

(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  
2017年8月10日(10.08.2017)



(10) 国際公開番号

WO 2017/134801 A1

- (51) 国際特許分類:  
G06Q 50/00 (2012.01) G06Q 10/10 (2012.01)  
G06F 17/24 (2006.01)
- (21) 国際出願番号: PCT/JP2016/053389
- (22) 国際出願日: 2016年2月4日(04.02.2016)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 富士通株式会社 (FUJITSU LIMITED)  
[JP/JP]; 〒2118588 神奈川県川崎市中原区上小田  
中4丁目1番1号 Kanagawa (JP).
- (72) 発明者: 米田 剛(MAITA, Tsuyoshi); 〒1080075 東京  
都港区港南一丁目2番70号 株式会社富士  
通システムズ・イースト内 Tokyo (JP). 野呂 述  
史(NORO, Nobumi); 〒1080075 東京都港区港南一  
丁目2番70号 株式会社富士通システムズ・  
イースト内 Tokyo (JP). 田中 哲(TANAKA, Tetsu);  
〒0368064 青森県弘前市大字東城北三丁目2番  
地の21 株式会社オフコ内 Aomori (JP).
- (74) 代理人: 特許業務法人酒井国際特許事務所  
(SAKAI INTERNATIONAL PATENT OFFICE); 〒

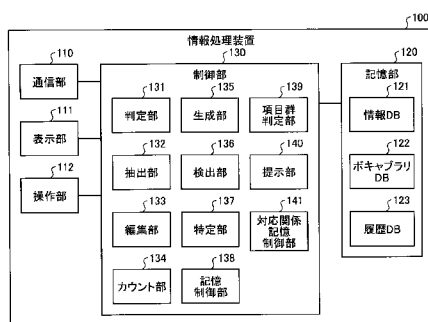
1000013 東京都千代田区霞が関3丁目8番1号  
虎の門三井ビルディング Tokyo (JP).

- (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, BN, 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,  
IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR,  
LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,  
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH,  
PL, PT, QA, RO, RS, RU, RW, SA, 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, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユー  
ラシア (AM, AZ, BY, KG, KZ, 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, KM,  
ML, MR, NE, SN, TD, TG).

[続葉有]

(54) Title: ITEM NAME ASSOCIATING PROCESS METHOD, ITEM NAME ASSOCIATING PROCESS PROGRAM, AND INFORMATION PROCESSING DEVICE

(54) 発明の名称: 項目名対応付け処理方法、項目名対応付け処理プログラム及び情報処理装置



- 100 Information processing device  
110 Communication unit  
111 Display unit  
112 Operation unit  
120 Storage unit  
121 Information database  
122 Vocabulary database  
123 History database  
130 Control unit  
131 Determination unit  
132 Extraction unit  
133 Editing unit  
134 Counting unit  
135 Generation unit  
136 Detection unit  
137 Specification unit  
138 Storage control unit  
139 Item group determination unit  
140 Presentation unit  
141 Association relationship storage control unit

(57) Abstract: Provided are an item name associating process method, an item name associating process program, and an information processing device capable of associating an item name with a standardized vocabulary. The item name associating process method includes the following processes performed by a computer (100). The item name associating process method extracts a plurality of item names from tabular data. Further, the item name associating process method refers to a storage unit in which a plurality of item groups is stored to determine an item group including an item name having a predetermined similarity relationship with each of the plurality of extracted item names. With respect to an item name among the plurality of item names for which a positive determination result has been obtained, the item name associating process method selects the item name having the predetermined similarity relationship as the object for association. With respect to an item name for which a negative determination result has been obtained, the item name associating process method presents an item group determined to include the item name having the predetermined similarity relationship with another item name as an association candidate.

(57) 要約:

[続葉有]



添付公開書類:

— 国際調査報告（条約第 21 条(3)）

項目名を標準化したボキャブラリに対応付けできる項目名対応付け処理方法、項目名対応付け処理プログラム及び情報処理装置を提供する。項目名対応付け処理方法は、以下の各処理をコンピュータ（１００）が実行する。項目名対応付け処理方法は、表形式データから複数の項目名を抽出する。また、項目名対応付け処理方法は、複数の項目群が記憶された記憶部を参照して、抽出した複数の項目名のそれぞれと所定の類似関係を有する項目名がどの項目群に含まれるか否かを判定する。また、項目名対応付け処理方法は、複数の項目名のうち、肯定的な判定結果が得られた項目名については、所定の類似関係を有する項目名を対応付け対象として選択する。また、項目名対応付け処理方法は、否定的な判定結果が得られた項目名については、他の項目名と所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する。

## 明 細 書

発明の名称：

項目名対応付け処理方法、項目名対応付け処理プログラム及び情報処理装置

### 技術分野

[0001] 本発明は、項目名対応付け処理方法、項目名対応付け処理プログラム及び情報処理装置に関する。

### 背景技術

[0002] 近年、例えば地方自治体が、当該地方自治体の地域内の観光地について様々な情報を集約し、インターネット上のホームページで公開することが行われている。当該地方自治体は、観光地の施設等から情報の提供を受けることで、観光地の情報を収集する。また、当該自治体の委託先企業が自治体からオープンデータとして観光地の情報の提供を受けて入力する場合もある。この場合、提供を受ける情報は、例えば、表形式データである各種表計算ソフトウェアのファイル形式、CSV（Comma-Separated Values）形式、TSV（Tab-Separated Values）形式等の様々なフォーマットに基づいた情報である。

### 先行技術文献

### 特許文献

[0003] 特許文献1：特開2013-015909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、収集された情報は、例えば、名前と氏名のように、項目名が統一されていない場合がある。このため、収集された情報の項目名を定義済みの標準化ボキャブラリに関連付けることで、項目名を統一することが考えられる。ところが、項目名に適切な標準化ボキャブラリを探し出すには、

相応の知識を有する者による探索の手間が発生する。

- [0005] 一つの側面では、本発明は、項目名を標準化したボキャブラリに対応付けできる項目名対応付け処理方法、項目名対応付け処理プログラム及び情報処理装置を提供することにある。

### 課題を解決するための手段

- [0006] 一つの態様では、項目名対応付け処理方法は、表形式データから複数の項目名を抽出する処理をコンピュータが実行する。また、項目名対応付け処理方法は、複数の項目群が記憶された記憶部を参照して、抽出した前記複数の項目名のそれぞれと所定の類似関係を有する項目名がどの項目群に含まれるか否か判定する処理をコンピュータが実行する。また、項目名対応付け処理方法は、前記複数の項目名のうち、肯定的な判定結果が得られた項目名については、前記所定の類似関係を有する項目名を対応付け対象として選択する処理をコンピュータが実行する。また、項目名対応付け処理方法は、否定的な判定結果が得られた項目名については、他の項目名と前記所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する処理をコンピュータが実行する。

### 発明の効果

- [0007] 項目名を標準化したボキャブラリに対応付けできる。

### 図面の簡単な説明

- [0008] [図1]図1は、実施例の情報処理装置の構成の一例を示すブロック図である。  
[図2]図2は、表形式データ及び表データの一例を示す図である。  
[図3]図3は、情報DBの一例を示す図である。  
[図4A]図4Aは、ボキャブラリDBの一例を示す図である。  
[図4B]図4Bは、ボキャブラリDBの一例を示す図である。  
[図4C]図4Cは、ボキャブラリDBの一例を示す図である。  
[図5]図5は、履歴DBの一例を示す図である。  
[図6]図6は、表データの抽出の一例を示す図である。  
[図7]図7は、表データの抽出の他の一例を示す図である。

[図8]図 8 は、編集処理の一例を示す図である。

[図9]図 9 は、編集処理の他の一例を示す図である。

[図10]図 10 は、編集処理の他の一例を示す図である。

[図11]図 11 は、項目行ではない行におけるセルの連結解除の一例を示す図である。

[図12]図 12 は、項目名の生成の一例を示す図である。

[図13]図 13 は、項目名の生成の他の一例を示す図である。

[図14]図 14 は、項目行の特定の一例を示す図である。

[図15]図 15 は、項目行の特定の他の一例を示す図である。

[図16]図 16 は、項目行の特定の他の一例を示す図である。

[図17]図 17 は、項目行の特定の他の一例を示す図である。

[図18]図 18 は、項目列の特定の一例を示す図である。

[図19]図 19 は、項目列の特定の他の一例を示す図である。

[図20]図 20 は、項目行の追加の一例を示す図である。

[図21]図 21 は、項目行の追加の他の一例を示す図である。

[図22]図 22 は、整形後の表データの一例を示す図である。

[図23]図 23 は、割当画面の一例を示す図である。

[図24]図 24 は、実施例の解析処理の一例を示すフローチャートである。

[図25]図 25 は、実施例の標準化処理の一例を示すフローチャートである。

[図26]図 26 は、項目名対応付け処理プログラムを実行するコンピュータの一例を示す図である。

### 発明を実施するための形態

[0009] 以下、図面に基づいて、本願の開示する項目名対応付け処理方法、項目名対応付け処理プログラム及び情報処理装置の実施例を詳細に説明する。なお、本実施例により、開示技術が限定されるものではない。また、以下の実施例は、矛盾しない範囲で適宜組みあわせてもよい。

### 実施例

[0010] 図 1 は、実施例の情報処理装置の構成の一例を示すブロック図である。図

1 に示す情報処理装置 100 は、表形式データから複数の項目名を抽出する。また、情報処理装置 100 は、複数の項目群が記憶された記憶部を参照して、抽出した複数の項目名のそれぞれと所定の類似関係を有する項目名がどの項目群に含まれるか否か判定する。また、情報処理装置 100 は、複数の項目名のうち、肯定的な判定結果が得られた項目名については、所定の類似関係を有する項目名を対応付け対象として選択する。また、情報処理装置 100 は、否定的な判定結果が得られた項目名については、他の項目名と所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する。これにより、情報処理装置 100 は、項目名を標準化したボキャブラリに対応付けできる。なお、以下の説明では、主に行方向に着目して説明するが、列方向についても同様に適用可能である。

[0011] 図 1 に示す情報処理装置 100 は、通信部 110 と、表示部 111 と、操作部 112 と、記憶部 120 と、制御部 130 とを有する。なお、情報処理装置 100 は、図 1 に示す機能部以外にも既知のコンピュータが有する各種の機能部、例えば各種の通信デバイス、入力デバイスや音声出力デバイス等の機能部を有することとしてもかまわない。情報処理装置 100 の一例としては、サーバ等の据置型のコンピュータを採用できる。情報処理装置 100 には、上記のサーバ等の据置型のコンピュータのみならず、可搬型や据置型のパーソナルコンピュータを情報処理装置 100 として採用することもできる。

[0012] 通信部 110 は、例えば、NIC (Network Interface Card) 等によって実現される。通信部 110 は、図示しないネットワークを介して図示しないユーザの端末装置と有線又は無線で接続され、端末装置との間で情報の通信を司る通信インタフェースである。通信部 110 は、端末装置から表形式データと、選択情報とを受信する。通信部 110 は、受信した表形式データと、選択情報とを制御部 130 に出力する。また、通信部 110 は、制御部 130 から割当画面が入力される。通信部 110 は、入力された割当画面を端末装置に送信する。

[0013] ここで、図2を用いて表形式データについて説明する。図2は、表形式データ及び表データの一例を示す図である。図2に示す表形式データ11は、例えば、複数の表データ12a、12bと、表形式データ11のタイトル等を含むデータである。なお、以下の説明では、例えば1つのファイル全体のデータを表形式データと表現し、表形式データ内の個々の表を表データと表現する。表形式データ11には、例えば、最上段の行に項目（ヘッダ）がある表データ、最上段の行と最左列の列に項目がある表データ、小項目を表するためにセルが連結されて項目行が2行に渡る表データ等が含まれる。なお、表データは、これらに限定されず、行列の形で表すことができるデータであれば、どのようなデータでもよい。また、表形式データは、例えば、官公庁や自治体から提供されるオープンデータを用いることができる。

[0014] 図1の説明に戻って、表示部111は、各種情報を表示するための表示デバイスである。表示部111は、例えば、表示デバイスとして液晶ディスプレイ等によって実現される。表示部111は、制御部130から入力された表示画面等の各種画面を表示する。

[0015] 操作部112は、情報処理装置100の管理者から各種操作を受け付ける入力デバイスである。操作部112は、例えば、入力デバイスとして、キーボードやマウス等によって実現される。操作部112は、管理者によって入力された操作を操作情報として制御部130に出力する。なお、操作部112は、入力デバイスとして、タッチパネル等によって実現されるようにしてもよく、表示部111の表示デバイスと、操作部112の入力デバイスとは、一体化されるようにしてもよい。

[0016] 記憶部120は、例えば、RAM（Random Access Memory）、フラッシュメモリ（Flash Memory）等の半導体メモリ素子、ハードディスクや光ディスク等の記憶装置によって実現される。記憶部120は、情報データベース121と、ボキャブラリデータベース122と、履歴データベース123とを有する。なお、以下の説明では、データベースをDBと省略して表現する。また、記憶部120は、制御部130での処理に用いる情報を記憶する。

[0017] 情報DB121は、表データについて、項目と値とボキャブラリとを対応付けて記憶する。図3は、情報DBの一例を示す図である。図3に示すように、情報DB121は、「行」、「項目」、「値」、「標準化ボキャブラリ」、「グループ」といった項目を有する。情報DB121は、例えば、表データを構成するセルごとに1レコードとして記憶する。

[0018] 「行」は、データが入力されているセルの行、つまりデータ行数を示す情報である。「項目」は、セルに対応する項目、つまり項目名を示す情報である。「値」は、セルに格納されているデータを示す情報である。「標準化ボキャブラリ」は、当該項目、つまり項目名に対応付けられた標準化ボキャブラリを示す情報である。「グループ」は、標準化ボキャブラリの属するグループを示す情報である。なお、グループは、項目群ともいう。図3の1行目の例では、表データの「1」行目の項目「x1」の値は「y1」となり、項目「x1」に対応付けられた標準化ボキャブラリは「B01」、グループは「G02」である。なお、図3の説明では、項目、値、標準化ボキャブラリ及びグループは、簡単のために記号と数字とで表しているが、実際には具体的な文字等が入る。例えば、あるレコードには、項目「住所」と、値「東京都・・・」と、標準化ボキャブラリ「住所」と、グループ「共通」とが対応付けられて記憶される。

[0019] 図1の説明に戻って、ボキャブラリDB122は、標準化ボキャブラリをグループごとに記憶する。図4Aは、ボキャブラリDBの一例を示す図である。図4Aは、標準化ボキャブラリのうち、全業種で共通する共通ボキャブラリを記憶するグループを示す。図4Aに示すように、ボキャブラリDB122の共通ボキャブラリグループ122aは、例えば、グループ名「共通ボキャブラリグループ」とともに、「タイトル」、「説明文」といった全業種で共通するボキャブラリを記憶する。なお、グループ名は省略して「共通」としてもよい。また、ボキャブラリDB122は、既存のデータベースがあれば、当該データベースを取り込んで用いてもよいし、既存のデータベースがなければ、新規に生成するようにしてもよい。



[0020] 図4Bは、ボキャブラリDBの一例を示す図である。図4Bは、標準化ボキャブラリのうち、業種ごとのボキャブラリを記憶するグループの一例である医薬品ボキャブラリグループを示す。図4Bに示すように、ボキャブラリDB122の医薬品ボキャブラリグループ122bは、例えば、グループ名「医薬品ボキャブラリグループ」とともに、「薬の名称」、「個別医薬品コード」といった医薬品に関して共通するボキャブラリを記憶する。なお、グループ名は省略して「医薬品」としてもよい。

[0021] 図4Cは、ボキャブラリDBの一例を示す図である。図4Cは、標準化ボキャブラリのうち、業種ごとのボキャブラリを記憶するグループの一例である取引ボキャブラリグループを示す。図4Cに示すように、ボキャブラリDB122の取引ボキャブラリグループ122cは、例えば、グループ名「取引ボキャブラリグループ」とともに、「取引の債権者」、「取引の債務者」といった取引に関して共通するボキャブラリを記憶する。なお、グループ名は省略して「取引」としてもよい。また、共通ボキャブラリグループ122a、医薬品ボキャブラリグループ122b及び取引ボキャブラリグループ122cは、所定の項目群の一例である。

[0022] 図1の説明に戻って、履歴DB123は、過去に手動判定で対応付けが行われた履歴を記憶する。図5は、履歴DBの一例を示す図である。図5に示すように、履歴DB123は、「項目名」、「標準化ボキャブラリ」、「グループ」といった項目を有する。履歴DB123は、例えば、項目名ごとに1レコードとして記憶する。言い換えると、履歴DB123は、採用された候補と否定的な判定結果が得られた項目名との対応関係を記憶する。なお、履歴DB123は、全てのユーザの履歴を用いることで、より適切な標準化ボキャブラリを提示させることができる。

[0023] 「項目名」は、表形式データから抽出され、手動判定が行われた項目名を示す情報である。「標準化ボキャブラリ」は、手動判定が行われた結果、採用された標準化ボキャブラリを示す情報である。「グループ」は、標準化ボキャブラリが属するグループを示す情報である。図5の1行目の例では、項

目名「TEL」に対して、グループ「共通」に属する標準化ボキャブラリ「電話番号」が手動判定にて採用されたことを示す。

[0024] 図1の説明に戻って、制御部130は、例えば、CPU (Central Processing Unit) やMPU (Micro Processing Unit) 等によって、内部の記憶装置に記憶されているプログラムがRAMを作業領域として実行されることにより実現される。また、制御部130は、例えば、ASIC (Application Specific Integrated Circuit) やFPGA (Field Programmable Gate Array) 等の集積回路により実現されるようにしてもよい。

[0025] 制御部130は、判定部131と、抽出部132と、編集部133と、カウント部134と、生成部135と、検出部136と、特定部137と、記憶制御部138とを有する。また、制御部130は、項目群判定部139と、提示部140と、対応関係記憶制御部141と有し、以下に説明する情報処理の機能や作用を実現又は実行する。なお、制御部130の内部構成は、図1に示した構成に限られず、後述する情報処理を行う構成であれば他の構成であってもよい。

[0026] 判定部131は、通信部110から表形式データが入力されると、入力された表形式データの各行又は各列について、データが入力されたセルの有無を判定する。すなわち、判定部131は、表形式データのデータ入力セルの有無を判定する。判定部131は、表形式データ及び判定結果を抽出部132に出力する。

[0027] 抽出部132は、判定部131から表形式データ及び判定結果が入力されると、判定結果に基づいて、表形式データから、データが入力されたセルが存在する複数の連続する行又は列の塊を1つの表データに関連する部分として抽出する。すなわち、抽出部132は、データが入力されたセルが存在しない1又は複数の連続する行又は列を挟んで、データが入力されたセルが存在する1又は複数の連続する行又は列の塊を2つ検出すると、該2つの塊のそれぞれを異なる表データとして抽出する。抽出部132は、表データを抽出すると、抽出した表データを第1表データとして編集部133及び生成部

135に出力する。また、抽出部132は、第1表データを記憶部120に記憶する。

[0028] ここで、図6及び図7を用いて表データの抽出について説明する。図6は、表データの抽出の一例を示す図である。図6は、表データが縦方向に複数存在する場合の一例である。図6の例では、表形式データ13について、各行のデータ入力数14を検出する。例えば、表形式データ13では、1行目は表形式データ13のタイトルが1つのセルに入力されているので、データ入力数14は「1」となる。また、2行目は、データが入力されているセルがないので、データ入力数14は「0」となる。以下同様に、各行のデータ入力数14を検出する。

[0029] 抽出部132は、データ入力数14が「0」である行を表データの切れ目と判定し、表形式データ13を切れ目で分割する。なお、以下の説明では、分割された表データに関する部分である塊をクラスタともいう。表形式データ13は、クラスタ15と、クラスタ16と、クラスタ17とに分割される。クラスタ15は、表形式データ13のタイトルである。クラスタ16は、1つ目の表データである。クラスタ17は、2つ目の表データである。抽出部132は、クラスタ16及びクラスタ17を第1表データとして抽出する。なお、抽出された第1表データは、例えばメモリ上で二次元配列等を用いて表形式に展開される。また、以下の説明において、第1表データに基づく各表データも同様である。

[0030] 図7は、表データの抽出の他の一例を示す図である。図7は、表データが横方向に複数存在する場合の一例である。図7の例では、表形式データ18について、各列のデータ入力数19を検出する。例えば、表形式データ18では、1列目はデータが入力されているセルがないので、データ入力数19は「0」となる。また、2列目は、1行目に表形式データ18のタイトル、2行目に「a」、3行目に「1」、4行目に「1」が入力されているので、データ入力数19は「4」となる。以下同様に、各列のデータ入力数19を検出する。

[0031] 抽出部132は、データ入力数19が「0」である列を表データの切れ目と判定し、表形式データ18を切れ目で分割する。表形式データ18は、クラスタ20と、クラスタ21とに分割される。クラスタ20は、1つ目の表データである。クラスタ21は、2つ目の表データである。抽出部132は、クラスタ20及びクラスタ21を第1表データとして抽出する。なお、クラスタ20は、クラスタ21と比べて、5行目にデータ入力セルがないが、5行目に空文字を付加して表の大きさを揃えている。

[0032] 図1の説明に戻って、編集部133は、抽出部132から第1表データが入力されると、入力された第1表データに対して編集処理を実行する。編集部133は、まず、第1表データのうち、タイトルのセルを除く表を構成するセルの中で、最も上側の行又は最も左側の列を項目行又は項目列と仮に決定する。なお、タイトルのセルは、第1表データの中で、抽出部132で用いたデータ入力数が「1」である最も上側又は最も左側の行又は列として判定できる。編集部133は、仮に決定した項目行又は項目列にセル連結処理が施された特定のセルが含まれる場合には、特定のセルを単位セルに分割する。また、編集部133は、分割された単位セルのそれぞれに特定のセルに入力されたデータと同じデータを入力する。編集部133は、編集処理を完了した表データを第2表データとしてカウント部134及び生成部135に出力する。また、編集部133は、仮に決定した項目行又は項目列にセル連結処理が施された特定のセルが含まれない場合には、入力された第1表データをそのまま第2表データとしてカウント部134及び生成部135に出力する。

[0033] ここで、図8から図11を用いて編集処理について説明する。図8は、編集処理の一例を示す図である。図8の例では、第1表データ22の1行目にセル連結処理が施されたセルが含まれている。つまり、値「a」、「b」のセルが、セル連結処理が施された特定のセルである。編集部133は、特定のセルを単位セルに分割し、値「a」、「b」を、分割された単位セルのそれぞれに入力する。編集部133は、編集処理を完了した第2表データ23

をカウント部 134 及び生成部 135 に出力する。

[0034] 図 9 は、編集処理の他の一例を示す図である。図 9 の例では、図 8 の例と同様に、第 1 表データ 24 のセル連結処理が施された特定のセルの値「a」、「b」を、分割された単位セルのそれぞれに入力して第 2 表データ 25 としている。

[0035] 図 10 は、編集処理の他の一例を示す図である。図 10 の例では、第 1 表データ 26 の 1 列目にセル連結処理が施されたセルが含まれている。つまり、値「g」、「h」のセルが、セル連結処理が施された特定のセルである。編集部 133 は、特定のセルを単位セルに分割し、値「g」、「h」を、分割された単位セルのそれぞれに入力する。編集部 133 は、編集処理を完了した第 2 表データ 27 をカウント部 134 及び生成部 135 に出力する。すなわち、編集部 133 は、行方向のセル連結処理が施された特定のセルと、列方向のセル連結処理が施された特定のセルとを単位セルに分割し、分割された単位セルのそれぞれに特定のセルの値を入力する。

[0036] 図 11 は、項目行ではない行におけるセルの連結解除の一例を示す図である。図 11 の例では、第 1 表データ 28 の最終行、つまり 4 行目にセル連結処理が施されたセルが含まれている。つまり、値「100」のセルが、セル連結処理が施された特定のセルである。編集部 133 は、第 1 表データ 28 の最終行が項目行ではないので、特定のセルを単位セルに分割し、値「100」を分割した単位セルのうち 1 つのセルに入力する。編集部 133 は、編集処理を完了した第 2 表データ 29 をカウント部 134 及び生成部 135 に出力する。なお、図 11 の例では、1 行目のセルの連結処理の解除の説明は省略している。なお、項目行ではない行におけるセルの連結解除は、特定部 137 において項目行又は項目列が特定された後に行ってもよい。

[0037] 図 1 の説明に戻って、カウント部 134 は、編集部 133 から第 2 表データが入力されると、第 2 表データのうち、データが入力されたセルの数を行ごと又は列ごとにカウントする。すなわち、カウント部 134 は、行又は列の塊のうち、データが入力されたセルの数を行ごと又は列ごとにカウントす

る。カウント部 134 は、行ごと又は列ごとにカウントしたセルの数をカウント値として検出部 136 に出力する。

[0038] 生成部 135 には、抽出部 132 から第 1 表データが入力され、編集部 133 から第 2 表データが入力される。生成部 135 は、まず、入力された第 1 表データのうち、タイトルのセルを除く表を構成するセルの中で、最も上側の行又は最も左側の列を項目行又は項目列と仮に決定する。なお、タイトルのセルは、編集部 133 と同様に判定できる。生成部 135 は、仮に決定した項目行又は項目列にセル連結処理が施された特定のセルが含まれる場合には、特定のセルを含む範囲を連続する複数の項目行又は連続する複数の項目列と仮に決定する。すなわち、生成部 135 は、特定のセルを分割したそれぞれの単位セルが含まれる行又は列と、当該行又は列の下側又は右側に隣接する行又は列とを、連続する複数の項目行又は連続する複数の項目列と仮に決定する。

[0039] 生成部 135 は、連続する複数の項目行又は連続する複数の項目列を仮に決定すると、編集部 133 から入力された第 2 表データに対して、項目名の生成を実行する。すなわち、生成部 135 は、仮に決定した連続する複数の項目行若しくは連続する複数の項目列について、同じ列若しくは同じ列のセルを含む連結セルの値、又は、同じ行若しくは同じ行のセルを含む連結セルの値を合成した値を項目名として生成する。なお、連結セルは、セル連結処理が施された特定のセルのことである。生成部 135 は、生成した項目名を適用した第 2 表データを第 3 表データとして検出部 136 に出力する。生成部 135 は、仮に決定した項目行又は項目列にセル連結処理が施された特定のセルが含まれない場合には、入力された第 2 表データをそのまま第 3 表データとして検出部 136 に出力する。

[0040] ここで、図 12 及び図 13 を用いて項目名の生成について説明する。図 12 は、項目名の生成の一例を示す図である。図 12 の例では、生成部 135 は、第 1 表データ 30 について、1 行目及び 2 行目を項目行、並びに、1 列目及び 2 列目を項目列と仮に決定する。次に、生成部 135 は、1 行目及び

2行目において、同じ列若しくは同じ列のセルを含む連結セルの値を合成した値を、項目行のそれぞれの項目名として生成する。なお、合成した値は、セル連結処理が施された特定のセルについてセル連結処理が解除された図示しない第2表データに基づいて生成される。生成部135は、例えば、第1表データ30の1行3列の「b」と、2行3列の「f」とを合成した「b／f」を、第3表データ31の1行2列の項目名として生成する。

[0041] また、生成部135は、1列目及び2列目において、同じ行若しくは同じ行のセルを含む連結セルの値を合成した値を、項目列のそれぞれの項目名として生成する。生成部135は、例えば、第1表データ30の3行1列の「j」と、3行2列の「m」とを合成した「j／m」を、第3表データ31の2行1列の項目名として生成する。なお、第1表データ30では、1行1列、1行2列、2行1列及び2行2列の4つのセルは連結されて値が「a」であるので、第3表データ31では、1行1列の項目名を「a」としている。

[0042] 図13は、項目名の生成の他の一例を示す図である。図13の例では、生成部135は、第1表データ32について、1行目及び2行目を項目行と仮に決定する。次に、生成部135は、1行目及び2行目において、同じ列若しくは同じ列のセルを含む連結セルの値を合成した値を、項目行のそれぞれの項目名として生成する。なお、合成した値は、セル連結処理が施された特定のセルについてセル連結処理が解除された図示しない第2表データに基づいて生成される。生成部135は、例えば、第1表データ32の1行1列の「a」と、2行1列の「d」とを合成した「a／d」を、第3表データ33の1行1列の項目名として生成する。また、生成部135は、例えば、第1表データ32の1行2列の「a」と、2行2列の「e」とを合成した「a／e」を、第3表データ33の1行2列の項目名として生成する。

[0043] 図1の説明に戻って、検出部136には、カウント部134からカウント値が入力され、生成部135から第3表データが入力される。検出部136は、入力された第3表データに対して、入力されたカウント値が最大の行又は列のうち、最も上側の行又は最も左側の列を検出する。検出部136は、

検出された最も上側の行又は最も左側の列を検出結果として、カウント値及び第3表データとともに特定部137に出力する。

[0044] 特定部137には、検出部136から検出結果、カウント値及び第3表データが入力される。特定部137は、カウント値及び第3表データに基づいて、カウント値が最大の行又は列のうち、最も上側の行又は最も左側の列を表の項目を示す行又は列として特定する。すなわち、特定部137は、項目行又は項目列を特定する。特定部137は、特定が完了した第3表データを第4表データとする。特定部137は、特定した項目行又は項目列と、第4表データとを記憶制御部138に出力する。

[0045] また、特定部137は、検出結果、カウント値及び第3表データに基づいて、項目行又は項目列を特定してもよい。特定部137は、検出した最も上側の行に隣接する下側の行に対応するカウント値が最大でない場合に、最も上側の行を表の項目を示す行として特定する。又は、特定部137は、検出した最も左側の列に隣接する右側の列に対応するカウント値が最大でない場合に、最も左側の列を表の項目を示す列として特定する。すなわち、特定部137は、項目行又は項目列を特定する。特定部137は、特定が完了した第3表データを第4表データとする。特定部137は、特定した項目行又は項目列と、第4表データとを記憶制御部138に出力する。

[0046] さらに、特定部137は、複数行が同じカウント値である場合に、非数値データが入力されたセルの割合に基づいて、項目行又は項目列を特定してもよい。特定部137は、検出した最も上側の行を含み連続する複数行が同じカウント値である場合に、複数行について行内のセルのうち非数値データが入力されたセルの割合に基づいて、項目を示す行を特定する。又は、特定部137は、検出した最も左側の列を含み連続する複数列が同じカウント値である場合に、複数列について列内のセルのうち非数値データが入力されたセルの割合に基づいて、項目を示す列として特定する。すなわち、特定部137は、項目行又は項目列を特定する。特定部137は、特定が完了した第3表データを第4表データとする。特定部137は、特定した項目行又は項目



列と、第4表データとを記憶制御部138に出力する。

[0047] また、特定部137は、編集部133で仮に決定された項目行又は項目列を採用して項目行又は項目列を特定してもよい。さらに、特定部137は、生成部135で仮に決定された連続する複数の項目行又は連続する複数の項目列を採用して、項目行又は項目列を特定してもよい。特定部137は、特定が完了した第3表データを第4表データとする。特定部137は、特定した項目行又は項目列と、第4表データとを記憶制御部138に出力する。

[0048] また、特定部137は、第3表データが、項目行又は項目列が存在しない表である場合に、最も上側の行又は最も左側の列を項目行又は項目列とみなして項目行又は項目列を特定してもよい。特定部137は、カウント値が最大の行又は列のうち、最も上側の行又は最も左側の列が、入力されたデータが項目名ではないセルを含む場合であっても、最も上側の行又は最も左側の列を、項目行又は項目列として特定する。特定部137は、特定が完了した第3表データを第4表データとする。特定部137は、特定した項目行又は項目列と、第4表データとを記憶制御部138に出力する。

[0049] さらに、特定部137は、入力されたデータが重複するセルを含む場合に、新たな項目行又は項目列を追加してもよい。特定部137は、カウント値が最大の行又は列のうち、最も上側の行又は最も左側の列が、入力されたデータが重複するセルを含む場合に、最も上側の行のさらに上側又は最も左側の列のさらに左側に新たな行又は列を追加する。特定部137は、追加した行又は列を項目行又は項目列として特定する。特定部137は、新たな行又は列が追加されて特定が完了した第3表データを第4表データとする。特定部137は、特定した項目行又は項目列と、第4表データとを記憶制御部138に出力する。

[0050] また、特定部137は、最も上側の行又は最も左側の列が空白のセルを含む場合に、新たな項目行又は項目列を追加してもよい。なお、空白のセルは、空文字（NULL）で表される。特定部137は、カウント値が最大の行又は列のうち、最も上側の行又は最も左側の列が、空白のセルを含む場合に

、最も上側の行のさらに上側又は最も左側の列のさらに左側に新たな行又は列を追加する。特定部 1 3 7 は、追加した行又は列を項目行又は項目列として特定する。特定部 1 3 7 は、新たな行又は列が追加されて特定が完了した第 3 表データを第 4 表データとする。特定部 1 3 7 は、特定した項目行又は項目列と、第 4 表データとを記憶制御部 1 3 8 に出力する。

[0051] ここで、図 1 4 から図 2 2 を用いて、項目行の特定について説明する。図 1 4 は、項目行の特定の一例を示す図である。図 1 4 の例は、カウント値が最大の行が 1 つの場合に、項目行を特定する場合である。第 3 表データ 3 4 は、カウント値 3 5 を見ると、2 行目が「5」であり最大である。特定部 1 3 7 は、2 行目に隣接する下側の行である 3 行目のカウント値が「4」であり最大でないので、2 行目を項目行と特定する。

[0052] 図 1 5 は、項目行の特定の他の一例を示す図である。図 1 5 の例は、カウント値が最大の行が複数の場合に、項目行を特定する場合である。第 3 表データ 3 7 は、カウント値 3 8 を見ると、2 行目及び 5 行目が「5」であり最大である。特定部 1 3 7 は、カウント値が最大の行のうち、最も上側の行である 2 行目を項目行と特定する。

[0053] 図 1 6 は、項目行の特定の他の一例を示す図である。図 1 6 の例は、非数値データが入力されたセルの割合に基づいて、項目行を特定する場合である。第 3 表データ 4 1 は、カウント値 4 2 を見ると、2 行目及び 3 行目が「5」であり最大である。なお、他の行のカウント値 4 2 は省略している。また、第 3 表データ 4 1 は、非数値データが入力されたセルの割合 4 3 は、2 行目が 1 0 0 %、3 行目が 4 0 % である。特定部 1 3 7 は、2 行目に隣接する 3 行目の当該割合が、例えば 5 0 % 以上であるか否かを判定する。特定部 1 3 7 は、3 行目の当該割合が 4 0 % であるので、3 行目は項目行でないと判定し、2 行目を項目行と特定する。

[0054] 図 1 7 は、項目行の特定の他の一例を示す図である。図 1 7 の例は、非数値データが入力されたセルの割合に基づいて、項目行を特定する場合である。第 3 表データ 4 6 は、カウント値 4 7 を見ると、2 行目及び 3 行目が「5

」であり最大である。なお、他の行のカウント値 47 は省略している。また、第 3 表データ 46 は、非数値データが入力されたセルの割合 48 は、2 行目が 100%、3 行目が 60% である。特定部 137 は、2 行目に隣接する 3 行目の当該割合が、例えば 50% 以上であるか否かを判定する。特定部 137 は、3 行目の当該割合が 60% であるので、3 行目は項目行であると判定し、2 行目及び 3 行目を項目行と特定する。なお、項目行に入る数値データは、例えば、交通手段の数等が挙げられる。

[0055] 図 18 は、項目列の特定の一例を示す図である。図 18 の例は、項目列が存在しない表である場合に、最も左側の列を項目列とみなして項目列を特定する場合である。第 3 表データ 51 は、1 行目が項目行であるが、2 行目以降の 1 列目のセルには、データが入力されている。この場合に、特定部 137 は、最も左側の列である 1 列目を項目列とみなして、1 列目を項目列と特定する。

[0056] 図 19 は、項目列の特定の他の一例を示す図である。図 19 の例は、項目列が存在しない表である場合に、最も左側の列を項目列とみなして項目列を特定する場合である。第 1 表データ 53 は、1 行目が項目行であるが、2 行目以降の 1 列目のセルには、データが入力されている。また、第 1 表データ 53 は、1 行 1 列のセルと 1 行 2 列のセルとが連結された特定のセルである。この場合に、特定部 137 は、1 列目に特定のセルが含まれるので、特定のセルが含まれる列、つまり 1 列目及び 2 列目を項目列とみなして、1 列目及び 2 列目を項目列と特定する。なお、特定部 137 は、検出結果、カウント値及び第 3 表データの他に、記憶部 120 に記憶された第 1 表データを参照して、項目列を特定する。

[0057] 図 20 は、項目行の追加の一例を示す図である。図 20 の例は、入力されたデータが重複するセルを含む場合に、新たな項目行又は項目列を追加する場合である。第 3 表データ 56 は、1 行 1 列と 1 行 2 列のデータがともに「a」であり、1 行目が入力されたデータが重複するセルを含む。この場合に、特定部 137 は、最も上側の行のさらに上側に新たな行を追加して第 4 表

データ 5 8 とする。特定部 1 3 7 は、第 4 表データ 5 8 の追加した行 5 9 を項目行として特定する。

[0058] 図 2 1 は、項目行の追加の他の一例を示す図である。図 2 1 の例は、最も上側の行が空白のセルを含む場合に、新たな項目行を追加する場合である。第 3 表データ 6 0 は、1 行 3 列のセルが空白である。この場合に、特定部 1 3 7 は、最も上側の行のさらに上側に新たな行を追加して第 4 表データ 6 2 とする。特定部 1 3 7 は、第 4 表データ 6 2 の追加した行 6 3 を項目行として特定する。なお、第 3 表データ 6 0 は、他の行にも空白のセルがあり、1 行目がカウント値が最大の行に含まれる場合である。このような場合には、2 行目以降の行が項目行と誤認識されないので適用可能である。

[0059] 図 2 2 は、整形後の表データの一例を示す図である。図 2 2 に示す第 4 表データ 6 4 は、特定部 1 3 7 で項目行又は項目列が特定された後、つまり整形後の表データである。第 4 表データ 6 4 は、項目行 6 5 と、データ行数 6 6 と、データ部分 6 7 とを有する。すなわち、第 4 表データ 6 4 は、データ（セルの値）1 つ 1 つに対して行数と項目名との関連付けが行われた状態である。なお、データ行数 6 6 は、第 4 表データ 6 4 に含まれなくてもよく、情報 DB 1 2 1 への記憶時に行数をカウントして付加してもよい。

[0060] 図 1 の説明に戻って、記憶制御部 1 3 8 には、特定部 1 3 7 から特定した項目行又は項目列と、第 4 表データとが入力される。記憶制御部 1 3 8 は、特定した項目行又は項目列と、第 4 表データとに基づいて、項目行又は項目列の各セルの入力データを項目名とし、各行又は各列の値を対応する項目名とデータ行数とに対応付けて、情報 DB 1 2 1 に記憶する。記憶制御部 1 3 8 は、データ行数と項目名と値とを対応付けて情報 DB 1 2 1 に記憶すると、項目群判定部 1 3 9 に判定指示を出力する。なお、判定部 1 3 1 から記憶制御部 1 3 8 の各部は、表形式データから特定した項目行又は項目列の各セルの入力データを項目名として抽出する項目名抽出部に対応する。

[0061] 項目群判定部 1 3 9 は、記憶制御部 1 3 8 から判定指示が入力されると、情報 DB 1 2 1、ボキャブラリ DB 1 2 2 及び履歴 DB 1 2 3 を参照して、

各項目名に対応する項目群（グループ）を判定する。すなわち、項目群判定部 139 は、複数の項目群が記憶されたボキャブラリ DB 122 及び履歴 DB 123 を参照して、情報 DB 121 に記憶された複数の項目名のそれぞれと所定の類似関係を有する項目名がどの項目群に含まれるか否かを判定する。

[0062] 具体的には、項目群判定部 139 は、例えば、情報 DB 121 の 1 レコード目から順に項目名を読み込み、項目名を整形する。項目群判定部 139 は、読み込んだ項目名から、例えば、カッコ書き等の注釈的な要素や、項目名の前後の空白を取り除くことで項目名を整形する。項目群判定部 139 は、例えば、「公共交通機関（JR）」という項目名を読み込むと、「公共交通機関」に整形する。

[0063] 項目群判定部 139 は、ボキャブラリ DB 122 を参照し、整形後の項目名を用いて、標準化ボキャブラリとのマッチングを実行する。項目群判定部 139 は、項目名が標準化ボキャブラリとマッチングしたか否かを判定する。項目群判定部 139 は、マッチング時に、項目名と標準化ボキャブラリとが完全一致した場合、又は、部分一致した場合に、マッチングしたと判定する。項目群判定部 139 は、例えば、項目名が「公共交通機関」である場合に、標準化ボキャブラリが「公共交通機関」であれば完全一致と判定し、標準化ボキャブラリが「交通機関」であれば部分一致と判定する。

[0064] 項目群判定部 139 は、項目名が標準化ボキャブラリとマッチングした場合には、マッチングした標準化ボキャブラリを採用する。項目群判定部 139 は、採用した標準化ボキャブラリと、標準化ボキャブラリが属するグループとを情報 DB 121 に記憶する。

[0065] 項目群判定部 139 は、項目名が標準化ボキャブラリとマッチングしない場合には、項目名を履歴 DB 123 と照合する。つまり、項目群判定部 139 は、項目名を過去の手動判定の対応付け履歴とのマッチングを実行する。項目群判定部 139 は、項目名が履歴 DB 123 の項目名とマッチングしたか否かを判定する。このとき、項目群判定部 139 は、項目名が履歴 DB 123 の項目名と完全一致した場合に、マッチングしたと判定する。項目群判

定部 1 3 9 は、例えば、項目名が「バス」である場合に、履歴 DB 1 2 3 の項目名の「バス」と完全一致するとマッチングしたと判定する。

[0066] 項目群判定部 1 3 9 は、項目名が履歴 DB 1 2 3 の項目名とマッチングした場合には、履歴 DB 1 2 3 の標準化ボキャブラリを採用する。例えば、項目群判定部 1 3 9 は、履歴 DB 1 2 3 の項目名「バス」に対応付けられている標準化ボキャブラリ「公共交通機関」を採用する。項目群判定部 1 3 9 は、採用した標準化ボキャブラリと、標準化ボキャブラリが属するグループとを情報 DB 1 2 1 に記憶する。項目群判定部 1 3 9 は、項目名が履歴 DB 1 2 3 の項目名とマッチングしない場合には、項目名を手動判定ストックに追加する。なお、手動判定ストックは、記憶部 1 2 0 に設けられる記憶領域である。

[0067] 言い換えると、項目群判定部 1 3 9 は、ボキャブラリ DB 1 2 2 及び履歴 DB 1 2 3 を参照し、項目名と完全一致又は部分一致したという所定の類似関係を有する標準化ボキャブラリが、どのグループ（項目群）に含まれるかを判定する。項目群判定部 1 3 9 は、肯定的な判定結果であると、採用した標準化ボキャブラリと、標準化ボキャブラリが属するグループとを情報 DB 1 2 1 に記憶する。項目群判定部 1 3 9 は、否定的な判定結果であると、標準化ボキャブラリとグループとを情報 DB 1 2 1 に記憶しない。なお、参照する履歴 DB 1 2 3 は、業種ごとのグループに応じた履歴を記憶するデータベースとしてもよい。この場合には、項目群判定部 1 3 9 は、当該業種の履歴を優先して判定できる。

[0068] 項目群判定部 1 3 9 は、全ての項目名のマッチングが完了したか否かを判定する。項目群判定部 1 3 9 は、全ての項目名のマッチングが完了していない場合には、情報 DB 1 2 1 の次のレコードの項目名についてマッチングを繰り返す。項目群判定部 1 3 9 は、全ての項目名のマッチングが完了した場合には、提示指示を提示部 1 4 0 に出力する。

[0069] 提示部 1 4 0 は、項目群判定部 1 3 9 から提示指示が入力されると、肯定的な判定結果が得られた項目名、つまり、情報 DB 1 2 1 に標準化ボキャブ

ラリ及びグループが記憶されている項目名は、記憶されている標準化ボキャブラリを対応付け対象として選択する。すなわち、提示部１４０は、項目名に対してボキャブラリＤＢ１２２又は履歴ＤＢ１２３に記憶される標準化ボキャブラリを自動的に選択する。提示部１４０は、選択した標準化ボキャブラリ及びグループと対応する項目名とを対応関係記憶制御部１４１に出力する。

[0070] 提示部１４０は、否定的な判定結果が得られた項目名、つまり、手動判定ストックに記憶された項目名については、表内の他の項目名と完全一致又は部分一致したという所定の類似関係を有する標準化ボキャブラリが含まれるグループを対応付け候補として提示する。すなわち、提示部１４０は、手動判定ストックに記憶された項目名に対して標準化ボキャブラリ候補、つまり対応付け候補を提示する割当画面を、通信部１１０を介して、図示しない端末装置に送信して表示させる。

[0071] 提示部１４０は、通信部１１０を介して、図示しない端末装置から選択情報を受信する。提示部１４０は、選択情報を受け付け、選択を受け付けた標準化ボキャブラリ及びグループと対応する項目名とを対応関係記憶制御部１４１に出力する。

[0072] また、提示部１４０は、対応付け候補を提示する際に、さらに、所定の項目群を他の対応付け候補として提示してもよい。提示部１４０は、例えば、グループ「医薬品」に属する標準化ボキャブラリに加えて、グループ「共通」に属する標準化ボキャブラリを対応付け候補として提示してもよい。

[0073] さらに、提示部１４０は、複数の項目名は、表形式データから検出された複数の表の中から抽出されたものである場合に、複数の表のうち同じ表から抽出された他の項目名と所定の類似関係を有する項目名が含まれるグループを対応付け候補として提示してもよい。すなわち、提示部１４０は、同じ表から抽出された他の項目名と完全一致又は部分一致する標準化ボキャブラリが含まれるグループを対応付け候補として提示してもよい。

[0074] また、提示部１４０は、同じ表に含まれる他の項目名でマッチングした標

標準化ボキャブラリが属するグループを優先して提示するようにしてもよい。すなわち、提示部 140 は、否定的な判定結果が得られた項目名と同じ表に含まれる項目名と所定の類似関係を有する第 1 の項目名が含まれると判定され、かつ、否定的な判定結果が得られた項目名と異なる表に含まれる第 2 の項目名と所定の類似関係を有する項目名が含まれると判定された場合に、否定的な判定結果が得られた項目名について、第 1 の項目名が含まれる項目群（グループ）を第 2 の項目名が含まれる項目群よりも優先して対応付け候補として提示する。

[0075] さらに、提示部 140 は、表形式データ又は他の表形式データから抽出した項目名が、履歴 DB 123 に記憶される対応関係によって特定の項目名に対応付けられている場合は、当該特定の項目名を対応付け対象として選択するようにしてもよい。なお、特定の項目名は、対応関係に係る標準化ボキャブラリである。

[0076] また、提示部 140 は、否定的な判定結果が得られた他の項目名の対応付け候補を提示する際に、特定の項目名（対応関係に係る標準化ボキャブラリ）が含まれる項目群（グループ）を、他の項目群よりも優先して対応付け候補として提示するようにしてもよい。

[0077] ここで、図 23 を用いて、割当画面の一例について説明する。図 23 は、割当画面の一例を示す図である。図 23 に示すように、割当画面 70 は、未確定項目名欄 71 と、ボキャブラリ候補欄 72 と、決定ボタン 73 とを有する。未確定項目名欄 71 には、手動判定ストックに記憶された項目名、つまり、対応する標準化ボキャブラリが未確定の項目名が表示される。ボキャブラリ候補欄 72 には、グループごとに標準化ボキャブラリの対応付け候補が表示される。決定ボタン 73 は、例えば、対応付け候補の先頭に設けられたラジオボタンが選択された状態で押下されると、当該ラジオボタンが選択された対応付け候補を、標準化ボキャブラリとして選択情報を送信するためのボタンである。

[0078] 図 23 の例では、ボキャブラリ候補欄 72 には、先頭に、医薬品ボキャブ



ラリグループ 74 に属する「薬の名称」、「個別医薬品コード」、「JAN コード」といった対応付け候補、つまり標準化ボキャブラリの候補が表示される。なお、医薬品ボキャブラリグループ 74 は、他の項目名においてマッチングしたボキャブラリグループである。また、ボキャブラリ候補欄 72 には、他の項目名においてマッチングしたボキャブラリグループが複数ある場合には、マッチングした数の多いグループから順に表示される。例えば、ボキャブラリ候補欄 72 には、医薬品ボキャブラリグループ 74 よりもマッチング数が少ない△△ボキャブラリグループ 75 が、医薬品ボキャブラリグループ 74 の次に表示される。すなわち、提示部 140 は、処理中の表にマッチングする可能性の高い標準化ボキャブラリのグループを、ボキャブラリ候補欄 72 に表示する。

[0079] また、ボキャブラリ候補欄 72 には、△△ボキャブラリグループ 75 の次に、共通ボキャブラリグループ 76 が表示される。すなわち、ボキャブラリ候補欄 72 には、他の項目名においてマッチングしたボキャブラリグループに続いて、次にマッチングする可能性が高いボキャブラリグループとして、共通ボキャブラリグループ 76 が表示される。

[0080] さらに、ボキャブラリ候補欄 72 には、共通ボキャブラリグループ 76 の次に、その他全てのボキャブラリグループ 77 が表示される。また、その他全てのボキャブラリグループ 77 は、選択しやすいように、例えば、取引ボキャブラリグループ 77 a、製品・物品ボキャブラリグループ 77 b といったように、ボキャブラリグループごとに括りを分けて表示される。

[0081] 図 1 の説明に戻って、対応関係記憶制御部 141 は、提示部 140 から選択された標準化ボキャブラリ及びグループと対応する項目名とが入力されると、項目名と標準化ボキャブラリ及びグループとを対応付けて情報 DB 121 に記憶する。なお、この場合には、項目名と標準化ボキャブラリ及びグループとが、項目群判定部 139 によって情報 DB 121 に記憶されているので、上書きしてもよいし、情報 DB 121 の該当するレコードを読み込んで確認するようにしてもよい。

[0082] 対応関係記憶制御部 141 は、提示部 140 から選択を受け付けた標準化ボキャブラリ及びグループと対応する項目名とが入力されると、項目名と標準化ボキャブラリ及びグループとを対応付けて情報 DB 121 に記憶する。また、対応関係記憶制御部 141 は、項目名と標準化ボキャブラリ及びグループとを対応付けて履歴 DB 123 に記憶する。すなわち、対応関係記憶制御部 141 は、否定的な判定結果が得られた項目名について提示した対応付け候補のうち、採用された候補と否定的な判定結果が得られた項目名との対応関係を履歴 DB 123 に記憶する。

[0083] 次に、実施例の情報処理装置 100 の動作について説明する。まず、解析処理について説明する。図 24 は、実施例の解析処理の一例を示すフローチャートである。

[0084] 情報処理装置 100 の通信部 110 は、図示しない端末装置から表形式データを受信する。通信部 110 は、受信した表形式データを制御部 130 に出力する。判定部 131 は、通信部 110 から表形式データが入力されると、入力された表形式データのデータ入力セルの有無を判定する（ステップ S1）。判定部 131 は、表形式データ及び判定結果を抽出部 132 に出力する。

[0085] 抽出部 132 は、判定部 131 から表形式データ及び判定結果が入力されると、判定結果に基づいて、表形式データから、データが入力されたセルが存在する複数の連続する行又は列の塊を 1 つの表データとして抽出する（ステップ S2）。抽出部 132 は、表データを抽出すると、抽出した表データを第 1 表データとして編集部 133 及び生成部 135 に出力する。また、抽出部 132 は、第 1 表データを記憶部 120 に記憶する。

[0086] 編集部 133 は、抽出部 132 から第 1 表データが入力されると、入力された第 1 表データに対して編集処理を実行する（ステップ S3）。編集部 133 は、編集処理を完了した表データを第 2 表データとしてカウント部 134 及び生成部 135 に出力する。

[0087] カウント部 134 は、編集部 133 から第 2 表データが入力されると、第

2表データのうち、データが入力されたセルの数を行ごと又は列ごとにカウントする（ステップS4）。カウント部134は、行ごと又は列ごとにカウントしたセルの数をカウント値として検出部136に出力する。

[0088] 生成部135には、抽出部132から第1表データが入力され、編集部133から第2表データが入力される。生成部135は、入力された第1表データに基づいて、項目行又は項目列を仮に決定する。生成部135は、仮に決定した項目行又は項目列にセル連結処理が施された特定のセルが含まれる場合には、特定のセルに対応する連続する複数の項目行又は連続する複数の項目列を仮に決定する。生成部135は、連続する複数の項目行又は連続する複数の項目列を仮に決定すると、編集部133から入力された第2表データに対して、項目名の生成を実行する（ステップS5）。生成部135は、生成した項目名を適用した第2表データを第3表データとして検出部136に出力する。生成部135は、仮に決定した項目行又は項目列にセル連結処理が施された特定のセルが含まれない場合には、入力された第2表データをそのまま第3表データとして検出部136に出力する。

[0089] 検出部136には、カウント部134からカウント値が入力され、生成部135から第3表データが入力される。検出部136は、入力された第3表データに対して、入力されたカウント値が最大の行又は列のうち、最も上側の行又は最も左側の列を検出する（ステップS6）。検出部136は、検出された最も上側の行又は最も左側の列を検出結果として、カウント値及び第3表データとともに特定部137に出力する。

[0090] 特定部137には、検出部136から検出結果、カウント値及び第3表データが入力される。特定部137は、検出結果、カウント値及び第3表データに基づいて、項目行又は項目列を特定する（ステップS7）。特定部137は、特定が完了した第3表データを第4表データとし、特定した項目行又は項目列と、第4表データとを記憶制御部138に出力する。

[0091] 記憶制御部138には、特定部137から特定した項目行又は項目列と、第4表データとが入力される。記憶制御部138は、特定した項目行又は項

目列と、第4表データとに基づいて、第4表データの各セルの値を項目名とデータ行数とに対応付けて、情報DB121に記憶する（ステップS8）。記憶制御部138は、データ行数と項目名と値とを対応付けて情報DB121に記憶すると、項目群判定部139に判定指示を出力する。これにより、情報処理装置100は、多様なフォーマットの表形式データを容易にデータベースに登録できる。

[0092] 次に、標準化処理について説明する。図25は、実施例の標準化処理の一例を示すフローチャートである。

[0093] 項目群判定部139は、記憶制御部138から判定指示が入力されると、情報DB121の1レコード目から順に項目名を読み込む。項目群判定部139は、読み込んだ項目名を整形する（ステップS11）。項目群判定部139は、ボキャブラリDB122を参照し、整形後の項目名を用いて、標準化ボキャブラリとのマッチングを実行する。項目群判定部139は、項目名が標準化ボキャブラリとマッチングしたか否かを判定する（ステップS12）。

[0094] 項目群判定部139は、項目名が標準化ボキャブラリとマッチングした場合には（ステップS12：肯定）、マッチングした標準化ボキャブラリを採用し（ステップS13）、ステップS18に進む。項目群判定部139は、採用した標準化ボキャブラリと、標準化ボキャブラリが属するグループとを情報DB121に記憶する。

[0095] 項目群判定部139は、項目名が標準化ボキャブラリとマッチングしない場合には（ステップS12：否定）、項目名を履歴DB123と照合する（ステップS14）。つまり、項目群判定部139は、項目名を過去の手動判定の対応付け履歴とのマッチングを実行する。項目群判定部139は、項目名が履歴DB123の項目名とマッチングしたか否かを判定する（ステップS15）。

[0096] 項目群判定部139は、項目名が履歴DB123の項目名とマッチングした場合には（ステップS15：肯定）、履歴DB123の標準化ボキャブラ

りを採用し（ステップS 1 6）、ステップS 1 8に進む。項目群判定部 1 3 9は、採用した標準化ボキャブラリと、標準化ボキャブラリが属するグループとを情報DB 1 2 1に記憶する。項目群判定部 1 3 9は、項目名が履歴DB 1 2 3の項目名とマッチングしない場合には（ステップS 1 5：否定）、項目名を手動判定ストックに追加する（ステップS 1 7）。

[0097] 項目群判定部 1 3 9は、全ての項目名のマッチングが完了したか否かを判定する（ステップS 1 8）。項目群判定部 1 3 9は、全ての項目名のマッチングが完了していない場合には（ステップS 1 8：否定）、ステップS 1 1に戻る。項目群判定部 1 3 9は、全ての項目名のマッチングが完了した場合には（ステップS 1 8：肯定）、提示指示を提示部 1 4 0に出力する。

[0098] 提示部 1 4 0は、項目群判定部 1 3 9から提示指示が入力されると、情報DB 1 2 1に標準化ボキャブラリ及びグループが記憶されている項目名は、記憶されている標準化ボキャブラリを対応付け対象として選択する。提示部 1 4 0は、選択した標準化ボキャブラリ及びグループと対応する項目名とを対応関係記憶制御部 1 4 1に出力する。

[0099] 提示部 1 4 0は、手動判定ストックに記憶された項目名に対して、標準化ボキャブラリ候補を提示する割当画面を図示しない端末装置に送信して表示させる（ステップS 1 9）。提示部 1 4 0は、図示しない端末装置から選択情報を受信する。提示部 1 4 0は、選択情報を受け付け、選択を受け付けた標準化ボキャブラリ及びグループと対応する項目名とを対応関係記憶制御部 1 4 1に出力する。

[0100] 対応関係記憶制御部 1 4 1には、提示部 1 4 0から選択された標準化ボキャブラリ及びグループと対応する項目名とが入力される。又は、対応関係記憶制御部 1 4 1には、提示部 1 4 0から選択を受け付けた標準化ボキャブラリ及びグループと対応する項目名とが入力される。対応関係記憶制御部 1 4 1は、選択された又は選択を受け付けた標準化ボキャブラリ及びグループを、項目名と対応付けて情報DB 1 2 1に記憶する（ステップS 2 0）。また、対応関係記憶制御部 1 4 1は、選択を受け付けた標準化ボキャブラリ及び

グループを項目名と対応付けて履歴DB123に記憶する。これにより、情報処理装置100は、項目名を標準化したボキャブラリに対応付けできる。また、情報処理装置100は、標準化ボキャブラリを用いることで、各種データを統合して利用することができる。さらに、情報処理装置100は、自動的に標準化ボキャブラリが採用されない項目名に対して、適切な標準化ボキャブラリを提示することができる。

[0101] このように、情報処理装置100は、表形式データから複数の項目名を抽出する。また、情報処理装置100は、複数の項目群が記憶されたボキャブラリDB122を参照して、抽出した複数の項目名のそれぞれと所定の類似関係を有する項目名がどの項目群に含まれるか否か判定する。また、情報処理装置100は、複数の項目名のうち、肯定的な判定結果が得られた項目名については、所定の類似関係を有する項目名を対応付け対象として選択する。また、情報処理装置100は、否定的な判定結果が得られた項目名については、他の項目名と所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する。その結果、項目名を標準化したボキャブラリに対応付けできる。

[0102] また、情報処理装置100は、否定的な判定結果が得られた項目名については、他の項目名と所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する際に、さらに、所定の項目群を他の対応付け候補として提示する。その結果、マッチングする可能性の高いグループの標準化ボキャブラリを提示できる。

[0103] また、情報処理装置100では、複数の項目群は、業種ごとに形成された項目群と、業種共通に形成された項目群とを含み、所定の項目群は、該業種共通に形成された項目群に対応する。その結果、マッチングする可能性の高いグループの標準化ボキャブラリを提示できる。

[0104] また、情報処理装置100では、複数の項目名は、表形式データから1つの表として検出された表の中から抽出されたものである。その結果、1つの表における項目名を標準化したボキャブラリに対応付けできる。

- [0105] また、情報処理装置１００では、複数の項目名は、表形式データから抽出された複数の表の中から抽出されたものである。また、情報処理装置１００は、否定的な判定結果が得られた項目名については、複数の表のうち同じ表から抽出された他の項目名と所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する。その結果、マッチングする可能性の高いグループの標準化ボキャブラリを提示できる。
- [0106] また、情報処理装置１００は、否定的な判定結果が得られた項目名と同じ表に含まれる項目名と所定の類似関係を有する第１の項目名が含まれると判定され、かつ、否定的な判定結果が得られた項目名と異なる表に含まれる第２の項目名と所定の類似関係を有する項目名が含まれると判定された場合に、否定的な判定結果が得られた項目名について、第１の項目名が含まれる項目群を第２の項目名が含まれる項目群よりも優先して対応付け候補として提示する。その結果、他の項目名がマッチングしたグループの標準化ボキャブラリを提示できる。
- [0107] また、情報処理装置１００は、否定的な判定結果が得られた項目名について提示した対応付け候補のうち、採用された候補と否定的な判定結果が得られた項目名との対応関係を履歴ＤＢ１２３に記憶する。また、情報処理装置１００は、表形式データ又は他の表形式データから抽出した項目名が対応関係によって特定の項目名に対応付けられている場合は、該特定の項目名を表形式データ又は他の表形式データから抽出した項目名の対応付け対象として選択する。その結果、履歴に含まれるグループの標準化ボキャブラリを提示できる。
- [0108] また、情報処理装置１００は、否定的な判定結果が得られた他の項目名の対応付け候補を提示する際に、特定の項目名が含まれる項目群を、他の項目群よりも優先して対応付け候補として提示する。その結果、履歴に含まれるグループの標準化ボキャブラリを優先して提示できる。
- [0109] また、情報処理装置１００は、対応付け対象として選択された項目名、又は、提示された項目群に含まれる項目名から選択を受け付けた項目名を、表

形式データの複数の項目名のそれぞれの項目名と対応付けて情報ＤＢ１２１に記憶する。その結果、項目名を標準化したボキャブラリに対応付けできる。

[0110] また、情報処理装置１００は、入力された表形式データの各行又は各列についてデータが入力されたセルの有無を判定する。また、情報処理装置１００は、データが入力されたセルが存在する複数の連続する行又は列の塊を１つの表に関連する部分として抽出する。また、情報処理装置１００は、行又は列の塊のうち項目行又は項目列を特定する。また、情報処理装置１００は、特定した項目行又は項目列の各セルの入力データを項目名として抽出する。その結果、表形式データから項目名を抽出できる。

[0111] なお、上記実施例では、表の本体部分の上部に表のタイトルが記載されている場合を一例として挙げたが、これに限定されない。例えば、表の本体部分の上部に数行に渡って見出しや注釈が記載されている場合であっても、上記実施例と同様に、表の本体部分を抽出することができる。

[0112] また、上記実施例では、情報ＤＢ１２１の形態として、表データを構成するセルごとに１レコードとしたが、これに限定されない。例えば、情報ＤＢ１２１は、元の表データを復元できれば、どのような形態のデータベースでもよい。

[0113] また、上記実施例では、表データを情報ＤＢ１２１に登録する際に、併せて標準化ボキャブラリ及びグループを決定したが、これに限定されない。例えば、標準化ボキャブラリ及びグループを決定する標準化処理を、情報ＤＢ１２１に記憶された表データを用いる際に行ってもよい。これにより、表データを使用するユーザが統一した基準で標準化ボキャブラリ及びグループを決定できる。さらに、例えば、他の自治体の各種データを登録する際に、他の自治体や業者が登録を支援することができる。

[0114] また、図示した各部の各構成要素は、必ずしも物理的に図示の如く構成されていることを要しない。すなわち、各部の分散・統合の具体的形態は図示のものに限られず、その全部又は一部を、各種の負荷や使用状況等に応じて



、任意の単位で機能的又は物理的に分散・統合して構成することができる。  
例えば、判定部 131 と抽出部 132 とを統合してもよい。また、図示した各処理は、上記の順番に限定されるものではなく、処理内容を矛盾させない範囲において、同時に実施してもよく、順序を入れ替えて実施してもよい。

[0115] さらに、各装置で行われる各種処理機能は、CPU（又はMPU、MCU（Micro Controller Unit）等のマイクロ・コンピュータ）上で、その全部又は任意の一部を実行するようにしてもよい。また、各種処理機能は、CPU（又はMPU、MCU等のマイクロ・コンピュータ）で解析実行されるプログラム上、又はワイヤードロジックによるハードウェア上で、その全部又は任意の一部を実行するようにしてもよいことは言うまでもない。

[0116] ところで、上記の実施例で説明した各種の処理は、予め用意されたプログラムをコンピュータで実行することで実現できる。そこで、以下では、上記の実施例と同様の機能を有するプログラムを実行するコンピュータの一例を説明する。図26は、項目名対応付け処理プログラムを実行するコンピュータの一例を示す図である。

[0117] 図26に示すように、コンピュータ200は、各種演算処理を実行するCPU201と、データ入力を受け付ける入力装置202と、モニタ203とを有する。また、コンピュータ200は、記憶媒体からプログラム等を読み取る媒体読取装置204と、各種装置と接続するためのインタフェース装置205と、他の情報処理装置等と有線又は無線により接続するための通信装置206とを有する。また、コンピュータ200は、各種情報を一時記憶するRAM207と、ハードディスク装置208とを有する。また、各装置201～208は、バス209に接続される。

[0118] ハードディスク装置208には、図1に示した判定部131、抽出部132、編集部133、カウント部134、生成部135、検出部136、特定部137及び記憶制御部138の各処理部と同様の機能を有する項目名対応付け処理プログラムが記憶される。また、ハードディスク装置208には、項目群判定部139、提示部140及び対応関係記憶制御部141の各処理

部と同様の機能を有する項目名対応付け処理プログラムが記憶される。また、ハードディスク装置２０８には、情報ＤＢ１２１、ボキャブラリＤＢ１２２、履歴ＤＢ１２３、及び、項目名対応付け処理プログラムを実現するための各種データが記憶される。入力装置２０２は、例えば、コンピュータ２００の管理者から操作情報、管理情報等の各種情報の入力を受け付ける。モニタ２０３は、例えば、コンピュータ２００の管理者に対して管理画面等の各種画面を表示する。インタフェース装置２０５は、例えば印刷装置等が接続される。通信装置２０６は、例えば、図１に示した通信部１１０と同様の機能を有し図示しないネットワークと接続され、図示しない端末装置と各種情報をやりとりする。

[0119] CPU２０１は、ハードディスク装置２０８に記憶された各プログラムを読み出して、RAM２０７に展開して実行することで、各種の処理を行う。また、これらのプログラムは、コンピュータ２００を図１に示した判定部１３１、抽出部１３２、編集部１３３、カウント部１３４、生成部１３５、検出部１３６、特定部１３７及び記憶制御部１３８として機能させることができる。また、これらのプログラムは、コンピュータ２００を図１に示した項目群判定部１３９、提示部１４０及び対応関係記憶制御部１４１として機能させることができる。

[0120] なお、上記の項目名対応付け処理プログラムは、必ずしもハードディスク装置２０８に記憶されている必要はない。例えば、コンピュータ２００が読み取り可能な記憶媒体に記憶されたプログラムを、コンピュータ２００が読み出して実行するようにしてもよい。コンピュータ２００が読み取り可能な記憶媒体は、例えば、ＣＤ－ＲＯＭやＤＶＤディスク、ＵＳＢ（Universal Serial Bus）メモリ等の可搬型記録媒体、フラッシュメモリ等の半導体メモリ、ハードディスクドライブ等が対応する。また、公衆回線、インターネット、ＬＡＮ等に接続された装置にこの項目名対応付け処理プログラムを記憶させておき、コンピュータ２００がこれらから項目名対応付け処理プログラムを読み出して実行するようにしてもよい。

## 符号の説明

|        |       |           |
|--------|-------|-----------|
| [0121] | 1 0 0 | 情報処理装置    |
|        | 1 1 0 | 通信部       |
|        | 1 1 1 | 表示部       |
|        | 1 1 2 | 操作部       |
|        | 1 2 0 | 記憶部       |
|        | 1 2 1 | 情報ＤＢ      |
|        | 1 2 2 | ボキャブラリＤＢ  |
|        | 1 2 3 | 履歴ＤＢ      |
|        | 1 3 0 | 制御部       |
|        | 1 3 1 | 判定部       |
|        | 1 3 2 | 抽出部       |
|        | 1 3 3 | 編集部       |
|        | 1 3 4 | カウント部     |
|        | 1 3 5 | 生成部       |
|        | 1 3 6 | 検出部       |
|        | 1 3 7 | 特定部       |
|        | 1 3 8 | 記憶制御部     |
|        | 1 3 9 | 項目群判定部    |
|        | 1 4 0 | 提示部       |
|        | 1 4 1 | 対応関係記憶制御部 |

## 請求の範囲

### [請求項1]

表形式データから複数の項目名を抽出し、

複数の項目群が記憶された記憶部を参照して、抽出した前記複数の項目名のそれぞれと所定の類似関係を有する項目名がどの項目群に含まれるか否か判定し、

前記複数の項目名のうち、肯定的な判定結果が得られた項目名については、前記所定の類似関係を有する項目名を対応付け対象として選択し、否定的な判定結果が得られた項目名については、他の項目名と前記所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する、

処理をコンピュータが実行することを特徴とする項目名対応付け処理方法。

### [請求項2]

前記提示する処理は、否定的な判定結果が得られた項目名については、他の項目名と前記所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する際に、さらに、所定の項目群を他の対応付け候補として提示する、

ことを特徴とする請求項1に記載の項目名対応付け処理方法。

### [請求項3]

前記複数の項目群は、業種ごとに形成された項目群と、業種共通に形成された項目群とを含み、前記所定の項目群は、該業種共通に形成された項目群に対応する、

ことを特徴とする請求項2に記載の項目名対応付け処理方法。

### [請求項4]

前記複数の項目名は、前記表形式データから1つの表として検出された表の中から抽出されたものである、

ことを特徴とする請求項1に記載の項目名対応付け処理方法。

### [請求項5]

前記複数の項目名は、前記表形式データから検出された複数の表の中から抽出されたものであり、

前記提示する処理は、否定的な判定結果が得られた項目名については、前記複数の表のうち同じ表から抽出された他の項目名と前記所定

の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する、

ことを特徴とする請求項 1 に記載の項目名対応付け処理方法。

[請求項6]

前記提示する処理は、前記否定的な判定結果が得られた項目名と同じ表に含まれる項目名と前記所定の類似関係を有する第 1 の項目名が含まれると判定され、かつ、前記否定的な判定結果が得られた項目名と異なる表に含まれる第 2 の項目名と前記所定の類似関係を有する項目名が含まれると判定された場合に、前記否定的な判定結果が得られた項目名について、前記第 1 の項目名が含まれる項目群を前記第 2 の項目名が含まれる項目群よりも優先して対応付け候補として提示する、

ことを特徴とする請求項 1 に記載の項目名対応付け処理方法。

[請求項7]

否定的な判定結果が得られた項目名について提示した対応付け候補のうち、採用された候補と前記否定的な判定結果が得られた項目名との対応関係を記憶部に記憶し、

前記提示する処理は、前記表形式データ又は他の表形式データから抽出した項目名が前記対応関係によって特定の項目名に対応付けられている場合は、該特定の項目名を前記表形式データ又は他の表形式データから抽出した項目名の対応付け対象として選択する、

ことを特徴とする請求項 1 に記載の項目名対応付け処理方法。

[請求項8]

前記提示する処理は、否定的な判定結果が得られた他の項目名の対応付け候補を提示する際に、前記特定の項目名が含まれる項目群を、他の項目群よりも優先して対応付け候補として提示する、

ことを特徴とする請求項 7 に記載の項目名対応付け処理方法。

[請求項9]

前記対応付け対象として選択された項目名、又は、前記提示された項目群に含まれる項目名から選択を受け付けた項目名を、前記表形式データの前記複数の項目名のそれぞれの項目名と対応付けて記憶部に記憶する、

ことを特徴とする請求項 1 に記載の項目名対応付け処理方法。

[請求項10]

前記項目名を抽出する処理は、

入力された表形式データの各行又は各列についてデータが入力されたセルの有無を判定し、

データが入力されたセルが存在する複数の連続する行又は列の塊を 1 つの表に関連する部分として抽出し、

前記行又は列の塊のうち項目行又は項目列を特定し、

特定した前記項目行又は前記項目列の各セルの入力データを前記項目名として抽出する、

ことを特徴とする請求項 1 に記載の項目名対応付け処理方法。

[請求項11]

表形式データから複数の項目名を抽出し、

複数の項目群が記憶された記憶部を参照して、抽出した前記複数の項目名のそれぞれと所定の類似関係を有する項目名がどの項目群に含まれるか否か判定し、

前記複数の項目名のうち、肯定的な判定結果が得られた項目名については、前記所定の類似関係を有する項目名を対応付け対象として選択し、否定的な判定結果が得られた項目名については、他の項目名と前記所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する、

処理をコンピュータに実行させることを特徴とする項目名対応付け処理プログラム。

[請求項12]

前記提示する処理は、否定的な判定結果が得られた項目名については、他の項目名と前記所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する際に、さらに、所定の項目群を他の対応付け候補として提示する、

ことを特徴とする請求項 1 1 に記載の項目名対応付け処理プログラム。

[請求項13]

前記複数の項目群は、業種ごとに形成された項目群と、業種共通に

形成された項目群とを含み、前記所定の項目群は、該業種共通に形成された項目群に対応する、

ことを特徴とする請求項 1 2 に記載の項目名対応付け処理プログラム。

[請求項14] 前記複数の項目名は、前記表形式データから 1 つの表として検出された表の中から抽出されたものである、

ことを特徴とする請求項 1 1 に記載の項目名対応付け処理プログラム。

[請求項15] 前記複数の項目名は、前記表形式データから検出された複数の表の中から抽出されたものであり、

前記提示する処理は、否定的な判定結果が得られた項目名については、前記複数の表のうち同じ表から抽出された他の項目名と前記所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する、

ことを特徴とする請求項 1 1 に記載の項目名対応付け処理プログラム。

[請求項16] 前記提示する処理は、前記否定的な判定結果が得られた項目名と同じ表に含まれる項目名と前記所定の類似関係を有する第 1 の項目名が含まれると判定され、かつ、前記否定的な判定結果が得られた項目名と異なる表に含まれる第 2 の項目名と前記所定の類似関係を有する項目名が含まれると判定された場合に、前記否定的な判定結果が得られた項目名について、前記第 1 の項目名が含まれる項目群を前記第 2 の項目名が含まれる項目群よりも優先して対応付け候補として提示する、

ことを特徴とする請求項 1 1 に記載の項目名対応付け処理プログラム。

[請求項17] 否定的な判定結果が得られた項目名について提示した対応付け候補のうち、採用された候補と前記否定的な判定結果が得られた項目名と

の対応関係を記憶部に記憶し、

前記提示する処理は、前記表形式データ又は他の表形式データから抽出した項目名が前記対応関係によって特定の項目名に対応付けられている場合は、該特定の項目名を前記表形式データ又は他の表形式データから抽出した項目名の対応付け対象として選択する、

ことを特徴とする請求項 11 に記載の項目名対応付け処理プログラム。

[請求項18] 前記提示する処理は、否定的な判定結果が得られた他の項目名の対応付け候補を提示する際に、前記特定の項目名が含まれる項目群を、他の項目群よりも優先して対応付け候補として提示する、

ことを特徴とする請求項 17 に記載の項目名対応付け処理プログラム。

[請求項19] 前記対応付け対象として選択された項目名、又は、前記提示された項目群に含まれる項目名から選択を受け付けた項目名を、前記表形式データの前記複数の項目名のそれぞれの項目名と対応付けて記憶部に記憶する、

ことを特徴とする請求項 11 に記載の項目名対応付け処理プログラム。

[請求項20] 前記項目名を抽出する処理は、

入力された表形式データの各行又は各列についてデータが入力されたセルの有無を判定し、

データが入力されたセルが存在する複数の連続する行又は列の塊を 1 つの表に関連する部分として抽出し、

前記行又は列の塊のうち項目行又は項目列を特定し、

特定した前記項目行又は前記項目列の各セルの入力データを前記項目名として抽出する、

ことを特徴とする請求項 11 に記載の項目名対応付け処理プログラム。



- [請求項21] 表形式データから複数の項目名を抽出する項目名抽出部と、  
複数の項目群が記憶された記憶部を参照して、抽出した前記複数の項目名のそれぞれと所定の類似関係を有する項目名がどの項目群に含まれるか否か判定する項目群判定部と、  
前記複数の項目名のうち、肯定的な判定結果が得られた項目名については、前記所定の類似関係を有する項目名を対応付け対象として選択し、否定的な判定結果が得られた項目名については、他の項目名と前記所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する提示部と、  
を有することを特徴とする情報処理装置。
- [請求項22] 前記提示部は、否定的な判定結果が得られた項目名については、他の項目名と前記所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する際に、さらに、所定の項目群を他の対応付け候補として提示する、  
ことを特徴とする請求項21に記載の情報処理装置。
- [請求項23] 前記複数の項目群は、業種ごとに形成された項目群と、業種共通に形成された項目群とを含み、前記所定の項目群は、該業種共通に形成された項目群に対応する、  
ことを特徴とする請求項22に記載の情報処理装置。
- [請求項24] 前記複数の項目名は、前記表形式データから1つの表として検出された表の中から抽出されたものである、  
ことを特徴とする請求項21に記載の情報処理装置。
- [請求項25] 前記複数の項目名は、前記表形式データから検出された複数の表の中から抽出されたものであり、  
前記提示部は、否定的な判定結果が得られた項目名については、前記複数の表のうち同じ表から抽出された他の項目名と前記所定の類似関係を有する項目名が含まれると判定された項目群を対応付け候補として提示する、

ことを特徴とする請求項 2 1 に記載の情報処理装置。

[請求項26]

前記提示部は、前記否定的な判定結果が得られた項目名と同じ表に含まれる項目名と前記所定の類似関係を有する第 1 の項目名が含まれると判定され、かつ、前記否定的な判定結果が得られた項目名と異なる表に含まれる第 2 の項目名と前記所定の類似関係を有する項目名が含まれると判定された場合に、前記否定的な判定結果が得られた項目名について、前記第 1 の項目名が含まれる項目群を前記第 2 の項目名が含まれる項目群よりも優先して対応付け候補として提示する、

ことを特徴とする請求項 2 1 に記載の情報処理装置。

[請求項27]

さらに、否定的な判定結果が得られた項目名について提示した対応付け候補のうち、採用された候補と前記否定的な判定結果が得られた項目名との対応関係を記憶部に記憶する対応関係記憶制御部を有し、

前記提示部は、前記表形式データ又は他の表形式データから抽出した項目名が前記対応関係によって特定の項目名に対応付けられている場合は、該特定の項目名を前記表形式データ又は他の表形式データから抽出した項目名の対応付け対象として選択する、

ことを特徴とする請求項 2 1 に記載の情報処理装置。

[請求項28]

前記提示部は、否定的な判定結果が得られた他の項目名の対応付け候補を提示する際に、前記特定の項目名が含まれる項目群を、他の項目群よりも優先して対応付け候補として提示する、

ことを特徴とする請求項 2 7 に記載の情報処理装置。

[請求項29]

さらに、前記対応付け対象として選択された項目名、又は、前記提示された項目群に含まれる項目名から選択を受け付けた項目名を、前記表形式データの前記複数の項目名のそれぞれの項目名と対応付けて記憶部に記憶する対応関係記憶制御部、

を有することを特徴とする請求項 2 1 に記載の情報処理装置。

[請求項30]

前記項目名抽出部は、

入力された表形式データの各行又は各列についてデータが入力され

たセルの有無を判定する判定部と、

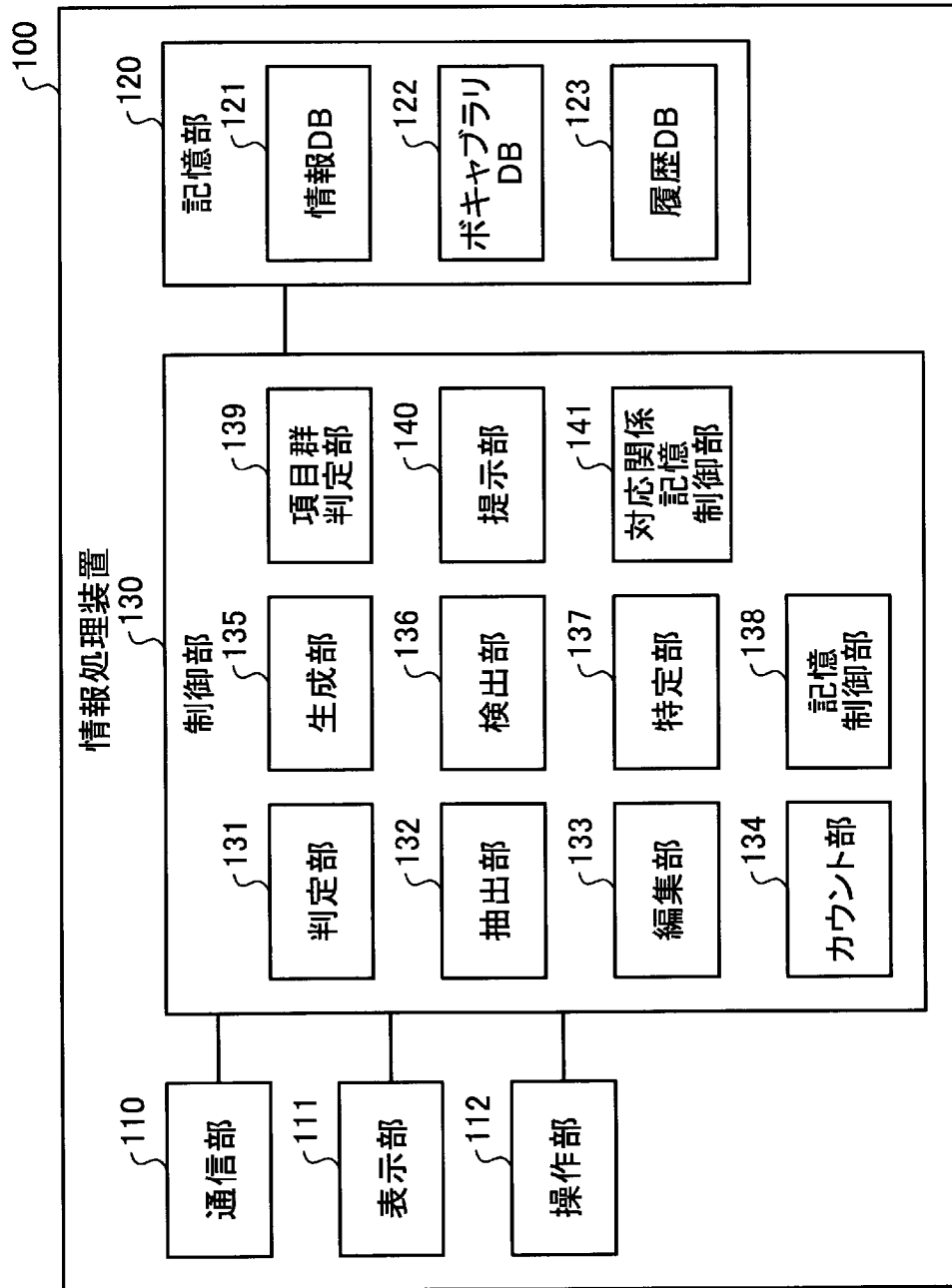
データが入力されたセルが存在する複数の連続する行又は列の塊を  
1つの表に関連する部分として抽出する抽出部と、

前記行又は列の塊のうち項目行又は項目列を特定する特定部と、を  
有し、

特定した前記項目行又は前記項目列の各セルの入力データを前記項  
目名として抽出する、

ことを特徴とする請求項21に記載の情報処理装置。

[図1]



[図2]

|    | A    | B        | C   | D   | E   | F | G | H        | I |
|----|------|----------|-----|-----|-----|---|---|----------|---|
| 1  | タイトル |          |     |     |     |   |   | 2015/4/6 |   |
| 2  |      |          |     |     |     |   |   |          |   |
| 3  |      | シンプルな表   |     |     |     |   |   |          |   |
| 4  |      | 項目1      | 項目2 | 項目3 | 項目4 |   |   |          |   |
| 5  |      | a        |     |     |     |   |   |          |   |
| 6  |      | b        |     |     |     |   |   |          |   |
| 7  |      | c        |     |     |     |   |   |          |   |
| 8  |      | d        |     |     |     |   |   |          |   |
| 9  |      | e        |     |     |     |   |   |          |   |
| 10 |      | f        |     |     |     |   |   |          |   |
| 11 |      | g        |     |     |     |   |   |          |   |
| 12 |      |          |     |     |     |   |   |          |   |
| 13 |      |          |     |     |     |   |   |          |   |
| 14 |      | セル連結がある表 |     |     |     |   |   |          |   |
| 15 |      | 項目1      |     | 項目3 | 項目4 |   |   |          |   |
| 16 |      | 1b       | 1b  |     |     |   |   |          |   |
| 17 | *    | a        | 1   | 2   | 3   |   |   |          |   |
| 18 | *    | b        | 1   | 2   | 3   |   |   |          |   |
| 19 |      | c        | 1   | 2   | 3   |   |   |          |   |
| 20 |      | d        | 1   | 2   | 3   |   |   |          |   |
| 21 |      | e        | 1   | 2   | 3   |   |   |          |   |
| 22 |      |          | 1   | 2   | 3   |   |   |          |   |
| 23 |      |          | 1   | 2   | 3   |   |   |          |   |
| 24 |      | ※補足事項。   |     |     |     |   |   |          |   |
| 25 |      |          |     |     |     |   |   |          |   |
| 26 |      |          |     |     |     |   |   |          |   |
| 27 |      |          |     |     |     |   |   |          |   |
| 28 |      |          |     |     |     |   |   |          |   |

11

12a

12b

[図3]

121  
↓

| 行 | 項目 | 値  | 標準化ボキャブラリ | グループ |
|---|----|----|-----------|------|
| 1 | x1 | y1 | B01       | G02  |
| 1 | x2 | 11 | B02       | G02  |
| 1 | x3 | 12 | B03       | G02  |
| 1 | x4 | 13 | B04       | G02  |
| 1 | x5 | 14 | B05       | G02  |
| 2 | x1 | y2 | B01       | G02  |
| 2 | x2 | 21 | B02       | G02  |
| 2 | x3 | 22 | B03       | G02  |
| 2 | x4 | 23 | B04       | G02  |
| 2 | x5 | 24 | B05       | G02  |
| 3 | x1 | y3 | B01       | G02  |
| 3 | x2 | 25 | B02       | G02  |
| 3 | x3 | 26 | B03       | G02  |
| 3 | x4 | 27 | B04       | G02  |
| 3 | x5 | 28 | B05       | G02  |

[図4A]

| 共通ボキャブラリグループ |        |
|--------------|--------|
| タイトル         | ← 122a |
| 説明文          |        |
| カテゴリ         |        |
| 住所           |        |
| 郵便番号         |        |
| ...          |        |

[図4B]

| 医薬品ボキャブラリグループ  |        |
|----------------|--------|
| 薬の名称           | ← 122b |
| 個別医薬品コード       |        |
| JANコード         |        |
| 医薬品のタイプ        |        |
| 薬の説明文          |        |
| 薬に関する補足情報・コメント |        |
| ...            |        |

[図4C]

| 取引ボキャブラリグループ |        |
|--------------|--------|
| 取引の債権者       | ← 122c |
| 取引の債務者       |        |
| 金額単位         |        |
| 領収書の宛先       |        |
| 店舗名          |        |
| 店舗住所         |        |
| 店舗TEL        |        |
| 精算済みフラグ      |        |
| ...          |        |

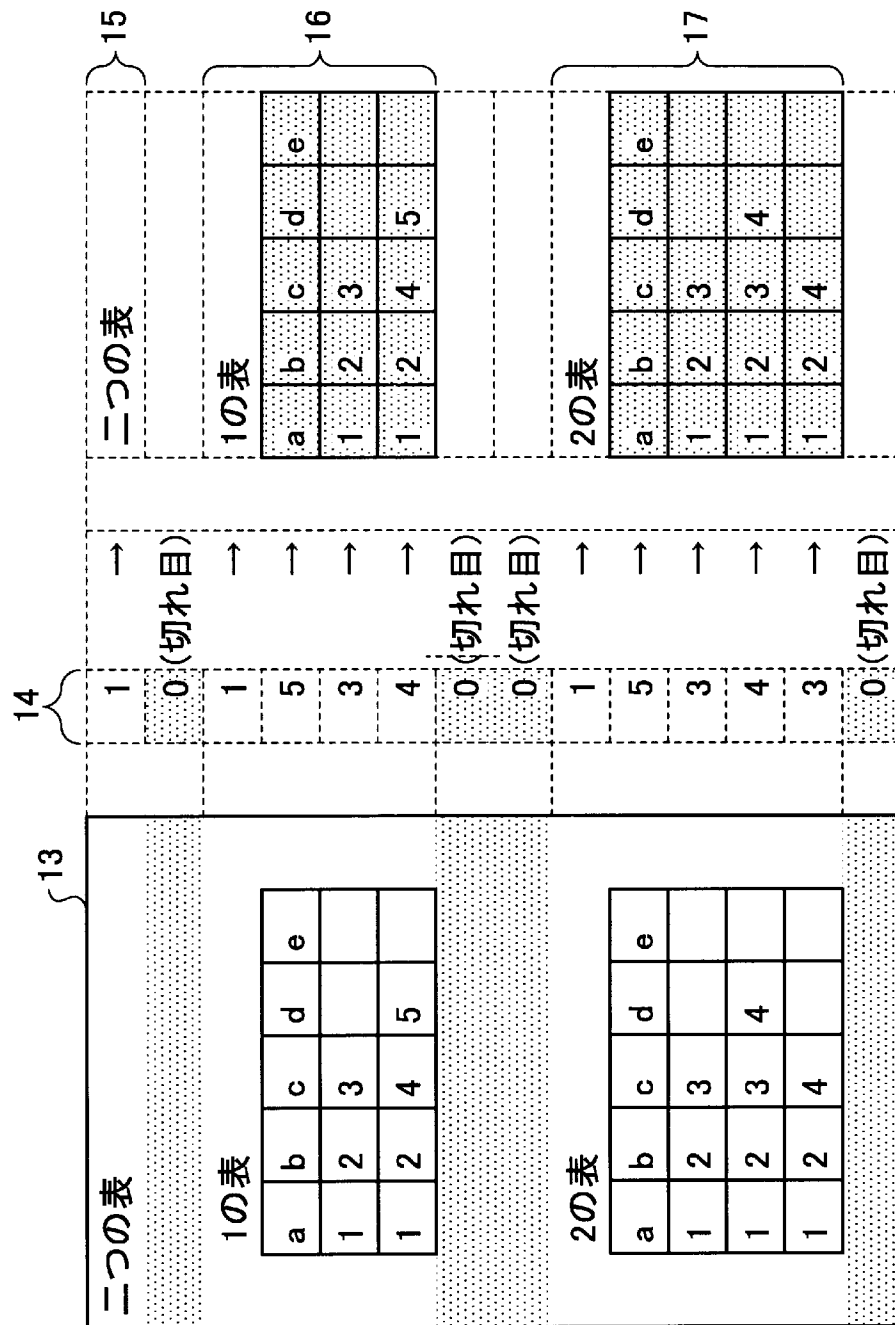


[図5]

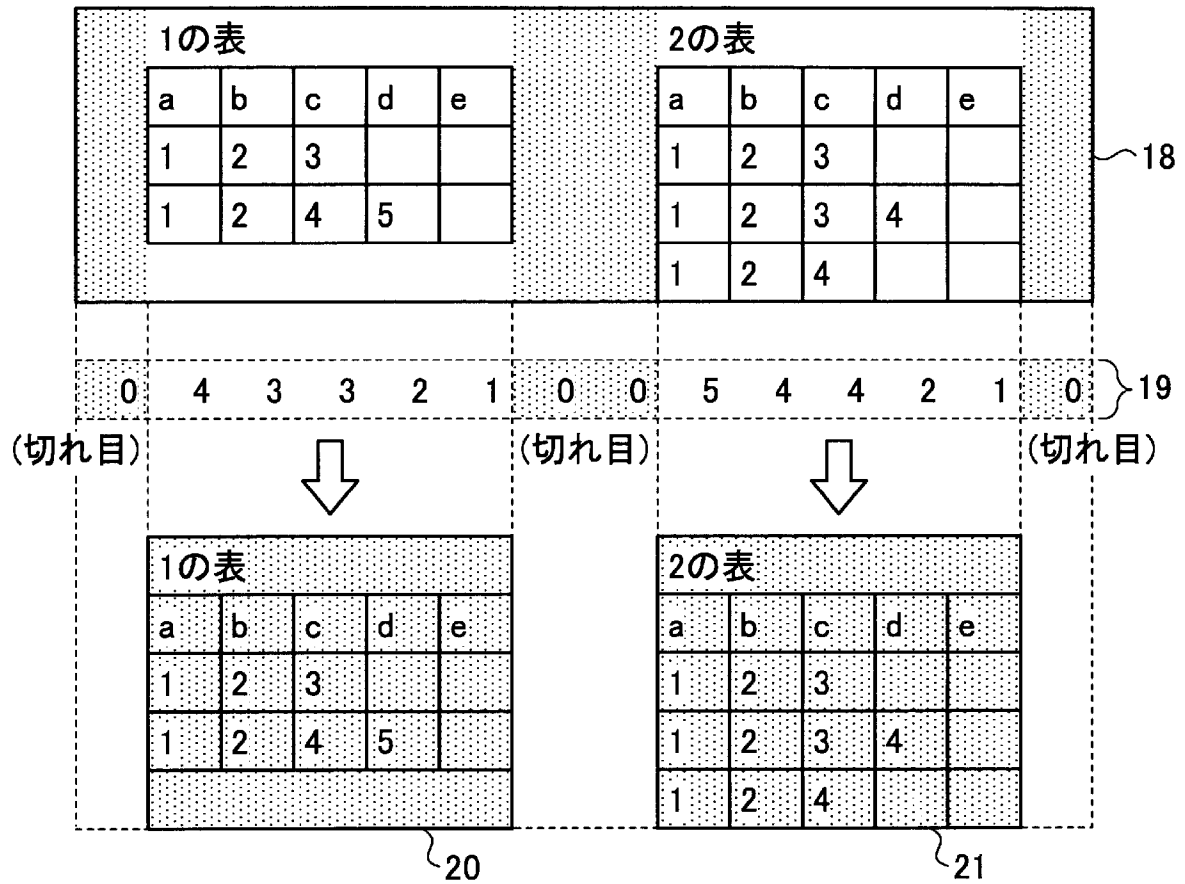
123  
↙

| 項目名 | 標準化ボキャブラリ | グループ |
|-----|-----------|------|
| TEL | 電話番号      | 共通   |
| バス  | 公共交通機関    | 共通   |
| 行き方 | 交通アクセス    | 共通   |
| ... | ...       | ...  |

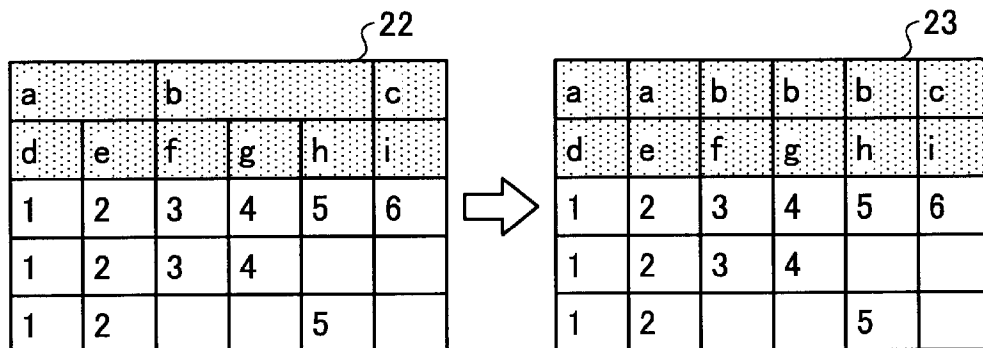
[図6]



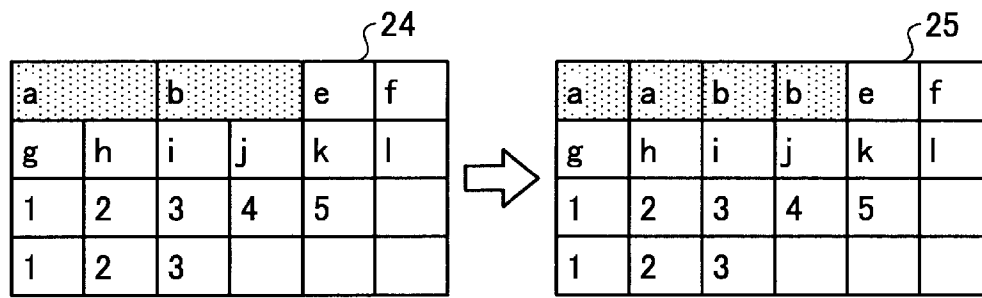
[図7]



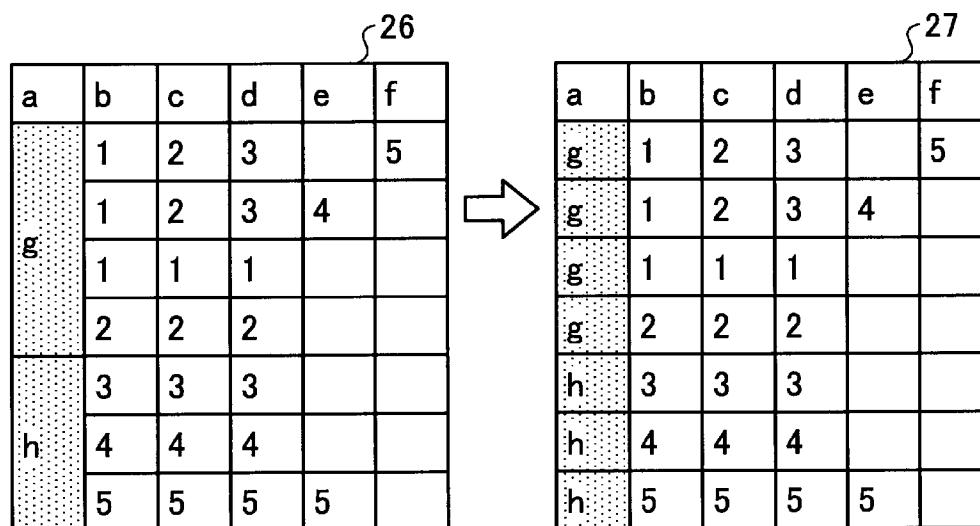
[図8]



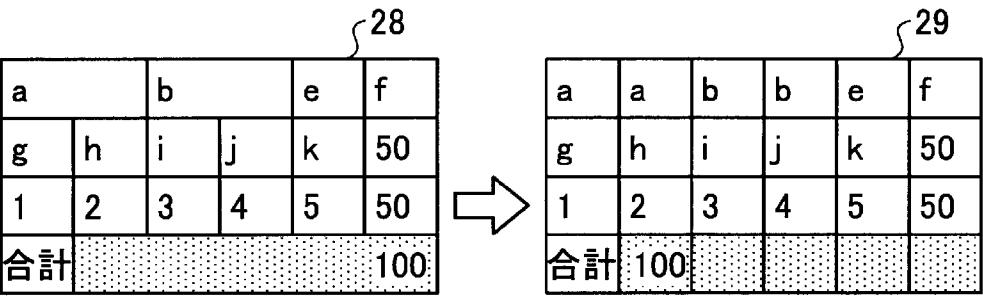
[図9]



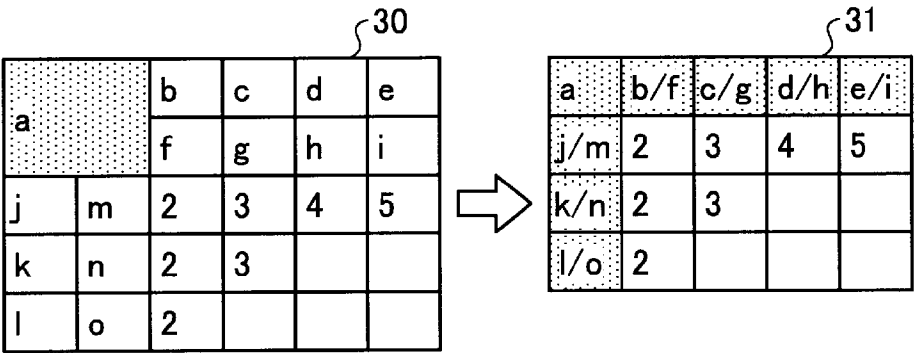
[図10]



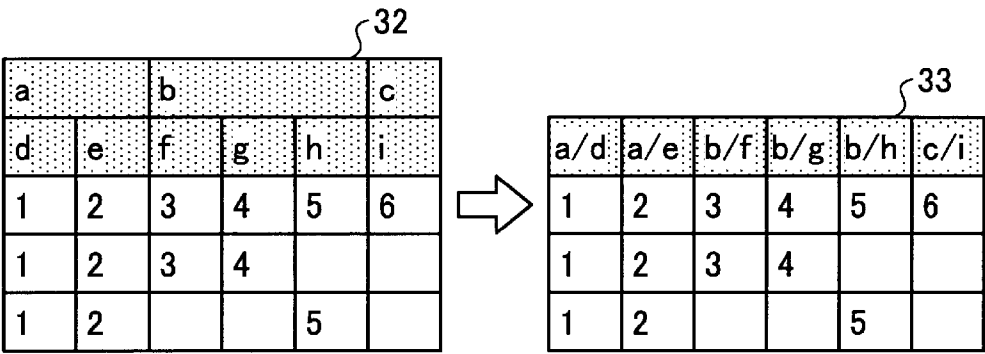
[図11]



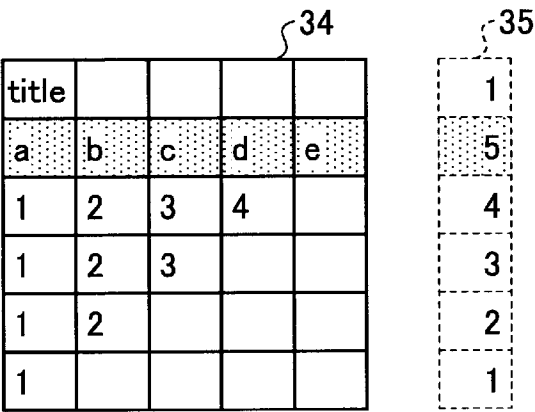
[図12]



[図13]



[図14]



[図15]

37

|       |   |   |   |   |
|-------|---|---|---|---|
| title |   |   |   |   |
| a     | b | c | d | e |
| 1     | 2 | 3 | 4 |   |
| 1     | 2 | 3 |   |   |
| f     | g | h | i | j |
| 1     | 2 |   |   |   |

38

|   |
|---|
| 1 |
| 5 |
| 4 |
| 3 |
| 5 |
| 2 |

[図16]

41

|       |   |   |   |   |
|-------|---|---|---|---|
| title |   |   |   |   |
| a     | b | c | d | e |
| f     | g | 1 | 2 | 3 |
| 1     | 2 | 3 |   |   |
| f     | g | h | i | j |
| 1     | 2 |   |   |   |

42 43

|   |      |
|---|------|
| 5 | 100% |
| 5 | 40%  |
| ⋮ |      |
| ⋮ |      |
| ⋮ |      |

[図17]

46

|       |   |   |   |   |
|-------|---|---|---|---|
| title |   |   |   |   |
| a     | b | c | d | e |
| d     | e | 1 | g | 2 |
| 1     | 2 | 3 |   |   |
| f     | g | h | i | j |
| 1     | 2 |   |   |   |

47 48

|   |      |
|---|------|
| 5 | 100% |
| 5 | 60%  |
| ⋮ |      |
| ⋮ |      |
| ⋮ |      |

[図18]

51

|   |   |   |   |   |
|---|---|---|---|---|
| a | b | c | d | e |
| 1 | 2 | 3 | 4 |   |
| 1 | 2 | 3 |   |   |
| f | g | h | i | j |
| 1 | 2 |   |   |   |

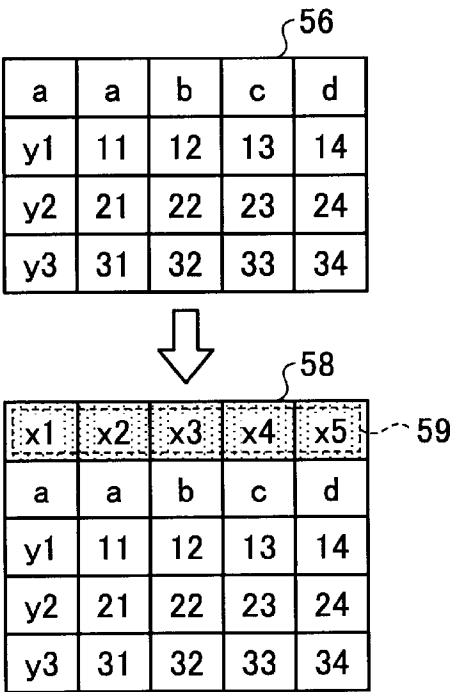


[図19]

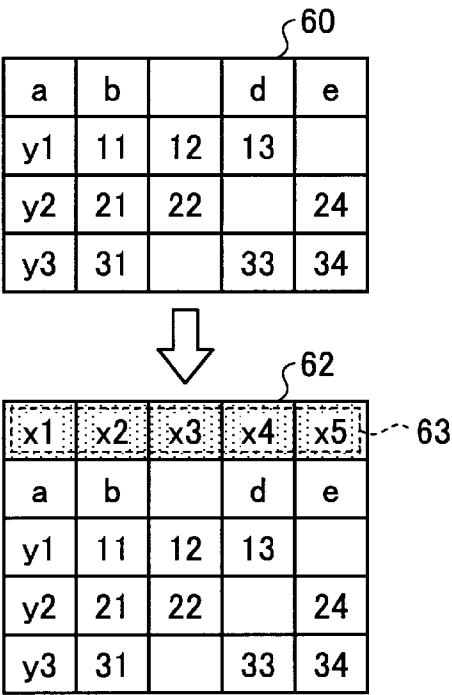
53

|   |   |   |   |   |
|---|---|---|---|---|
|   | a | c | d | e |
| 1 | 2 | 3 | 4 |   |
| 1 | 2 | 3 |   |   |
| f | g | h | i | j |
| 1 | 2 |   |   |   |

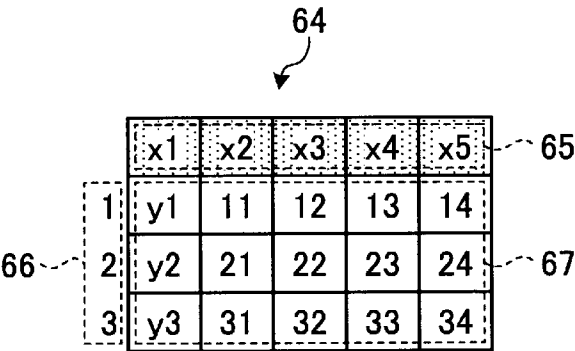
[図20]



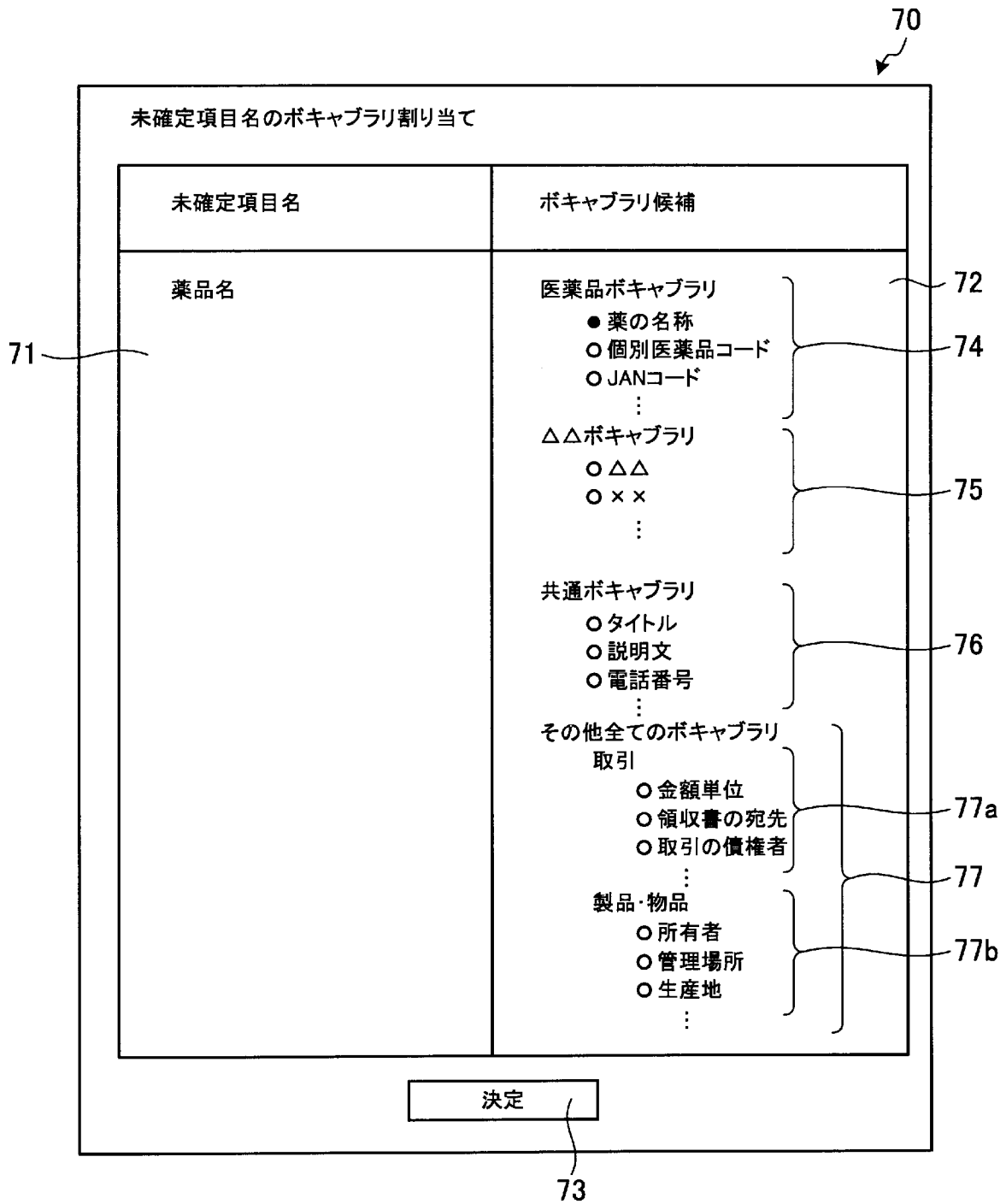
[図21]



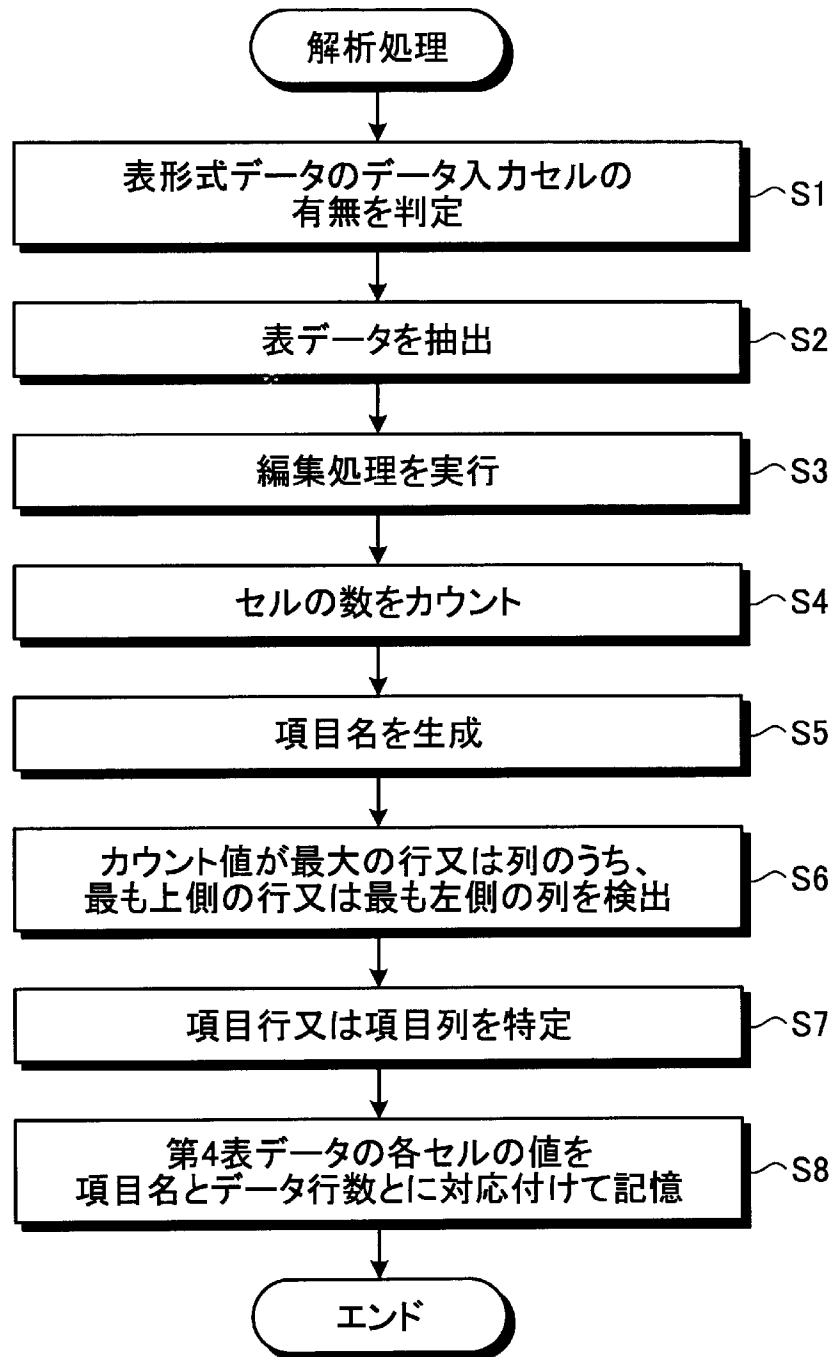
[図22]



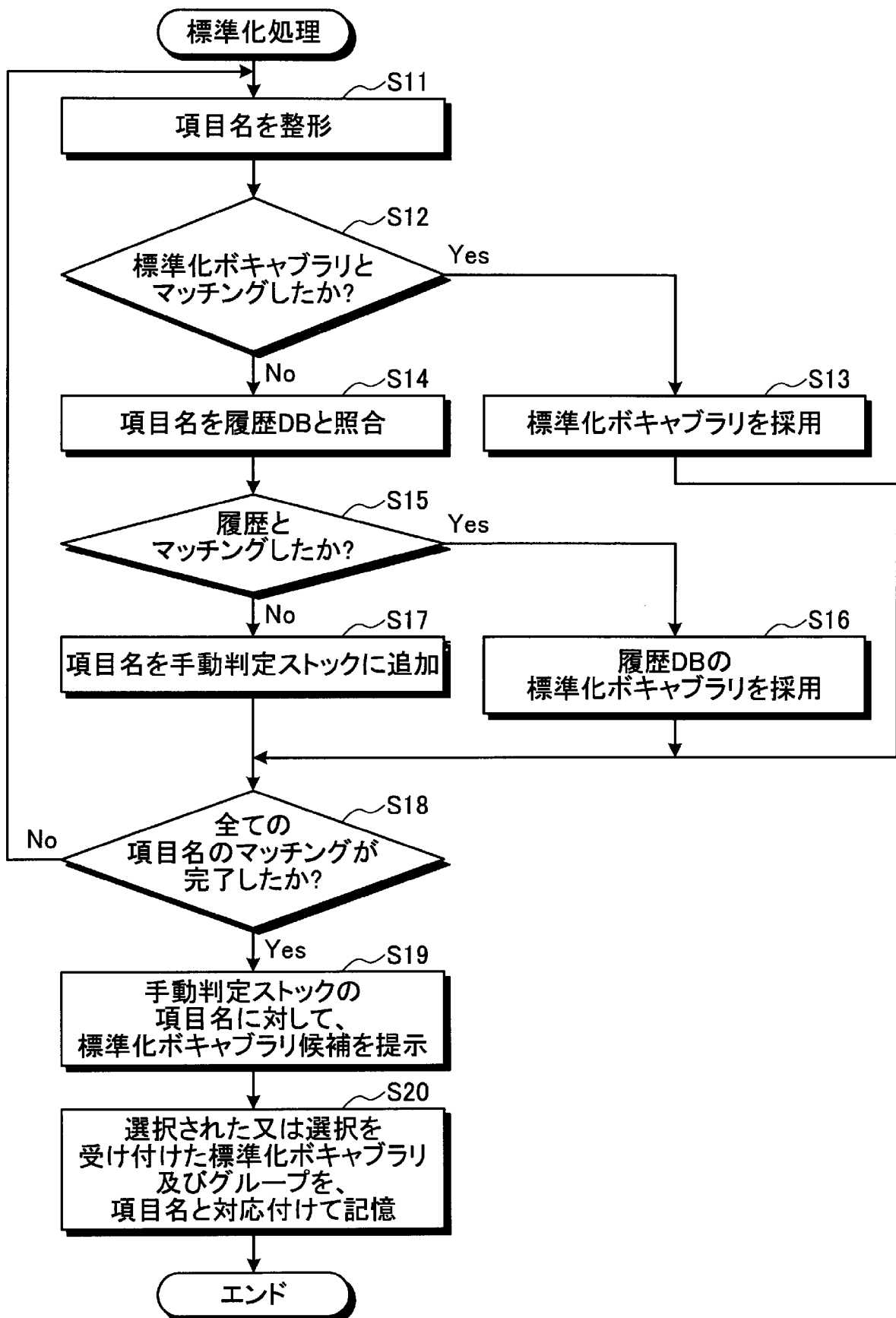
[図23]



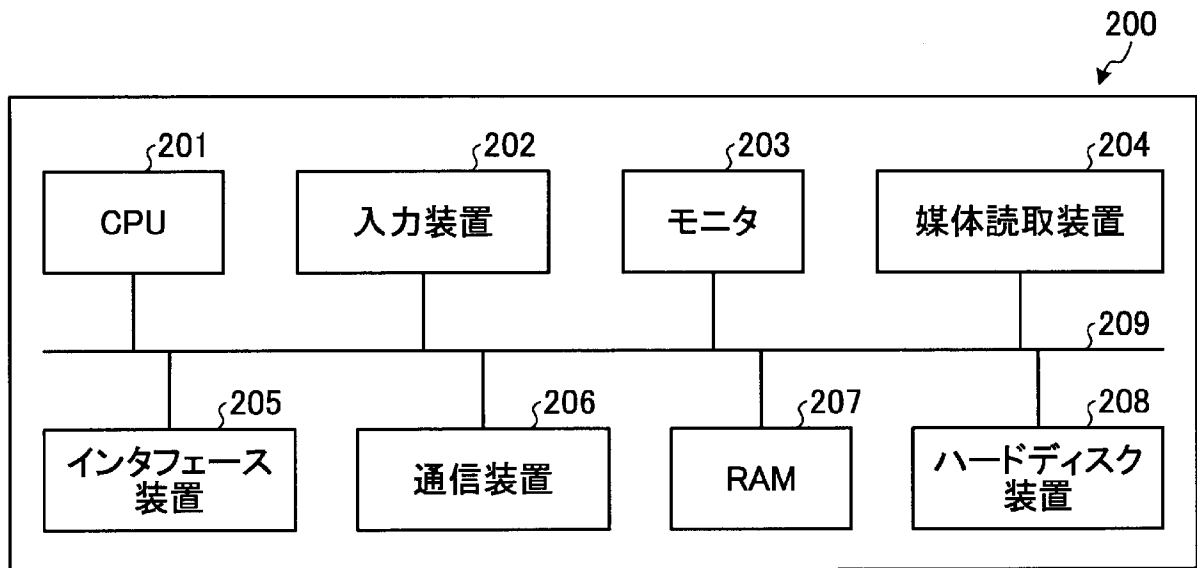
[図24]



[図25]



[図26]



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2016/053389

## A. CLASSIFICATION OF SUBJECT MATTER

G06Q50/00(2012.01)i, G06F17/24(2006.01)i, G06Q10/10(2012.01)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)

G06Q50/00, G06F17/24, G06Q10/10

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 | 1996-2016 |
| Kokai Jitsuyo Shinan Koho | 1971-2016 | Toroku Jitsuyo Shinan Koho | 1994-2016 |

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. |
|-----------|------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | JP 2000-305824 A (Justsystem Corp.),<br>02 November 2000 (02.11.2000),<br>paragraphs [0003] to [0136]<br>(Family: none)      | 1-9, 11-19,<br>21-29  |
| Y         |                                                                                                                              | 10, 20, 30            |
| Y         | JP 2005-242587 A (Fuji Xerox Co., Ltd.),<br>08 September 2005 (08.09.2005),<br>paragraphs [0009] to [0083]<br>(Family: none) | 10, 20, 30            |

☐ 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  
26 February 2016 (26.02.16)

Date of mailing of the international search report  
08 March 2016 (08.03.16)

Name and mailing address of the ISA/  
Japan Patent Office  
3-4-3, Kasumigaseki, Chiyoda-ku,  
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

## A. 発明の属する分野の分類（国際特許分類（IPC））

Int.Cl. G06Q50/00(2012.01)i, G06F17/24(2006.01)i, G06Q10/10(2012.01)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. G06Q50/00, G06F17/24, G06Q10/10

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 | 1971-2016年 |
| 日本国実用新案登録公報 | 1996-2016年 |
| 日本国登録実用新案公報 | 1994-2016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<br>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<br>請求項の番号       |
|-----------------|-------------------------------------------------------------------|----------------------|
| X               | JP 2000-305824 A（株式会社ジャストシステム）2000.11.02,<br>段落0003-0136（ファミリーなし） | 1-9, 11-19,<br>21-29 |
| Y               |                                                                   | 10, 20, 30           |
| Y               | JP 2005-242587 A（富士ゼロックス株式会社）2005.09.08,<br>段落0009-0083（ファミリーなし）  | 10, 20, 30           |

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

26.02.2016

国際調査報告の発送日

08.03.2016

国際調査機関の名称及びあて先

日本国特許庁（ISA/J P）

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）

塩田 徳彦

5 L

4533

電話番号 03-3581-1101 内線 3562