複数有し、複数の統合処理方法のうちの選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合し、統合処理方法を選択する際には、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、下位のノードを統合するために用いる統合処理方法を選択することを特徴とする。
- [0020] 本発明によるデータ統合処理プログラムは、複数のグラフを統合するためのデータ統合処理プログラムであって、コンピュータに、入力されたグラフ

群の統合に用いる統合処理方法を選択する統合処理方法選択処理と、統合処理方法を複数有し、複数の統合処理方法のうち選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合する統合処理実行処理とを実行させ、統合処理方法選択処理で、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、下位のノードを統合するために用いる処理統合処理方法を選択する処理を実行させることを特徴とする。

### 発明の効果

[0021] 本発明によれば、提供する統合機能への制限を低く抑えながら、高いスループットを実現することができる。

### 図面の簡単な説明

[0022] [図1]本発明によるデータ統合処理システムの第1の実施形態の全体構成の一例を示す機能ブロック図である。

[図2]データ統合処理システムに含まれる解析手段別特性記憶手段が格納する一致重複頻度テーブルのデータ構造の一例を示す概念図である。

[図3]データ統合処理システムに含まれる解析手段別特性記憶手段が格納する矛盾重複頻度テーブルのデータ構造の一例を示す概念図である。

[図4]データ統合処理システムに含まれる解析手段別特性記憶手段が格納するクラスプロパティ出現頻度テーブルのデータ構造の一例を示す概念図である。

[図5]データ統合処理システムに含まれる統合処理方法選択ルール記憶手段記憶手段が格納する選択ルールテーブルのデータ構造の一例を示す概念図である。

[図6]データ統合処理システムが実行するデータ統合処理の流れの一例を示すフローチャートである。

[図7]データ統合処理システムが実行する統合処理選択処理の流れの一例を示すフローチャートである。

[図8]データ統合処理システムが実行する特性学習処理の流れの一例を示すフ

ローチャートである。

[図9]データ統合処理システムの第2の実施形態の全体構成の一例を示す機能ブロック図である。

[図10]データ統合処理システムが実行するデータ統合処理の流れの一例を示すフローチャートである。

[図11]データ統合処理システムが実行するグラフ分割処理の流れの一例を示すフローチャートである。

[図12]本発明の実施例における顧客情報検索エンジンが出力するモデルの一例を示す概念図である。

[図13]社員情報検索エンジンが出力するモデルの一例を示す概念図である。

[図14]人物動線検索エンジンが出力するモデルの一例を示す概念図である。

[図15]第1の関連技術として、特許文献1に記載のデータ統合処理システムの構成を概略示すブロック図である。

[図16]データ統合処理装置の最小の構成例を示す機能ブロック図である。

### 発明を実施するための形態

#### [0023] 実施形態1.

次に、本発明の第1の実施形態について図面を参照して詳細に説明する。

図1は、本発明によるデータ統合処理システムの第1の実施形態の全体構成の一例を示す機能ブロック図である。図1に示すように、データ統合処理システムは、データ統合処理装置1と解析処理装置2とを含む。なお、本実施形態では、データ統合処理装置1と解析処理装置2とが異なる装置として構成されている例について説明するが、これに限らず、単一の装置によって構成されていてもよい。

[0024] 解析処理装置2は、具体的には、プログラムに従って動作するパーソナルコンピュータ等の情報処理装置によって実現される。解析処理装置2は、データを解析する複数の解析手段（図示せず）を備えている。

[0025] データ統合処理装置1は、具体的には、プログラムに従って動作するパーソナルコンピュータ等の情報処理装置によって実現される。データ統合処理

装置 1 は、解析手段別特性記憶手段 5 と、統合処理方法選択ルール記憶手段 8 と、統合処理方法選択手段 4 と、統合処理実行手段 6 と、特性学習手段 7 とを備えている。なお、各手段は、統合制御手段（図示せず）によって制御されているものとする。

[0026] 解析手段別特性記憶手段 5 は、解析処理装置 2 が備えている解析手段ごとに、解析結果を表現したグラフ内の部分グラフの特性情報を格納する。解析手段別特性記憶手段 5 は、具体的には、光ディスク装置や磁気ディスク装置等の記憶装置によって実現される。

[0027] 統合処理方法選択ルール記憶手段 8 は、グラフデータに最適な統合処理方法を選択するためのルールを示すルール情報（例えば、選択ルールテーブル 810 等）を格納する。統合処理方法選択ルール記憶手段 8 は、具体的には、光ディスク装置や磁気ディスク装置等の記憶装置によって実現される。

[0028] 統合処理方法選択手段 4 は、解析処理装置 2 から解析結果を表現したグラフ群とグラフ群を出力した解析手段群の情報とを受け取り、解析手段別特性記憶手段 5 が格納する特性情報と、統合処理方法選択ルール記憶手段 8 が格納するルール情報とに基づいて、適切な統合処理方法を選択する機能を備えている。統合処理方法選択手段 4 は、具体的には、プログラムに従って動作する情報処理装置の CPU によって実現される。

[0029] 統合処理実行手段 6 は、統合処理方法選択手段 4 が選択した統合処理方法に従って統合処理を実行することでグラフの統合を実施し、結果を解析処理装置 2 に送信する機能を備えている。統合処理実行手段 6 は、具体的には、プログラムに従って動作する情報処理装置の CPU によって実現される。

[0030] また、統合処理実行手段 6 は、それぞれ異なる方法でグラフの統合処理を実行する第 1 の統合処理方法実行手段 9、第 2 の統合処理方法実行手段 10 および第 3 の統合処理方法実行手段 11 を含む。本実施形態では、3 つの異なる統合処理方法実行手段を含む例について説明するが、これに限らず、2 つ以上であればいくつでもよい。また、各統合処理方法実行手段は、例えば情報処理装置の CPU が既存のグラフ統合アルゴリズムに基づいて処理を実

行することによって実現される。

[0031] 特性学習手段7は、統合処理方法選択手段4からグラフ群とグラフ群を出力した解析手段群の情報とを受け取り、解析手段別特性記憶手段5が記憶する情報を更新する機能を備えている。特性学習手段7は、具体的には、プログラムに従って動作する情報処理装置のCPUによって実現される。

[0032] 解析手段別特性記憶手段5は、一致重複頻度テーブル510と、矛盾重複頻度テーブル520と、クラスプロパティ出現頻度テーブル530とを記憶している。

[0033] 一致重複頻度テーブル510は、図2に示すように、解析処理装置2が備えている解析手段のIDを縦軸、横軸に持つ。ここで、解析手段IDの表現方法については、数字に限定する必要はなく、任意の文字列やURIなど、解析手段を一意に特定できる表現であれば、任意の表現を用いることができる。

[0034] 一致重複頻度テーブル510内の各セルは、セルの縦軸、横軸の解析手段IDに対応する2つの解析手段がそれぞれ出力する解析結果において一致重複が発生する箇所とその頻度とを格納する。つまり、2つの解析手段がそれぞれ出力するグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する箇所とその頻度とを格納する。なお、ここでは、2つのグラフについて、解析結果の要素データが重複している場合に、要素データを構成する情報が一致していることを一致重複と呼ぶ。一致重複が発生する箇所の表現には、処理対象となるグラフデータのノードのクラスや、グラフデータのエッジのラベルであるプロパティ名を用いることができる。

[0035] 矛盾重複頻度テーブル520は、図3に示すように、解析処理装置2が備えている解析手段のIDを縦軸、横軸に持つ。ここで、解析手段IDの表現方法については、数字に限定する必要はなく、任意の文字列やURIなど、解析手段を一意に特定できる表現であれば、任意の表現を用いることができる。

[0036] 矛盾重複頻度テーブル520内の各セルは、セルの縦軸、横軸の解析手段

IDに対応する2つの解析手段がそれぞれ出力する解析結果において矛盾重複が発生する箇所とその頻度とを格納する。つまり、2つの解析手段がそれぞれ出力するグラフ中の各ノードについて、上位のノードが一致するが、下位のノードが一致しない箇所とその頻度とを格納する。なお、ここでは、2つのグラフについて、解析結果の要素データが重複している場合に、要素データを構成する情報が異なっていることを矛盾重複と呼ぶ。矛盾重複が発生する箇所の表現には、処理対象となるグラフデータのノードのクラスや、グラフデータのエッジのラベルであるプロパティ名を用いることができる。

[0037] クラスプロパティ出現頻度テーブル530は、図4に示すように、解析手段ID531と、クラス／プロパティID532と、頻度533とを格納する。

[0038] 解析手段ID531は、一致重複管理テーブル510や矛盾重複頻度テーブル520と同様に、解析処理装置2が備えている解析手段を一意に特定するために用いられる。

[0039] また、クラス／プロパティID532は、解析結果のグラフ中のデータに含まれるクラスやプロパティを一意に特定するために用いられる。ここで、クラス／プロパティID532の表現方法については、英文字に限定する必要はなく、任意の文字列やURIなど、クラスやプロパティを一意に特定できる表現であれば、任意の表現を用いることができる。

[0040] 頻度533は、解析手段ID531で特定される解析手段が解析結果として出力したグラフ中に含まれるクラスまたはプロパティ全体を母数として、クラス／プロパティID532で特定されるクラスまたはプロパティの出現頻度を算出した値である。ここで、頻度533の表現方法については、百分率に限定する必要はなく、任意の数値表現を用いることができる。

[0041] 統合処理方法選択ルール記憶手段8は、選択ルールテーブル810を記憶している。選択ルールテーブル810は、図5に示すように、選択ルールID811と、ルール812と、統合処理方法ID813とを格納する。

[0042] 選択ルールID811は、選択ルールを一意に識別するためのIDであり

、選択ルールテーブル 810 の主キーである。ここで、選択ルール ID 811 の表現方法については、数字に限定する必要はなく、任意の文字列や URL など、選択ルールを一意に特定できる表現であれば、任意の表現を用いることができる。

[0043] ルール 812 は、統合処理方法を選択するために用いられる。ルール 812 が含む条件に入力データが合致する場合には、使用する統合処理方法をルール 812 に対応付けられた統合処理方法 ID 813 で指定する。

[0044] ルール 812 には、例えば「一致重複頻度が高く、矛盾重複頻度が低いものは低機能で高速な統合処理方法を選択する。」や、「一致重複頻度が低く、矛盾重複頻度が高いものは高機能な統合処理方法を選択する。」といったルールが記述される。ルールの他の例としては、「一致重複頻度に基づく値が所定値よりも高い場合は、低機能で高速な統合処理方法を選択する。」や、「一致重複頻度に基づく値が所定値よりも低い場合は、高機能で低速な統合処理方法を選択する。」といったルールであってもよく、「一致重複頻度に基づく値が所定値よりも高く、矛盾重複頻度に基づく値が所定値よりも低い場合は、低機能で高速な統合処理方法を選択する。」や、「一致重複頻度に基づく値が所定値よりも低く、矛盾重複頻度に基づく値が所定値よりも高い場合は、高機能な統合処理方法を選択する。」といったルールであってもよい。例えば、図 5 に示すルールは、「一致重複頻度に後述する出現頻度をかけた値が所定値よりも高く、矛盾重複頻度に後述する出現頻度をかけた値が所定値以下である場合には、低機能で高速な統合処理方法を選択する。」や、「一致重複頻度に後述する出現頻度をかけた値が所定値以下であり、矛盾重複頻度に後述する出現頻度をかけた値が所定値よりも高い場合には、高機能な統合処理方法を選択する。」といったルールである。なお、ルール 812 の記述方法については、論理式に限定する必要はなく、決定木などの記述を用いることもできる。

[0045] 統合処理方法 ID は、統合処理方法（具体的には、統合処理方法実行手段（9～11））を一意に特定するために用いられる。ここで、統合処理方法

IDの表現方法については、文字列に限定する必要はなく、任意の文字列やURIなど、統合処理方法を一意に特定できる表現であれば、任意の表現を用いることができる。

[0046] 次に、データ統合処理システムの第1の実施形態の動作例について説明する。

[0047] 第1の実施形態のデータ統合処理システムは、データ統合処理Saと、統合処理方法選択処理Sbと、特性学習処理Scとを実行する。

[0048] データ統合処理Saでは、データ統合処理システムは、解析処理装置2の要求に対して一連の統合処理を実行し、統合処理後の結果を返信する。

[0049] また、統合処理方法選択処理Sbでは、統合処理方法選択手段4が、統合対象のグラフ群に最適な統合処理方法群を選択する。

[0050] また、特性学習処理Scでは、特性学習手段7が、統合処理方法選択手段4から統合対象のグラフ群とグラフ群を出力した解析手段群の情報とを受け取り、解析手段別特性記憶手段5が記憶する情報を更新する。

[0051] まず、データ統合処理Saについて、図を用いて説明する。図6に、データ統合処理システムが実行するデータ統合処理の流れの一例を示すフローチャートを示す。

[0052] データ統合処理Saでは、まず、統合処理方法選択手段4は、解析処理装置2からデータ統合処理の要求を受信する（ステップSa1）。

[0053] 次に、統合処理方法選択手段4は、解析処理装置2から、解析結果群を表現した解析結果グラフ群と、グラフを出力した解析手段のID群とを受信する（ステップSa2）。

[0054] 次に、統合処理方法選択手段4は、統合処理方法選択処理（ステップSb）を行い、統合処理方法ID群を選択する。統合処理方法選択処理（ステップSb）の詳細については後述する。

[0055] 次に、統合処理方法選択手段4は、解析結果グラフ群と、統合処理方法ID群とを、統合処理実行手段6に出力する。

[0056] 次に、統合処理実行手段6は、すべての解析結果グラフに対して、解析結

果グラフに対応する統合処理方法IDを統合処理方法ID群から抽出し、統合処理方法IDに対応する統合処理方法実行手段9～11のいずれかに2つのグラフの統合処理を実行させる（ステップSa3～Sa5）。具体的には、統合処理実行手段6は、解析結果グラフに対応する統合処理方法IDを抽出し、抽出した統合処理方法IDによって特定される統合処理方法実行手段9～11のいずれかに統合処理要求を出力する。すると、統合処理方法実行手段（9～11のいずれか）は、要求に従って、2つのグラフを統合する処理を実行する。

[0057] 最後に、統合処理実行手段6は、解析処理装置2に統合処理後のグラフを送信する（ステップSa6）。

[0058] 次に、統合処理方法選択処理Sbについて、図を用いて説明する。図7に、データ統合処理システムが実行する統合処理選択処理の流れの一例を示すフローチャートを示す。

[0059] 統合処理選択処理Sbでは、まず、統合処理方法選択手段4は、すべての解析結果グラフと解析結果グラフを出力した解析手段のIDとに対して、対応する特性情報を解析手段別特性記憶手段5から抽出する。そして、統合処理方法選択手段4は、抽出した特性情報に基づいて、統合処理方法選択ルール記憶手段8が記憶するルール情報から適合するルールを特定する（ステップSb3、ステップSb4）。具体的には、統合処理方法選択手段4は、抽出した特性情報に含まれる一致重複頻度および矛盾重複頻度とその出現頻度とに基づいて、ルール812に記載された条件と一致する選択ルールIDを特定する。

[0060] 次に、統合処理方法選択手段4は、特定したルールに対応する統合処理方法ID813を統合処理方法選択ルール記憶手段8から取得し、内部的に保持する（ステップSb5）。具体的には、統合処理方法選択手段4は、特定した選択ルールID811に対応する統合処理方法ID813を示す情報を統合処理方法選択ルール記憶手段8から抽出し、抽出した情報を記憶部に一時的に記憶させる。

- [0061] 最後に、統合処理方法選択手段4は、統合処理方法ID群を統合制御手段に出力する（ステップS b 6）。具体的には、統合処理方法選択手段4は、抽出した統合処理方法ID 8 1 3を示す情報をデータ統合処理装置2の各手段を制御する統合制御手段に出力する。
- [0062] 次に、特性学習処理S cについて、図を用いて詳細に説明する。図8に、データ統合処理システムが実行する特性学習処理の流れの一例を示すフローチャートを示す。
- [0063] 特性学習処理S cでは、まず、特性学習手段7は、統合処理方法選択手段4から解析結果のグラフ群と、グラフ群を出力した解析処理装置2の解析手段のID群とを受け取る（ステップS c 1）。
- [0064] 次に、特性学習手段7は、受け取ったグラフ群のすべてのペアの組合せについて、以下の処理を実行する（ステップS c 2）。
- [0065] まず、特性学習手段7は、グラフのペアについて、解析結果グラフ間での一致重複頻度を算出する（ステップS c 3）。
- [0066] 次に、特性学習手段7は、グラフのペアについて、解析結果グラフ間でのクラス／プロパティ出現頻度を算出する（ステップS c 4）。
- [0067] 次に、特性学習手段7は、グラフのペアについて、解析結果グラフ間での矛盾重複頻度を算出する（ステップS c 5）。
- [0068] ここで、ステップS c 3、S c 4、S c 5の処理については、並列実行することが可能であり、実行順序は問わない。
- [0069] 次に、特性学習手段7は、ステップS c 3～S c 5で算出した各頻度について、解析手段別特性記憶手段5から対応する頻度を示す情報を抽出し、それぞれ加重平均を求める（ステップS c 6）。
- [0070] 最後に、特性学習手段7は、加重平均を求めた各頻度値を、解析手段別特性記憶手段5に記憶させる（ステップS c 7）。
- [0071] 特性学習手段7は、すべてのグラフの組合せについて、ステップS c 3からS c 7までの処理を実行すると、学習処理を終了する（ステップS c 2）。データ統合処理システムは、このような特性学習処理S cを所定期間ごと

に実行することにより、解析手段別特性記憶手段 5 が記憶している特性情報を随時更新する。

[0072] 次に、本実施形態の効果について説明する。

[0073] 本実施形態の効果は、2つのグラフの統合時に要求される機能を提供する統合処理方法を用いて統合処理を実行することにより、機能を制限することなく、統合処理のスループットを向上できることにある。

[0074] この理由は、解析手段別特性記憶手段 5 が記憶しているグラフ間の統計情報に基づいて、統合処理方法選択手段 4 が適切な統合処理方法を選択することで、不用意に高度な処理をすべてのグラフへ適用することを防止し、無駄な処理を削減できるためである。

[0075] 実施形態 2.

次に、本発明の第 2 の実施形態について説明する。図 9 に、データ統合処理システムの第 2 の実施形態の全体構成の一例を示す機能ブロック図を示す。

[0076] 第 2 の実施形態のデータ統合処理システムは、グラフ分割手段 1 2 を含み、グラフ分割手段 1 2 が統合前のグラフをサブグラフ群に分割する点で、第 1 の実施形態と相違している。

[0077] 図 9 において、グラフ分割手段 1 2 以外の構成要素については、第 1 の実施形態と同様である。第 1 の実施形態と同様の構成要素については、図 1 と同一の符号を付し、詳細な説明を省略する。

[0078] グラフ分割手段 1 2 は、統合処理方法選択手段 4 から、統合対象のグラフ群とそのグラフ群を出力した解析手段 ID 群とを受け取る。そして、グラフ分割手段 1 2 は、解析手段別特性記憶手段 5 内の特性情報に基づいて、グラフ内で同じ特性を示す部分をサブグラフとして抽出し、統合対象のグラフ群をサブグラフ群に分割する。グラフ分割手段 1 2 は、具体的には、プログラムに従って動作する情報処理装置の CPU によって実現される。

[0079] 次に、第 2 の実施形態の動作例について説明する。

[0080] 本実施形態では、第 1 の実施形態と同様に、データ統合処理 S d と、統合

処理選択処理 S b と、特性学習処理 S c とを実行する。ただし、本実施形態では、サブグラフ分割処理 S e を実行し、データ統合処理 S d において、グラフ分割手段 1 2 が、入力グラフをサブグラフへ分割する点で、第 1 の実施形態と異なる。

[0081] 統合処理選択処理 S b と、特性学習処理 S c とについては、第 1 の実施形態と同様であるため、詳細な説明を省略する。

[0082] データ統合処理 S d について、図を用いて説明する。図 10 に、データ統合処理システムが実行するデータ統合処理 S d の流れの一例を示すフローチャートを示す。なお、図 10 において第 1 の実施形態と同様の動作要素については、図 6 と同様の符号を付し、詳細な説明を省略する。

[0083] まず、統合処理方法選択手段 4 は、解析処理装置 2 からデータ統合処理の要求を受け付ける（ステップ S a 1）。

[0084] 次に、統合処理方法選択手段 4 は、解析処理装置 2 から、解析結果群を表現した解析結果グラフ群と、グラフを出力した解析手段の ID 群とを受信する（ステップ S a 2）。

[0085] 次に、統合処理方法選択手段 4 は、受信した解析結果グラフ群と、グラフを出力した解析手段の ID 群とをグラフ分割手段 1 2 に出力する。すると、グラフ分割手段 1 2 は、グラフをサブグラフに分割する（ステップ S e）。ステップ S e の詳細については後述する。

[0086] 以降の処理については、第 1 の実施形態における処理（ステップ S b ~ S a 6）と同様であるため、説明を省略する。

[0087] サブグラフ分割処理 S e について、図を用いて説明する。図 11 に、データ統合処理システムが実行するサブグラフ分割処理 S e の流れの一例を示すフローチャートを示す。

[0088] サブグラフ分割処理 S e では、まず、グラフ分割手段 1 2 は、統合処理方法選択手段 4 から、解析結果グラフ群と、グラフを出力した解析処理装置 2 の解析手段の ID 群とを受け取り、すべてのグラフに対して、以下の処理を実施する（ステップ S e 1）。

- [0089] まず、グラフ分割手段12は、グラフの中に含まれるクラスとプロパティについて、解析手段別特性記憶手段5から、一致重複度テーブル510と矛盾重複度テーブル520とを参照し、一致重複頻度が高く、矛盾重複頻度が低いものを列挙する（ステップS e 2）。
- [0090] 次に、グラフ分割手段12は、一致重複頻度が高く、矛盾重複頻度が低いクラス・プロパティを多く含む部分グラフを、高一致サブグラフとして抽出する（ステップS e 3）。
- [0091] 次に、グラフ分割手段12は、グラフの中に含まれるクラスとプロパティについて、解析手段別特性記憶手段5から、一致重複度テーブル510と矛盾重複度テーブル520とを参照し、一致重複頻度が低く、矛盾重複頻度が高いものを列挙する（ステップS e 4）。
- [0092] 次に、グラフ分割手段12は、一致重複頻度が低く、矛盾重複頻度が高いクラス・プロパティを多く含む部分グラフを、高矛盾サブグラフとして抽出する（ステップS e 5）。
- [0093] 次に、グラフ分割手段12は、グラフ中において、高一致サブグラフと高矛盾サブグラフとのいずれにも含まれていないデータをサブグラフとして抽出する（ステップS e 6）。
- [0094] 最後に、これらの処理をすべてのグラフに対して実行した後、グラフ分割手段12は、抽出したサブグラフ群を統合処理方法選択手段4に出力する（ステップS e 8）。
- [0095] 次に、本実施形態の効果について説明する。
- [0096] 本実施形態の効果は、グラフのサイズが大きくなった際に、グラフを分割することで統合処理を高速化して処理時間を短縮できることにある。
- [0097] この理由は、グラフ分割手段12が、グラフを類似した特性を持ったデータをまとめたサブグラフに分割できるため、各統合処理方法に従って効率的に統合処理を行えるためである。
- [0098] 実施形態3.

次に、本発明の第3の実施形態について説明する。ここでは、解析処理装

置 2 は、下記に示す解析手段を備えているものとする。

[0099] (1) 入力された名前に対応する、顧客の電子メールアドレス、住所、名前を出力する顧客情報検索エンジン

[0100] (2) 入力された名前に対応する、社員の電子メールアドレス、住所、名前、社員 ID を出力する社員情報検索エンジン

[0101] (3) 社員 ID のオフィスでの移動軌跡を出力する人物動線検索エンジン

[0102] ここで、(1) から (3) の解析手段がそれぞれ出力する解析結果データの例を図 1 2 から図 1 4 に示す。なお、(1) から (3) の解析手段が出力する解析結果データについての特徴情報の内容は、図 2 から図 4 に示すものであるとする。

[0103] また、本実施形態では、データ統合処理装置は、下記に示す統合処理方法に従って統合処理を実行可能であるものとする。

[0104] (い) 第 1 の統合処理方法：入力された 2 つのグラフについて、ノードの ID や値が同じノードを同一とみなして単純に統合する統合処理を実行する方法

[0105] (ろ) 第 2 の統合処理方法：入力された 2 つのグラフについて、同じノード ID が持つプロパティが異なるノードの値が矛盾している（異なっている）場合には、2 つの入力モデルを走査してどちらの値を使用するかを判定する統合処理を実行する方法

[0106] ここで統合処理の計算量について比較すると、(い) の統合処理方法では、2 つのグラフ間でノードの一致を走査する必要があるため、ノード数を  $N$  とすると、 $N^2$  の計算量が必要である。一方、(ろ) の統合処理方法では、対象ノードのプロパティ統合時に、矛盾したノードの矛盾を解決するためにモデルを走査するため、 $N^2 \times N^2$  の計算量が必要である。

[0107] 上記の (1) から (3) の解析手段が出力する解析結果データの組合せと、必要とする統合機能のレベルとは、下記のような関係になる。

[0108] (1) - (2) を統合：異なるプロパティがあった場合にはこの矛盾を解決しながら統合する

- [0109] (1) - (3) を統合：単純にすべて統合
- [0110] (2) - (3) を統合：単純にすべて統合
- [0111] また、統合処理方法選択ルール記憶手段 8 には、図 5 に示したルールが設定されているものとする。具体的には、統合処理方法選択ルール記憶手段 8 は、ルール情報として、図 5 に示す選択ルールテーブル 810 を記憶している。
- [0112] 次に、具体的な動作例を示す。統合処理方法選択手段 4 は、解析処理装置 2 から、(1) および (2) の解析手段の解析結果データを受け取ると、次のように動作する。
- [0113] まず、統合処理方法選択手段 4 は、受信した解析結果データに含まれる解析手段 ID 群（この場合、顧客情報検索エンジンと社員情報検索エンジン）に基づいて、解析手段別特性記憶手段 5 から、一致重複頻度（Person 70%）と矛盾重複頻度（e-mail 80%）とを抽出する。
- [0114] さらに、統合処理方法選択手段 4 は、受信した解析結果データに含まれる解析手段 ID に基づいて、解析手段別特性記憶手段 5 から、クラス／プロパティの出現頻度（Person 33%、e-mail 33%）を抽出する。
- [0115] 次に、統合処理方法選択手段 4 は、統合処理方法選択ルール記憶手段 8 が記憶するルール情報と抽出した上記の頻度情報とに基づいて、ルールに一致するか否かの判定を行う。
- [0116] ここでは、統合処理方法選択手段 4 は、ルール ID：002（図 5 参照）のルールに一致すると判定し、ルール ID：002 に対応付けられた第 2 の統合処理方法を選択する。
- [0117] 最後に、統合処理方法選択手段 4 は、統合制御手段に第 2 の統合処理方法を特定する統合処理方法 ID を出力する。この場合、データ統合処理装置 1 は、異なるプロパティがあった場合には、この矛盾を解決しながら統合する機能を提供する。そのため、計算時間（計算量）は、 $N^2 \times N^2$  となる。
- [0118] また、統合処理方法選択手段 4 は、解析処理装置 2 から (1) および (3)

)の解析手段の解析結果データを受け取ると、次のように動作する。

- [0119] まず、統合処理方法選択手段4は、受信した解析結果データに含まれる解析手段ID群(この場合、顧客情報検索エンジンと人物動線検索エンジン)に基づいて、解析手段別特性記憶手段5から、一致重複頻度(Person 100%)と矛盾重複頻度(なし)とを抽出する。
- [0120] さらに、統合処理方法選択手段4は、受信した解析結果データに含まれる解析手段IDに基づいて、解析手段別特性記憶手段5から、クラス/プロパティの出現頻度(Person 33%)を抽出する。
- [0121] 次に、統合処理方法選択手段4は、統合処理方法選択ルール記憶手段8が記憶するルール情報と抽出した上記の頻度情報とに基づいて、ルールに一致するか否かの判定を行う。
- [0122] ここでは、統合処理方法選択手段4は、ルールID:001(図5参照)のルールに一致すると判定し、ルールID:001に対応付けられた第1の統合処理方法を選択する。
- [0123] 最後に、統合処理方法選択手段4は、統合制御手段に第1の統合処理方法を特定する統合処理方法IDを出力する。この場合、データ統合処理装置1は、単純なID一致型の統合機能を提供する。そのため、計算時間(計算量)は、 $N^2$ となる。
- [0124] 上記のように、本実施形態では、統合機能として(ろ)が定義する機能を提供しつつ、その機能を必要としない場合には、より高速な(い)を適用する。そのため、システム全体のスループットを(ろ)のみを使用するシステムに対して向上させることができる。
- [0125] 具体的には、(1)と(2)との解析結果の統合要求が20%、(1)と(3)との解析結果の統合要求が80%のシステムであれば、平均処理時間は、 $N^2 \times (0.8 + 0.2 \times N^2)$ となる。そのため、(ろ)のみを使用するシステムの平均処理時間( $N^2 \times N^2$ )より高速となる。
- [0126] 以上のように、本発明は、複数の機能群を統合し動作する処理基盤における処理結果データの統合性能を向上させるためのものである。

- [0127] 次に、本発明によるデータ統合処理装置の最小構成について説明する。図16は、データ統合処理装置の最小の構成例を示すブロック図である。図16に示すように、データ統合処理装置は、最小の構成要素として、入力されたグラフ群の統合に用いる統合処理方法を選択する統合処理方法選択手段4と、統合処理方法を複数有する統合処理実行手段6とを備えている。
- [0128] 図16に示す最小構成のデータ統合処理装置では、統合処理方法選択手段4は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、下位のノードを統合するために用いる統合処理方法を選択する。そして、統合処理実行手段6は、複数の統合処理方法のうち統合処理方法選択手段4が選択した統合処理方法に従って統合処理を実行することにより、入力された複数のグラフを統合する。
- [0129] 従って、最小構成のデータ統合処理装置によれば、2つのグラフの統合時に要求される機能を提供する統合処理方法を適用することにより、機能を制限することなく、統合処理のスループットを向上できることにある。
- [0130] なお、本実施形態では、以下の(1)～(6)に示すようなデータ統合処理装置の特徴的構成が示されている。
- [0131] (1) データ統合処理装置は、複数のグラフ（例えば、解析処理装置2の解析手段による解析結果データ）を統合するデータ統合処理装置（例えば、データ統合処理装置1によって実現される）であって、入力されたグラフ群の統合に用いる統合処理方法（例えば、(い)第1の統合処理方法など）を選択する統合処理方法選択手段（例えば、統合処理方法選択手段4によって実現される）と、統合処理方法を複数有し、複数の統合処理方法のうち統合処理方法選択手段が選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合する統合処理実行手段（例えば、統合処理実行手段6によって実現される）とを備え、統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、下位のノードを統合するために用いる統合処理方法を選択することを特徴とする。

[0132] (2) データ統合処理装置において、統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度である一致重複頻度と、下位のノードが存在するが一致しない頻度である矛盾重複頻度とに基づいて、下位のノードを統合するために用いる統合処理方法を選択するように構成されていてもよい。

[0133] (3) データ統合処理装置において、統合処理実行手段は、統合機能が高く処理速度が遅い第2の統合処理方法または統合機能が低く処理速度が速い第1の統合処理方法に従って統合処理を実行可能であり、統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度が高い場合には、第1の統合処理方法を選択し、頻度が低い場合には、第2の統合処理方法を選択するように構成されていてもよい。

[0134] (3-1) 具体的に、統合処理実行手段は、所定の統合機能を有すると共に所定の処理速度で統合処理可能な第1の統合処理方法、または、当該第1の統合処理方法よりも統合機能が高いが処理速度が遅い第2の統合処理方法、に従って統合処理を実行可能であり、

統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に基づく値が所定値よりも高い場合には、第1の統合処理方法を選択し、頻度に基づく値が所定値よりも低い場合には、第2の統合処理方法を選択するよう構成されていてもよい。

[0135] (3-2) さらに、統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に基づく値が所定値よりも高く、一致しない頻度に基づく値が所定値よりも低い場合には、第1の統合処理方法を選択し、一致する頻度に基づく値が所定値よりも低く、一致しない頻度に基づく値が所定値よりも高い場合には、第2の統合処理方法を選択するよう構成されていてもよい。

[0136] (4) データ統合処理装置において、入力されたグラフを複数のサブグラ

フに分割するグラフ分割手段（例えば、グラフ分割手段12によって実現される）を備え、グラフ分割手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に基づいてグラフをサブグラフに分割し、統合処理方法選択手段は、グラフ分割手段が分割したサブグラフ単位で統合処理方法を選択し、統合処理実行手段は、グラフ分割手段が分割したサブグラフ単位で統合処理を実行するように構成されていてもよい。

[0137] （5）データ統合処理装置において、過去に入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する統計頻度を、グラフを出力した解析手段と対応付けて格納する解析手段別特性記憶手段（例えば、解析手段別特性記憶手段5によって実現される）を備え、統合処理方法選択手段は、入力されたグラフを出力した解析手段に基づいて統計頻度を解析手段別特性記憶手段から抽出し、抽出した統計頻度に基づいてグラフを統合するために用いる統合処理方法を選択するように構成されていてもよい。

[0138] （6）データ統合処理装置において、過去に入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する統計頻度を算出し、解析手段別特性記憶手段に格納させる特性学習手段（例えば、特性学習手段7によって実現される）を備え、特性学習手段は、入力されたグラフを出力した解析手段を示す情報を統合処理方法選択手段から入力し、入力した情報に基づいて統計頻度を算出し、解析手段別特性記憶手段が格納する情報を順次更新するように構成されていてもよい。

[0139] なお、上記各実施形態においてプログラムは、記憶装置に記憶されていたり、コンピュータが読み取り可能な記録媒体に記録されている。例えば、記録媒体は、フレキシブルディスク、光ディスク、光磁気ディスク、及び、半導体メモリ等の可搬性を有する媒体である。

[0140] 以上、上記各実施形態を参照して本願発明を説明したが、本願発明は、上述した実施形態に限定されるものではない。本願発明の構成や詳細には、本

願発明の範囲内で当業者が理解しうる様々な変更をすることができる。

[0141] なお、本発明は、日本国にて2010年9月13日に特許出願された特願2010-204210の特許出願に基づく優先権主張の利益を享受するものであり、当該特許出願に記載された内容は、全て本明細書に含まれるものとする。

### 産業上の利用可能性

[0142] 本発明は、複数のグラフデータを統合する処理基盤における統合処理のスループットを向上させるための、データ統合処理装置やデータ統合処理装置をコンピュータに実現するためのプログラムといった用途に適用できる。

### 符号の説明

- [0143]
- 1 データ統合処理装置
  - 2 解析処理装置
  - 4 統合処理方法選択手段
  - 5 解析手段別特性記憶手段
  - 6 統合処理実行手段
  - 7 特性学習手段
  - 8 統合手段選択ルール記憶手段
  - 9 第1の統合処理方法実行手段
  - 10 第2の統合処理方法実行手段
  - 11 第3の統合処理方法実行手段
  - 12 グラフ分割手段
  - 510 一致重複頻度テーブル
  - 520 矛盾重複頻度テーブル
  - 530 クラスプロパティ出現頻度テーブル
  - 810 選択ルールテーブル

## 請求の範囲

- [請求項1] 複数のグラフを統合するデータ統合処理装置であって、  
入力されたグラフ群の統合に用いる統合処理方法を選択する統合処理方法選択手段と、  
統合処理方法を複数有し、該複数の統合処理方法のうちの前記統合処理方法選択手段が選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合する統合処理実行手段とを備え、  
前記統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、前記下位のノードを統合するために用いる統合処理方法を選択する  
ことを特徴とするデータ統合処理装置。
- [請求項2] 統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度である一致重複頻度と、下位のノードが存在するが一致しない頻度である矛盾重複頻度とに基づいて、前記下位のノードを統合するために用いる統合処理方法を選択する  
請求項1記載のデータ統合処理装置。
- [請求項3] 統合処理実行手段は、統合機能が高く処理速度が遅い第2の統合処理方法または統合機能が低く処理速度が速い第1の統合処理方法に従って統合処理を実行可能であり、  
統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度が高い場合には、前記第1の統合処理方法を選択し、前記頻度が低い場合には、前記第2の統合処理方法を選択する  
請求項1又は請求項2記載のデータ統合処理装置。
- [請求項4] 統合処理実行手段は、所定の統合機能を有すると共に所定の処理速度で統合処理可能な第1の統合処理方法、または、当該第1の統合処

理方法よりも統合機能が高いが処理速度が遅い第2の統合処理方法、に従って統合処理を実行可能であり、

統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に基づく値が所定値よりも高い場合には、前記第1の統合処理方法を選択し、前記頻度に基づく値が所定値よりも低い場合には、前記第2の統合処理方法を選択する

請求項1又は請求項2記載のデータ統合処理装置。

[請求項5]

入力されたグラフを複数のサブグラフに分割するグラフ分割手段を備え、

前記グラフ分割手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に基づいて前記グラフをサブグラフに分割し、

統合処理方法選択手段は、前記グラフ分割手段が分割したサブグラフ単位で統合処理方法を選択し、

統合処理実行手段は、前記グラフ分割手段が分割したサブグラフ単位で統合処理を実行する

請求項1から請求項4のうちのいずれか1項に記載のデータ統合処理装置。

[請求項6]

過去に入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する統計頻度を、該グラフを出力した解析手段と対応付けて格納する解析手段別特性記憶手段を備え、

統合処理方法選択手段は、入力されたグラフを出力した解析手段に基づいて前記統計頻度を前記解析手段別特性記憶手段から抽出し、抽出した統計頻度に基づいて該グラフを統合する統合処理方法を選択する

ことを特徴とした請求項1から請求項5のうちのいずれか1項に記載のデータ統合処理装置。

[請求項7] 過去に入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する統計頻度を算出し、解析手段別特性記憶手段に格納させる特性学習手段を備え、

前記特性学習手段は、入力されたグラフを出力した解析手段を示す情報を統合処理方法選択手段から入力し、入力した情報に基づいて統計頻度を算出し、前記解析手段別特性記憶手段が格納する情報を順次更新する

請求項6記載のデータ統合処理装置。

[請求項8] 複数のグラフを統合するデータ統合処理システムであって、

入力されたグラフ群の統合に用いる統合処理方法を選択する統合処理方法選択手段と、

統合処理方法を複数有し、該複数の統合処理方法のうちの前記統合処理方法選択手段が選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合する統合処理実行手段とを含み、

前記統合処理方法選択手段は、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、前記下位のノードを統合するために用いる統合処理方法を選択する

ことを特徴とするデータ統合処理システム。

[請求項9] 複数のグラフを統合するデータ統合処理方法であって、

入力されたグラフ群の統合に用いる統合処理方法を選択し、

統合処理方法を複数有し、該複数の統合処理方法のうちの前記選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合し、

統合処理方法を選択する際には、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、前記下位のノードを統合するために用いる統合処理方法を選択する

ことを特徴とするデータ統合処理方法。

[請求項10]

複数のグラフを統合するためのデータ統合処理プログラムであって

、

コンピュータに、

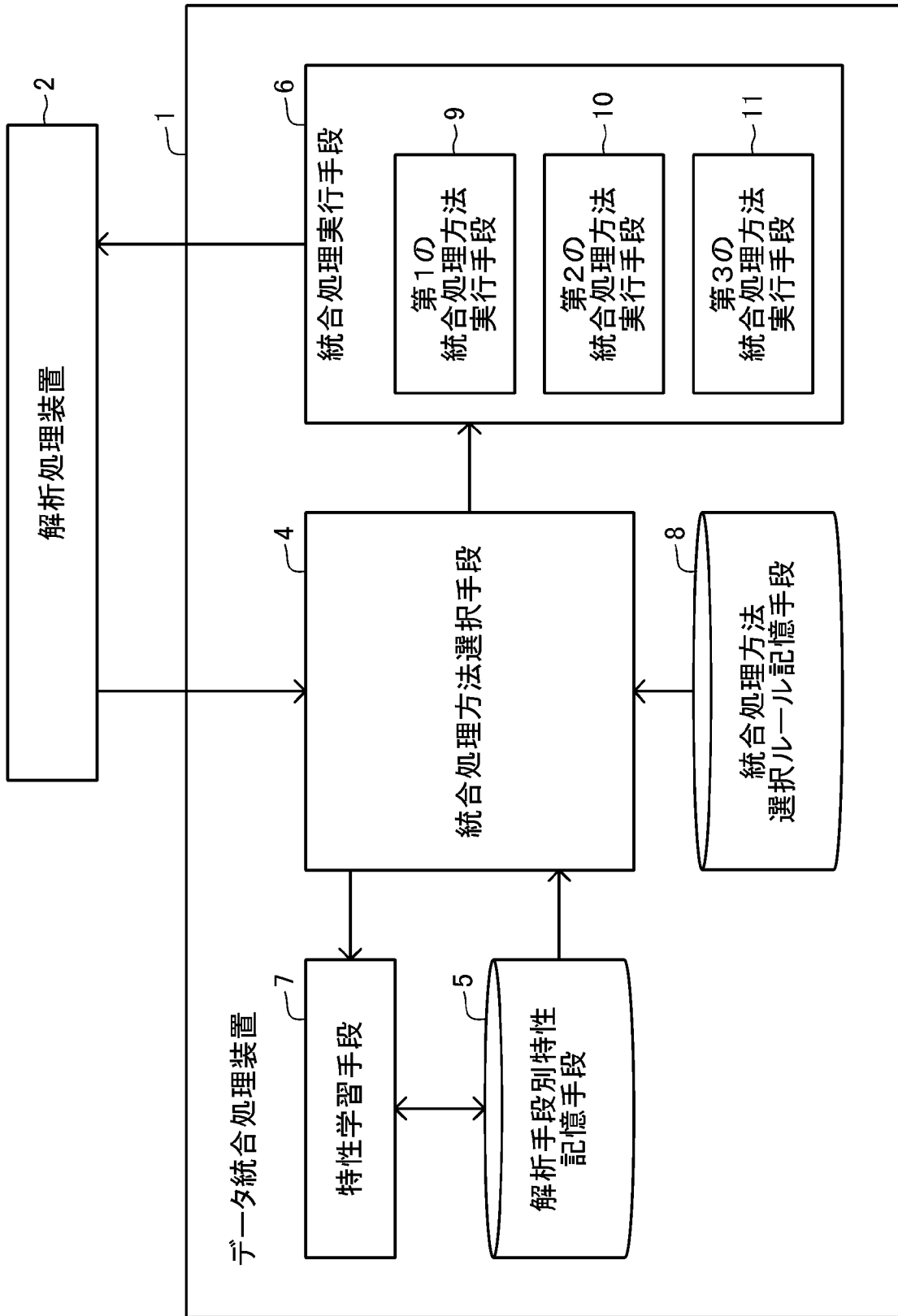
入力されたグラフ群の統合に用いる統合処理方法を選択する統合処理方法選択処理と、

統合処理方法を複数有し、該複数の統合処理方法のうちの選択した統合処理方法に従って統合処理を実行することにより複数のグラフを統合する統合処理実行処理とを実行させ、

前記統合処理方法選択処理で、入力されたグラフ中の各ノードについて、上位のノードが一致する際に、下位のノードが一致する頻度に応じて、前記下位のノードを統合するために用いる処理統合処理方法を選択する処理を

実行させるためのデータ統合処理プログラム。

[図1]



[図2]

510:一致重複頻度テーブル

解析手段 解析手段	社員情報 検索エンジン	顧客情報 検索エンジン	人物動線 検索エンジン
社員情報 検索エンジン		Name(90%)	Person(100%)
顧客情報 検索エンジン			-

[図3]

520:矛盾重複頻度テーブル

解析手段 解析手段	社員情報 検索エンジン	顧客情報 検索エンジン	人物動線 検索エンジン
社員情報 検索エンジン		e-mail(80%)	-
顧客情報 検索エンジン			-

[図4]

530: クラスプロパティ出現頻度テーブル

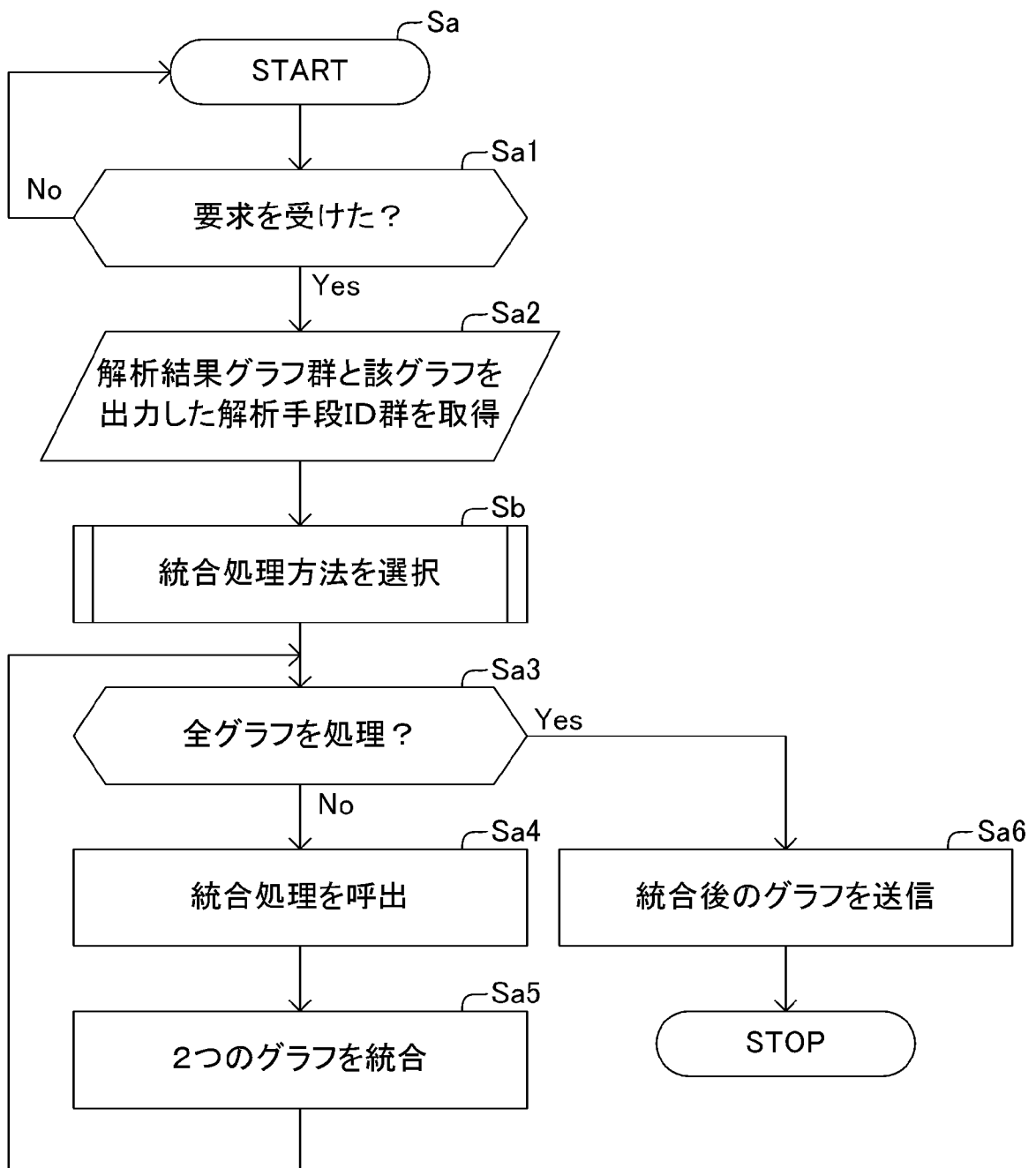
解析手段ID	クラス／プロパティID	頻度
社員情報検索エンジン	Person	33%
社員情報検索エンジン	e-mail	33%
顧客情報検索エンジン	Person	33%
顧客情報検索エンジン	e-mail	12%
⋮	⋮	⋮

[図5]

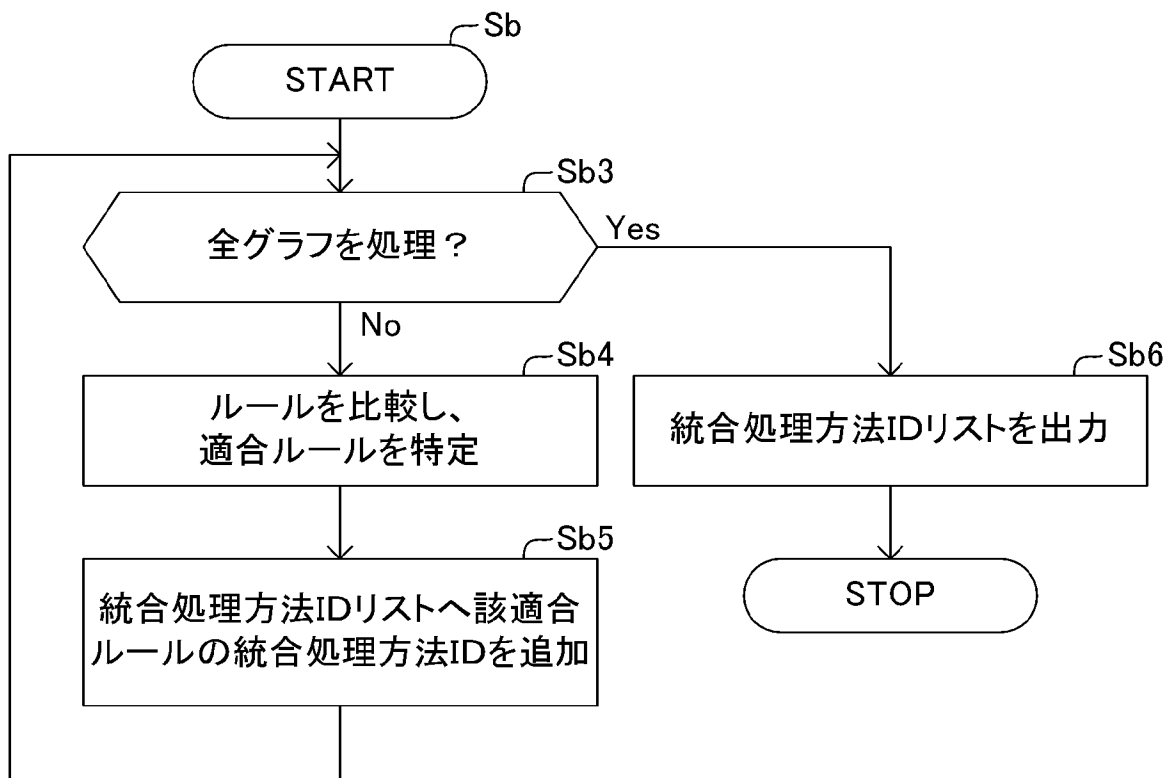
810: 選択ルールテーブル

選択ルールID	ルール	統合処理方法ID
001	一致重複頻度 * 出現頻度 > 0.3 ∧ 矛盾重複頻度 * 出現頻度 ≤ 0.1	第1の統合処理方法
002	一致重複頻度 * 出現頻度 ≤ 0.3 ∧ 矛盾重複頻度 * 出現頻度 > 0.1	第2の統合処理方法

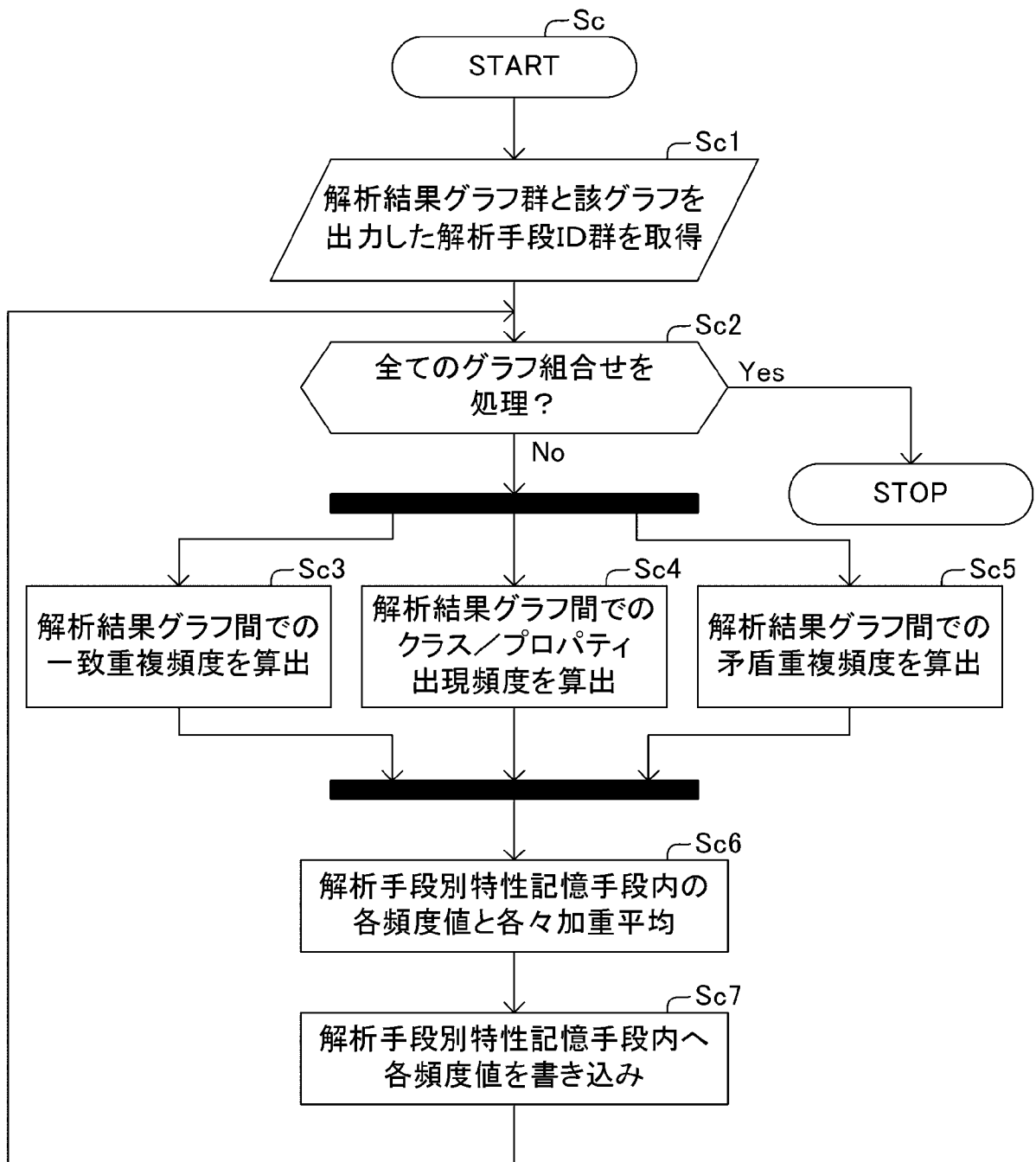
[図6]



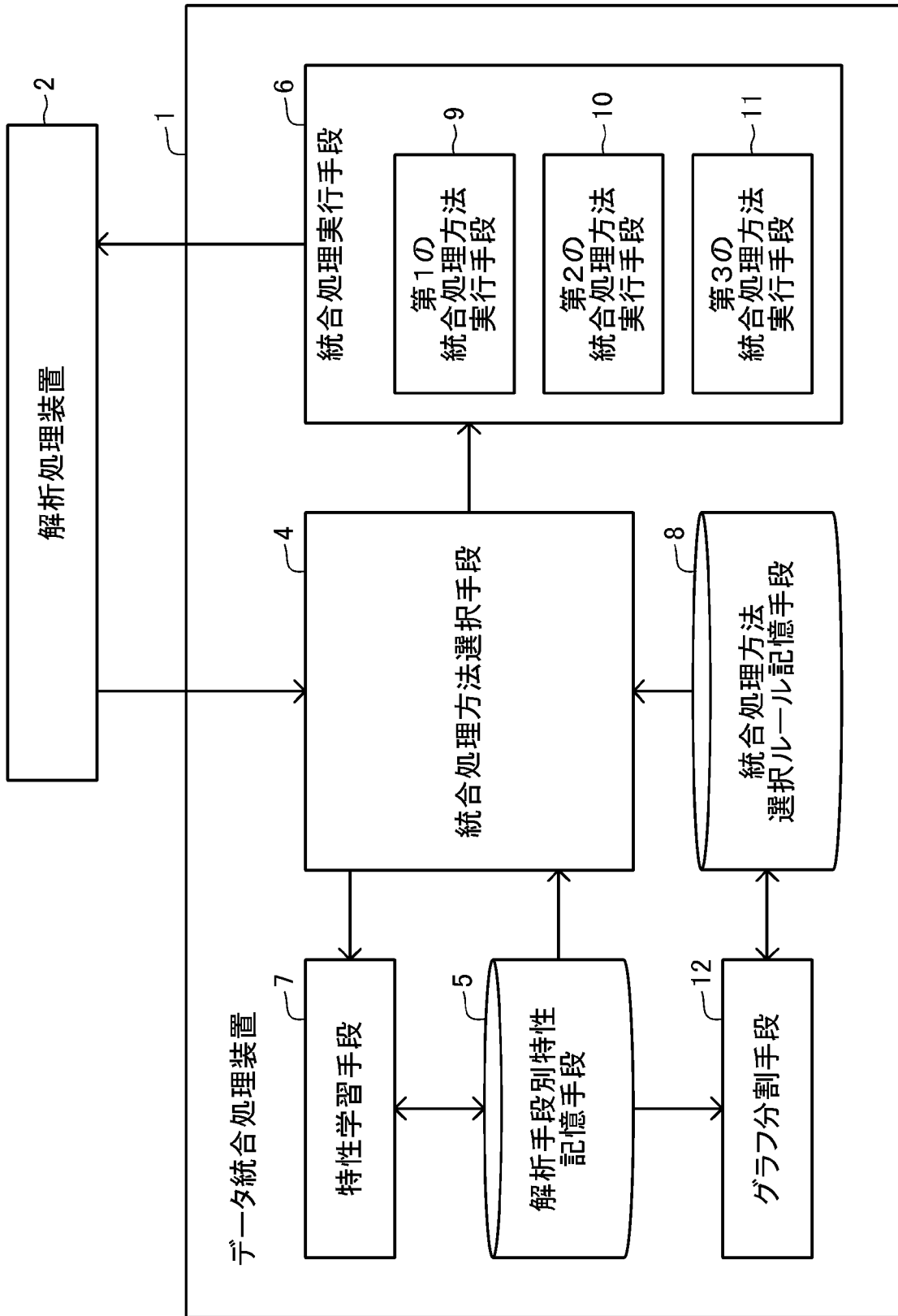
[図7]



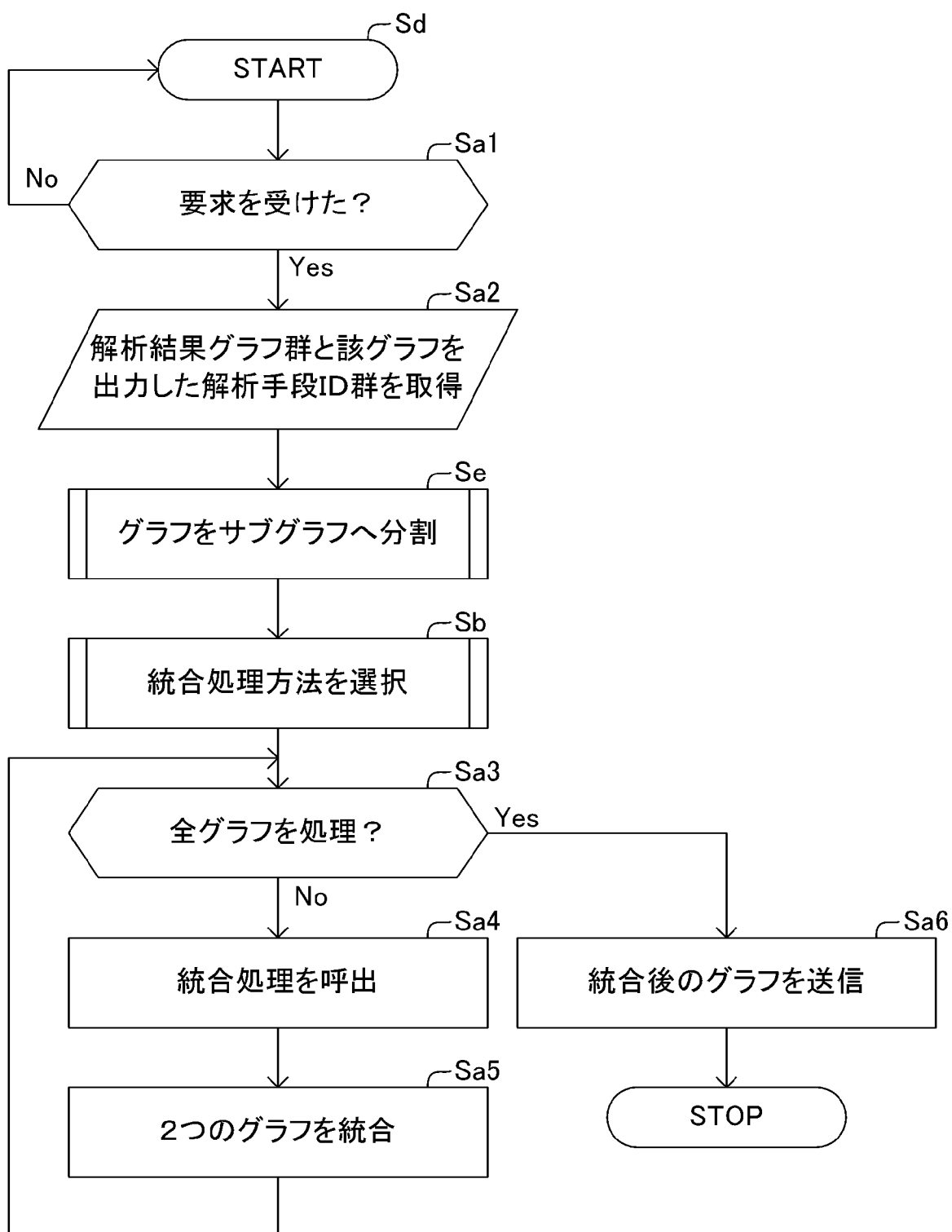
[図8]



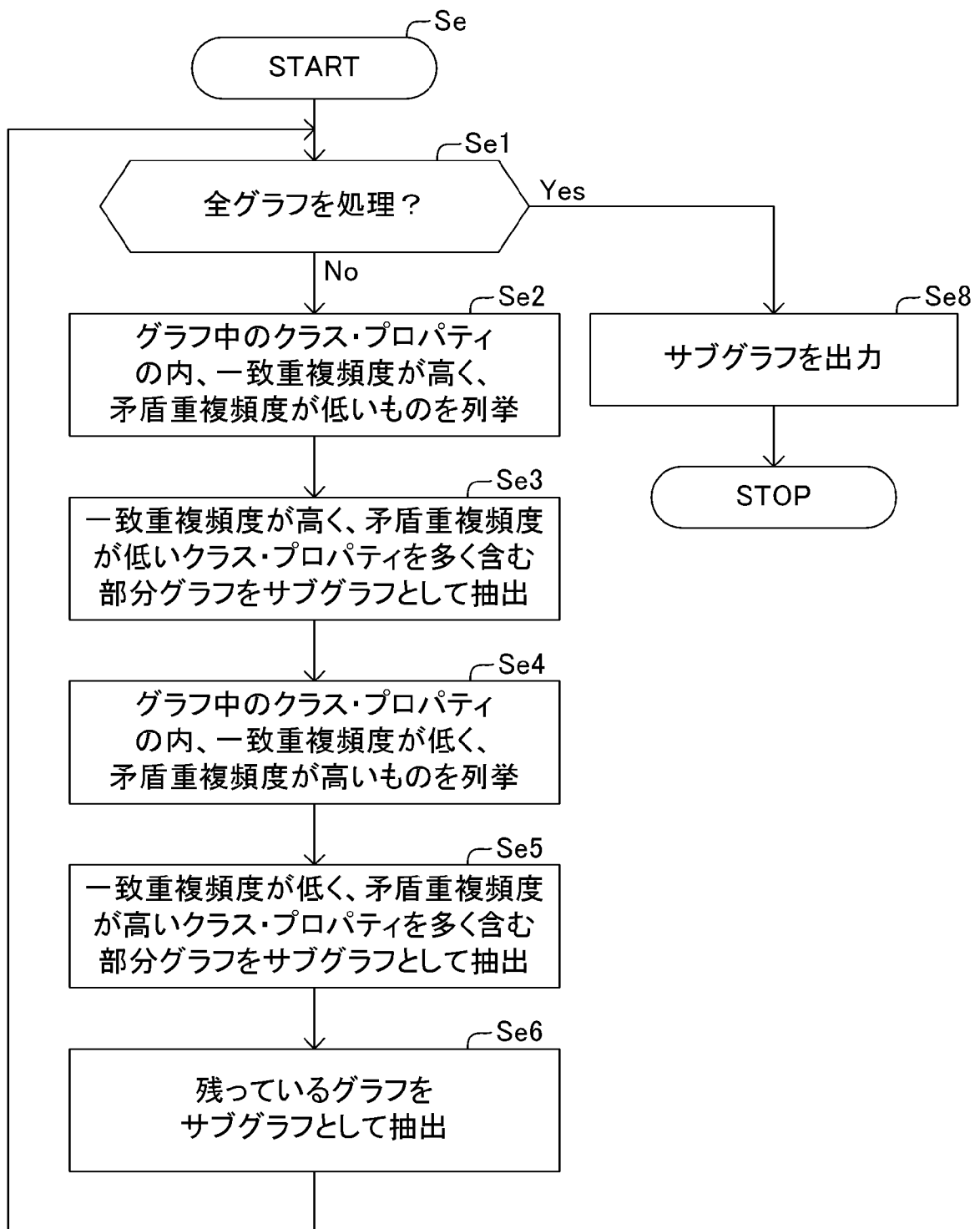
[図9]



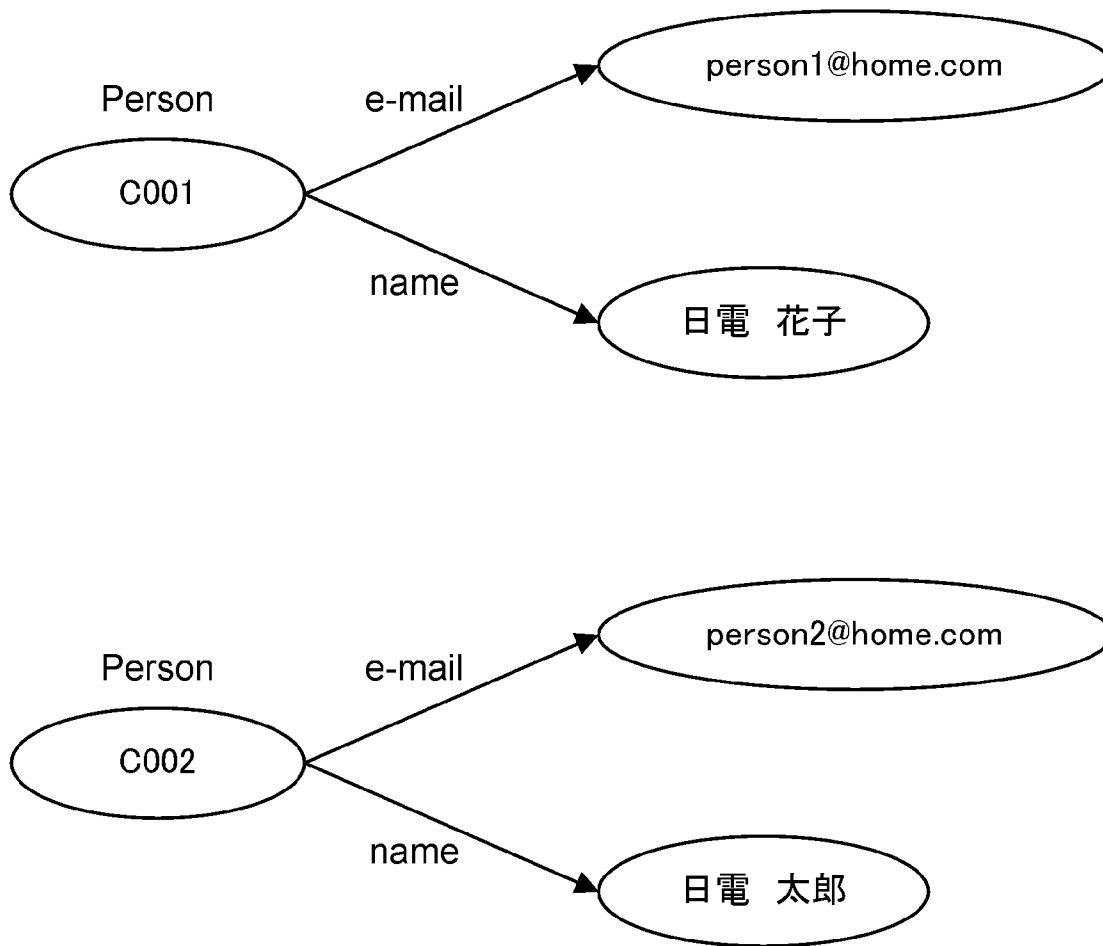
[図10]



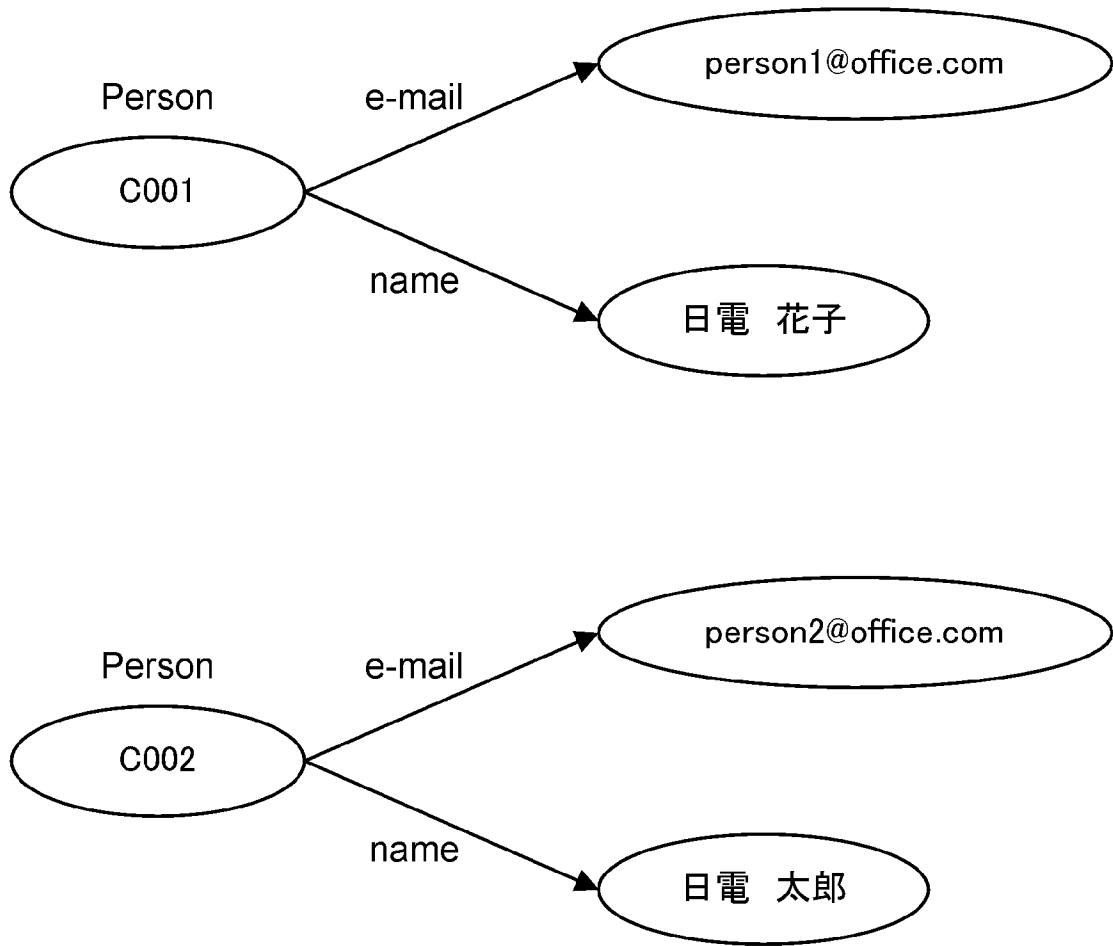
[図11]



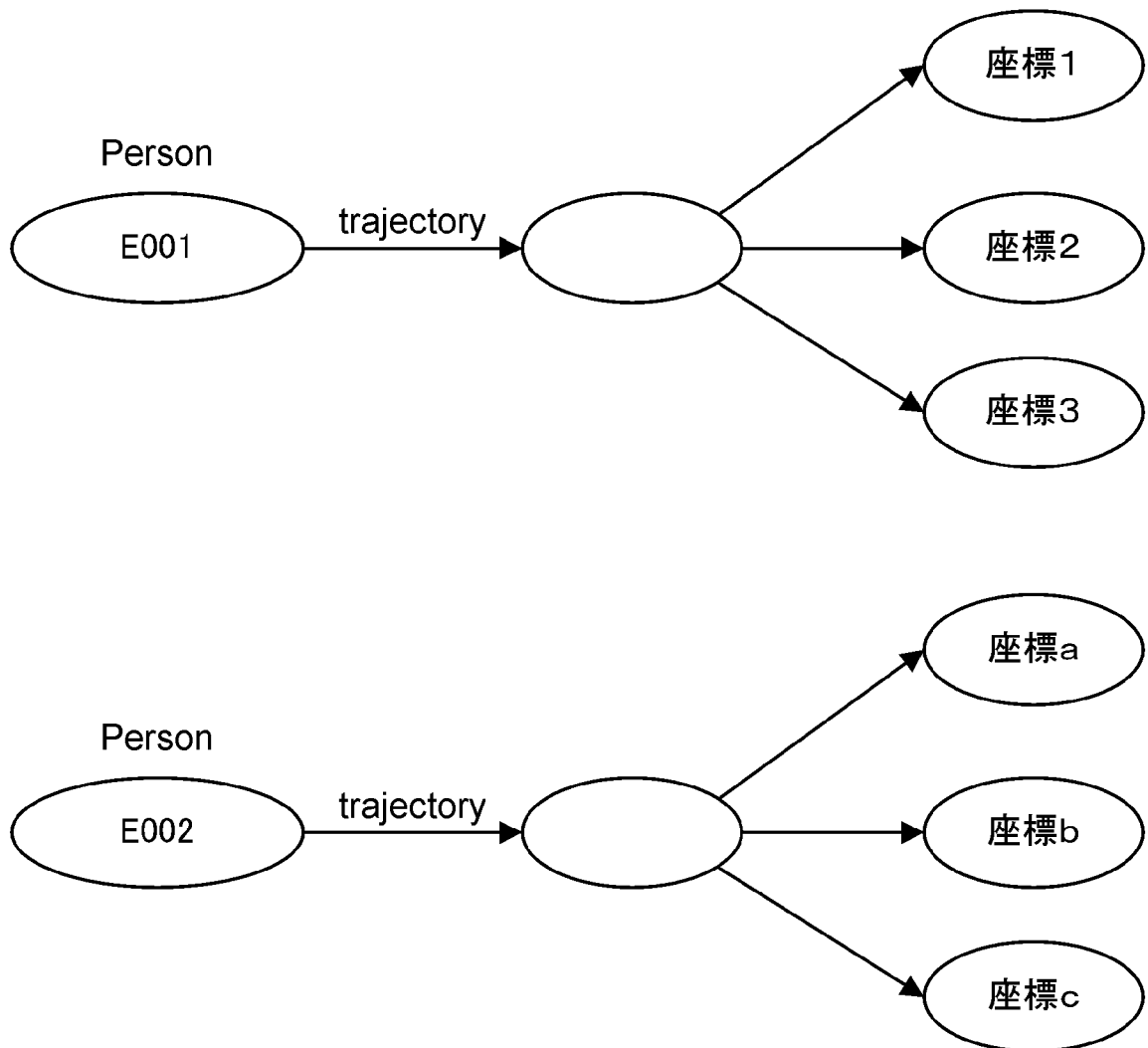
[図12]



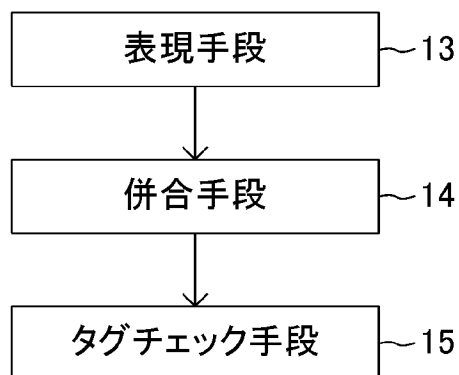
[図13]



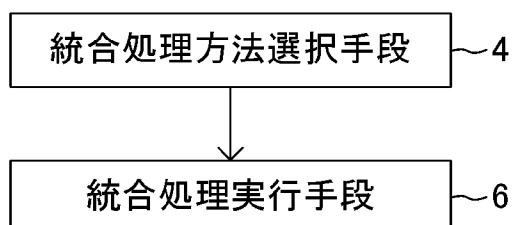
[図14]



[図15]



[図16]



**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2011/005129

**A. CLASSIFICATION OF SUBJECT MATTER**

G06F17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2011
Kokai Jitsuyo Shinan Koho	1971-2011	Toroku Jitsuyo Shinan Koho	1994-2011

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2008-84114 A (Toshiba Corp.), 10 April 2008 (10.04.2008), entire text; all drawings & US 2008/0082574 A1 & US 2010/0299330 A1 & EP 1906315 A1	1-10
A	WO 2008/146807 A1 (NEC Corp.), 04 December 2008 (04.12.2008), entire text; all drawings & US 2010/0121885 A1	1-10
A	JP 2001-14166 A (Fujitsu Ltd.), 19 January 2001 (19.01.2001), entire text; all drawings (Family: none)	1-10

Further documents are listed in the continuation of Box C.

See patent family annex.

\* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search  
02 December, 2011 (02.12.11)

Date of mailing of the international search report  
13 December, 2011 (13.12.11)

Name and mailing address of the ISA/  
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/JP2011/005129

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2005-352874 A (Nippon Telegraph and Telephone Corp.), 22 December 2005 (22.12.2005), entire text; all drawings (Family: none)	1-10

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G06F17/30(2006.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int.Cl. G06F17/30

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2011年
日本国実用新案登録公報	1996-2011年
日本国登録実用新案公報	1994-2011年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2008-84114 A (株式会社東芝) 2008.04.10, 全文, 全図 & US 2008/0082574 A1 & US 2010/0299330 A1 & EP 1906315 A1	1-10
A	WO 2008/146807 A1 (日本電気株式会社) 2008.12.04, 全文, 全図 & US 2010/0121885 A1	1-10
A	JP 2001-14166 A (富士通株式会社) 2001.01.19, 全文, 全図 (ファミリーなし)	1-10

C欄の続きにも文献が列挙されている。

パテントファミリーに関する別紙を参照。

\* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的な技術水準を示すもの  
 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの  
 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)  
 「O」口頭による開示、使用、展示等に言及する文献  
 「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献  
 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの  
 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの  
 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの  
 「&」同一パテントファミリー文献

国際調査を完了した日

02.12.2011

国際調査報告の発送日

13.12.2011

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)  
 郵便番号100-8915  
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

吉田 誠

5M

3659

電話番号 03-3581-1101 内線 3599

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2005-352874 A (日本電信電話株式会社) 2005. 12. 22, 全文, 全図 (ファミリーなし)	1 - 10