

(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
2020年9月3日(03.09.2020)



(10) 国際公開番号

WO 2020/175163 A1

- (51) 国際特許分類:
G06K 9/20 (2006.01)
- (21) 国際出願番号: PCT/JP2020/005631
- (22) 国際出願日: 2020年2月13日(13.02.2020)
- (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 2019-034980 2019年2月27日(27.02.2019) JP
- (71) 出願人: 日本電信電話株式会社 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目5番1号 Tokyo (JP).
- (72) 発明者: 高木 郁子(TAKAGI, Ikuko); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 知的財産センタ内 Tokyo (JP). 小笠原 志朗(OGASAWARA, Shiro); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 知的財

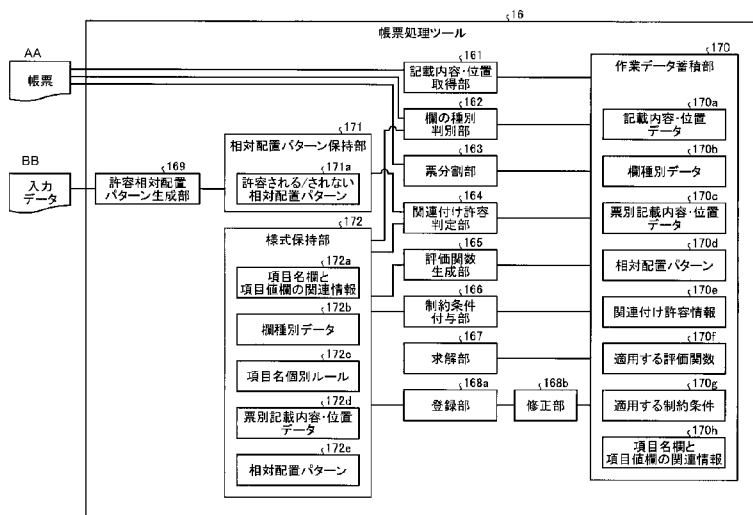
産センタ内 Tokyo (JP). 辻 幸嗣(TSUJI, Koji); 〒1808585 東京都武蔵野市緑町3丁目9-1 NTT 知的財産センタ内 Tokyo (JP).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, KE, KG, KH, KN, KP, KR, KW, 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,

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

(54) 発明の名称: 情報処理装置、関連付け方法および関連付けプログラム



- 16 Business form processing tool
- 161 Written content/position acquisition unit
- 162 Column type determination unit
- 163 Form dividing unit
- 164 Association permission determination unit
- 165 Evaluation function generation unit
- 166 Constraint condition addition unit
- 167 Solution-finding unit
- 168a Registration unit
- 168b Correction unit
- 169 Permissible relative arrangement pattern generation unit
- 170 Work data accumulation unit
- 170a Written content/position data
- 170b Column type data
- 170c 172dBy-form written content/position data
- 170d, 172e Relative arrangement pattern
- 170e Association permission information
- 170f Applied evaluation function
- 170g Applied constraint condition
- 170h, 172a Item name and item value columns association information
- 171 Relative arrangement pattern holding unit
- 171a Permissible/not-permissible relative arrangement pattern
- 172c Style holding unit
- 172c Item name individual rule
- AA Business form
- BB Input data

(57) Abstract: An information processing device (10) registers a set of item names associated with the item name column which is associated with the item value column of a prescribed business form, and specifies, on the basis of the registered set of item names, the association of the item name and item value columns included in the business form to be processed.

(57) 要約: 情報処理装置 (10) は、所定の帳票における項目値欄と関連付けられた項目名欄に対応付けられた項目名のセットを登録し、登録した項目名のセットに基づいて、処理対象の帳票に含まれる項目名欄と項目値欄との関連付けを特定する。

WO 2020/175163 A1

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, 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).

添付公開書類:

- 一 国際調査報告 (条約第21条(3))

明 細 書

発明の名称：

情報処理装置、関連付け方法および関連付けプログラム

技術分野

[0001] 本発明は、情報処理装置、関連付け方法および関連付けプログラムに関する。

背景技術

[0002] 業務において、電子ファイルまたは紙で作成された帳票が様々なシーンで用いられている。ここで帳票とは、電子ファイルまたは紙で作成されたドキュメントやシステム画面のなかでも、複数の項目について、それぞれ、その項目の名前（項目名）と値（項目値）が記載され、かつその関係性が2次元平面上の配置によって表現されているものをいう。

[0003] このような帳票は、記載内容の多くが同じであっても帳票の様式を定める組織によって様式が少しずつ異なることが多い。また、業務運用の過程で変更が加えられることも多い。例えば、従来技術として、帳票の様式を定めた組織の違いや、運用中の改定等により、レイアウトが異なっても、帳票上に記載された自動的に処理したい項目値の位置を正しく特定し、データを抽出できるようにするために、欄の種別（欄内の記載内容が項目名なのか、項目値なのか）を判別するための情報を予め与えることで、欄が項目名であるかどうかを判別し、その上で、項目名欄と項目値欄の対応関係を認識することで、指定された自動処理対象項目の項目名に対応する項目値欄を、帳票の欄の中から特定できるようにする方法がある。

先行技術文献

特許文献

[0004] 特許文献1：特開2018-10489号公報

発明の概要

発明が解決しようとする課題

[0005] しかしながら、従来の手法では、関連付けた結果が欄の配置変更に対応できない場合があるという課題があった。例えば、従来技術は、項目名が縦と横の両方に配置されている場合に縦と横が入れ替わった場合や、同一の項目値に関連付けられる複数の項目名間の親子関係の順序性が変わった場合に、項目値欄を特定できないという課題あった。

課題を解決するための手段

[0006] 上述した課題を解決し、目的を達成するために、本発明の情報処理装置は、所定の帳票における、所望の項目値欄と関連付けられた項目名欄に対応付けられた項目名のセットを登録する登録部と、前記登録部によって登録された項目名のセットに基づいて、処理対象の帳票に含まれる項目名欄と項目値欄を関連付け、所望の項目値欄を特定する求解部とを有することを特徴とする。

発明の効果

[0007] 本発明によれば、帳票の欄の配置に変更があった場合であっても、帳票に含まれる項目名欄と項目値欄とを関連付け、所望の項目値欄を特定することができるという効果を奏する。

図面の簡単な説明

[0008] [図1]図1は、第1の実施形態に係る情報処理装置の構成例を示すブロック図である。

[図2]図2は、第1の実施形態に係る情報処理装置における帳票処理ツールの論理構成を示すブロック図である。

[図3]図3は、相対配置パターンについて説明する図である。

[図4]図4は、記載内容・位置データの一例を示す図である。

[図5]図5は、欄種別データの一例を示す図である。

[図6]図6は、票を分割する処理を説明する図である。

[図7]図7は、相対配置パターンのデータ例を示す図である。

[図8]図8は、許容される／されない相対配置パターンの一例を示す図である。

[図9]図9は、関連付け許容情報の一例を示す図である。

[図10]図10は、項目値欄基準での相対配置の判定結果と、項目名欄基準での相対配置の判定結果の両方を総合的に判断する処理例を説明する図である。

[図11]図11は、項目値欄基準での相対配置の判定結果と、項目名欄基準での相対配置の判定結果の両方を総合的に判断する処理例を説明する図である。

[図12]図12は、相対配置パターンを生成する処理例を説明する図である。

[図13]図13は、相対配置パターンを生成する処理例を説明する図である。

[図14]図14は、制約条件と評価関数を数式で表現する際に使用する記号について説明するための図である。

[図15]図15は、様式登録時における制約条件の一例を示す図である。

[図16]図16は、様式登録時における評価関数の一例について説明する図である。

[図17]図17は、登録済み様式利用時における制約条件について説明する図である。

[図18]図18は、登録済み様式利用時における評価関数の一例を示す図である。

[図19]図19は、第1の実施形態に係る情報処理装置による求解例を示す図である。

[図20]図20は、項目名欄と項目値欄の関連情報を示す図である。

[図21A]図21Aは、第1の実施形態に係る情報処理装置における全体的な処理の流れの一例を示すフローチャートである。

[図21B]図21Bは、第1の実施形態に係る情報処理装置における全体的な処理の流れの一例を示すフローチャートである。

[図22]図22は、第1の実施形態に係る情報処理装置の許容相対配置パターン生成部における処理の流れの一例を示すフローチャートである。

[図23]図23は、第1の実施形態に係る情報処理装置の記載内容・位置取得

部における処理の流れの一例を示すフローチャートである。

[図24]図24は、第1の実施形態に係る情報処理装置の記載内容・位置取得部における処理の流れの一例を示すフローチャートである。

[図25]図25は、第1の実施形態に係る情報処理装置の記載内容・位置取得部における処理の流れの一例を示すフローチャートである。

[図26]図26は、特定プログラムを実行するコンピュータを示す図である。

発明を実施するための形態

[0009] 以下に、本願に係る情報処理装置、関連付け方法および関連付けプログラムの実施の形態を図面に基づいて詳細に説明する。なお、この実施の形態により本願に係る情報処理装置、関連付け方法および関連付けプログラムが限定されるものではない。

[0010] [第1の実施形態]

以下の実施の形態では、第1の実施形態に係る情報処理装置10の構成、帳票処理ツール16の論理構成、情報処理装置10の処理の流れを順に説明し、最後に第1の実施形態による効果を説明する。

[0011] [情報処理装置の構成]

まず、図1を用いて、本実施形態の情報処理装置10を含むシステムの構成例を説明する。図1は、第1の実施形態に係る情報処理装置の構成例を示すブロック図である。なお、図1に示す構成は一例にすぎず、具体的な構成は特に限定されない。

[0012] 情報処理装置10は、PC (Personal Computer) やスマートフォン等の装置であって、例えば、帳票の記載内容を機械的に処理する。ここで、帳票とは、例えば、アプリケーションで作成された電子データ、PDFファイル、システム画面の構造データ、画像データ（例えば、電子ファイルやシステム画面を画像データに変換したもの、紙をスキャンして電子化したもの）が対象である。

[0013] また、情報処理装置10は、帳票の記載内容を機械的に処理するのに先立ち、許容される相対配置パターンおよび許容されない相対配置パターン（以

下、「許容される／されない相対配置パターン」と記載する)を登録する処理を行う。情報処理装置10は、帳票の様式登録の際、および登録済みの様式情報を利用する際は、登録した許容される／されない相対配置パターンを用いて、帳票に含まれる項目名欄と項目値欄との関連付けの可能性を判定する。なお、帳票の欄は矩形でなくてもよく、単純多角形であればよいものとする。また、帳票の欄は罫線で囲われることもあれば、囲われていないこともある。このため、罫線で囲われていない場合には、オブジェクト取得等で取得した矩形を欄として扱うようにしてもよい。

[0014] ここで、相対配置パターンについて、図3を用いて説明する。例えば、帳票における欄iの欄jに対する相対配置パターンとは、図3のように、平面を、欄jに外接する最小矩形を中心として、その左上、上、右上、左、右、左下、下、右下、同位の9個の領域に分けた時(以下、これらの領域を、「欄jを中心とする領域群」と呼ぶ)、欄iがどの領域に重なりをもつかを表したものである。欄iが広がる領域は、単一の場合もあるし、複数になる場合もある。ただし、欄は、単純多角形であるため、欄iの広がる領域が複数になる場合には、それらの領域同士は、領域の境界となる辺のうち少なくとも1個以上を共有している。

[0015] 許容される／されない相対配置パターンについて、例えば、日本国内で用いられる帳票においては、項目名欄iが、項目値欄jの項目名欄であるならば、項目名欄iは、項目値欄jを中心とする領域群のうち、少なくとも左、上のどちらかの領域と重なりを持つ(左、上以外の領域と重なりを持つかどうかは問わない)か、あるいは、同位の領域とのみ重なりを持つことを、仮定できる。言い換えれば、下記(1)～(3)の領域とのみ、重なりをもつ項目名欄iは、項目値欄jの項目名欄とはならない。

- ・ (1) 項目値欄jの左上の領域
- ・ (2) 項目値欄jの右上、右、右下、下、左下の領域、またはこれらのいくつか、あるいはすべてを結合した領域
- ・ (3) 上記(1)、(2)領域と、項目値欄jの同位の領域を結合した領

域

[0016] 許容される／されない相対配置パターンについて、例えば、日本国内で用いられる帳票においては、項目値欄 j が、項目名欄 i の項目値欄であるならば、項目値欄 j は、項目名欄 i を中心とする領域群のうち、少なくとも右、下のどちらかの領域と重なりを持つ（右、下以外の領域と重なりを持つかどうかは問わない）か、あるいは、同位の領域とのみ重なりを持つことを、仮定できる。言い換えれば、下記（１）～（３）の領域とのみ、重なりをもつ項目値欄 j は、項目名欄 i の項目値欄とはならない。

- ・（１）項目名欄 i の右下の領域
- ・（２）項目名欄 i の左下、左、左上、上、右上の領域、またはこれらを結合した領域
- ・（３）上記（１）、（２）の領域と、項目値欄 i の同位の領域を結合した領域

[0017] ここでは、許容される相対配置パターン、または許容されない配置パターンを、具体的に列挙することで指定しているが、「領域群のうち、〇〇領域と××領域の少なくともいずれかと重なりを持てば、許容される。」という許容条件や、非許容条件を表す論理式等により、指定してもよい。

[0018] 以上は、任意の項目名欄と項目値欄に関する「汎用モデル」である。特定の様式における、特定の項目名をもつ項目名欄と項目値欄について、これとは異なる仮定が必要な場合には、「項目名個別ルール」として、その項目名欄と項目値欄の相対配置パターンについて、許容できるもの、または許容できないものを決めておくようにしてもよい。

[0019] 情報処理装置 10 は、操作入力部 11 a、帳票入力部 11 b、表示部 12、OS (Operating System) 13、記憶部 14、エディタ 15 および帳票処理ツール 16 を有する。なお、これらの構成要素が同一装置上にある場合に限定されず、複数の装置に分散して配置されてもよい。

[0020] 操作入力部 11 a は、帳票の様式登録、機械的な処理などの実行命令や、項目名欄と項目値欄の関連情報の修正や編集などツールに対する操作を入力

するものであり、キーボードやマウス、タッチパネルなどを備えて構成される。帳票入力部 11b は、帳票のデータなどを入力するものであり、カメラ、スキャナなどを備えて構成される。また、表示部 12 は、各種処理結果を表示するものであり、ディスプレイやタッチパネル、スピーカなどを備えて構成される。

[0021] また、情報処理装置 10 本体は、OS 13 が実行され、OS 13 の処理にともない各種パラメータ等を編集するためのエディタ 15 や帳票を処理するための帳票処理ツール 16 が起動される。

[0022] 記憶部 14 は、例えば、HDD (Hard Disk Drive)、SSD (Solid State Drive)、光ディスク等の記憶装置である。記憶部 14 は、RAM (Random Access Memory)、フラッシュメモリ、NVS RAM (Non Volatile Static Random Access Memory) 等のデータを書き換え可能な半導体メモリであってもよい。記憶部 14 は、例えば、後述する欄の種別の判別情報を記憶する。

[0023] 次に、図 2 の例を用いて、帳票処理ツール 16 の論理構成について説明する。図 2 は、第 1 の実施形態に係る情報処理装置における帳票処理ツールの論理構成を示すブロック図である。図 2 に例示するように、帳票処理ツール 16 は、記載内容・位置取得部 161、欄の種別判別部 162、票分割部 163、関連付け許容判定部 164、評価関数生成部 165、制約条件付与部 166、求解部 167、登録部 168a、修正部 168b および許容相対配置パターン生成部 169 を有する。

[0024] 記載内容・位置取得部 161 は、帳票のデータを受け付け、帳票における欄の構成領域と欄に含まれる文字列を取得する。そして、記載内容・位置取得部 161 は、取得した各帳票中の各欄の文字列と各欄の構成領域を示す記載内容・位置データ 170a を作業データ蓄積部 170 に格納する。ここで、記載内容・位置データ 170a について図 4 を用いて説明する。図 4 に例示するように、記載内容・位置データ 170a は、欄を識別する「欄 ID」と、欄に含まれる「文字列」と、文字列を構成する「文字の種類」と、「欄

の構成領域」とが対応付けられている。欄の構成領域は、例えば、ベクターデータのように、欄の輪郭線、欄の頂点集合に関する情報であってもよいし、ラスタデータのように、帳票全体を行と列の格子状に並んだセルに分割して、欄がどのセルを含むかどうかの情報であってもよい。

[0025] 欄の種別判別部 162 は、帳票の各欄の種別を判別し、判別結果を欄種別データ 170b として作業データ蓄積部 170 に格納する。なお、帳票の各欄の種別を判別する方法はどのような方法であってもよい。また、欄の種別判別部 162 は、既知の帳票については、予め設定された欄の種別を判別するための情報を参照して、帳票の各欄の種別を判別するようにしてもよい。また、例えば、欄の種別判別部 162 は、レイアウトが同一の複数の帳票を取得し、各帳票における同一位置の各欄の記載内容を比較し、各帳票における同一位置の欄の記載内容が同一である場合には、当該欄の種別を項目名と判別し、異なる欄がある場合には、当該欄の種別を項目値と判別するようにしてもよい。

[0026] ここで、欄種別データ 170b について図 5 を用いて説明する。図 5 に例示するように、欄種別データ 170b は、欄を識別する「ID」と欄の種別を示す「種別」とが対応付けられている。図 5 の例では、種別が「TRUE」である場合には、欄が「項目名」であることを意味し、種別が「FALSE」である場合には、欄が「項目値」であることを意味する。

[0027] 票分割部 163 は、項目名欄と項目値欄の関連付けの効率化のために、票を分割する。票分割部 163 は、例えば、図 6 に例示するように、罫線で囲われている領域と、囲われていない領域で分割し、各領域における票をそれぞれ一つの票として扱う。そして、票分割部 163 は、記載内容・位置データ 170a のデータを票ごとに分けて票別記載内容・位置データ 170c として作業データ蓄積部 170 に格納する。なお、票分割部 163 による処理は省略してもよく、票を分割しない場合には、帳票全体を一つの票として扱い、以後の処理を行う。以降では、特に断りのない限り、分割した場合には分割して得られる各票を、分割しない場合には帳票全体を、単に「帳票」と

して記載し、区別しない。

[0028] 関連付け許容判定部 164 は、処理対象の帳票中の各欄を基準にして他の欄の存在領域の分布パターンである相対配置パターンをそれぞれ求め、該相対配置パターンと、予め設定された「許容される／されない相対配置パターン」のうちのいずれか一つまたは両方と比較し、処理対象の帳票中の項目名欄と項目値欄との関連付く可能性の有無を判定する。つまり、関連付け許容判定部 164 は、「許容される／されない相対配置パターン」に基づき、各項目値欄と項目名欄とが関連付く可能性があるかどうかを判定する。

[0029] 具体的には、まず、関連付け許容判定部 164 は、処理対象の帳票中の各欄を基準にして他の欄の存在領域の分布パターンである相対配置パターンを求め、「相対配置パターン」170dとして作業データ蓄積部 170に格納する。ここで、図7の例を用いて、相対配置パターンのデータ例について説明する。図7に例示するように、相対配置パターンは、項目名を識別する「項目名欄ID」と、項目値を識別する「項目値欄ID」と、項目名と項目値のいずれを基準とするかを示す「基準」と、相対配置パターンを識別する「相対配置パターンID」とが対応付けられている。

[0030] そして、関連付け許容判定部 164 は、相対配置パターン保持部 171に記憶される許容される／されない相対配置パターン171aを参照し、各欄の相対配置パターンが各項目値欄と項目名欄とが関連付く可能性があるかどうかを判定する。具体的には、項目名欄と項目値欄の各組合せについて、項目名欄を基準とする項目値欄の相対配置パターンが許容されるかどうか、および項目値欄を基準とする項目名欄の相対配置パターンが許容されるかどうかを、許容される／されない相対配置パターン171aを参照して判定し、どちらを基準とした場合にも許容される場合には、関連付く可能性があると判定し、そうでない場合には、関連付く可能性がないと判定する。ここで、図8の例を用いて、許容される／されない相対配置パターンのデータ例について説明する。許容される／されない相対配置パターン171aは、相対配置パターンIDを示す「ID」と、項目名と項目値のいずれを基準とするか

を示す「基準」と、左上、上、右上、左、同位、右、左下、下、右下のそれぞれに配置されているかどうか（1／0）を上記の順に2進数で表現した「相対配置パターン」と、許容されるか否かを示す「許容可否」とが対応付けられている。

[0031] 図7および図8の例を挙げて具体例に説明すると、例えば、関連付け許容判定部164は、図7に例示する相対配置パターンの1行目の、項目名欄ID「1」、項目値欄ID「3」、基準「項目名」について、相対配置パターンIDが「8」であるため、図8の許容される／されない相対配置パターン171aのID「8」に対応する許容可否を参照する。図8の例では、ID「8」に対応する許容可否が「○」であるため、関連付け許容判定部164は、許容されるものと判定する。また、関連付け許容判定部164は、項目名欄ID「1」、項目値欄ID「3」、基準「項目値」について、相対配置パターンIDが「548」であるため、許容される／されない相対配置パターン171aのID「548」に対応する許容可否を参照する。ここで、ID「548」に対応する許容可否も「○」であるため、関連付け許容判定部164は、項目名欄ID「1」の項目名欄と項目値欄ID「3」の項目値欄とは関連付く可能性があるとして判定し、判定結果を関連付け許容情報170eとして作業データ蓄積部170に格納する。また、もし、ID「548」に対応する許容可否が「×」であれば、関連付け許容判定部164は、項目名欄ID「1」の項目名欄と項目値欄ID「3」の項目値欄とは関連付く可能性がないと判定し、判定結果を関連付け許容情報170eとして作業データ蓄積部170に格納する。

[0032] ここで、図9の例を用いて、関連付け許容情報のデータ例について説明する。関連付け許容情報170eは、図9に例示するように、「項目名欄ID」および「項目値欄ID」の組み合わせと許容可能かどうかを示す「許容可能」とが対応付けられている。関連付け許容情報170eにおける許容可能が「1」である場合には、対応する「項目名欄ID」および「項目値欄ID」の組み合わせが関連付く可能性があることを意味する。上記の例のように

、関連付け許容判定部164は、項目名欄ID「1」の項目名欄と項目値欄ID「3」の項目値欄とは関連付く可能性があるとは判定した場合には、関連付け許容情報として、項目名欄ID「1」と項目値欄ID「3」と許容可能「1」とを対応付けて作業データ蓄積部170に格納する。

[0033] このように、関連付け許容判定部164は、欄の形状が矩形ではない場合であっても、項目値欄基準での相対配置の判定結果と、項目値名欄基準での相対配置の判定結果の両方を総合的に判断することにより、補完し合い、不要な対応関係の可能性を否定でき、誤って対応付けられる場合を減らすことが可能である。

[0034] ここで、図10および図11を用いて、項目値欄基準での相対配置の判定結果と、項目名欄基準での相対配置の判定結果の両方を総合的に判断する処理について説明する。図10および図11は、項目値欄基準での相対配置の判定結果と、項目名欄基準での相対配置の判定結果の両方を総合的に判断する処理例を説明する図である。

[0035] 例えば、図10に例示するように、関連付け許容判定部164は、項目値欄「j」基準で相対配置を判定すると、項目名欄「i」は「同位」の領域内にのみ存在しており、「左」「上」の領域には広がっていないものの、「同位」でよくわからないため、対応関係は否定できないものと判定する。一方、関連付け許容判定部164は、項目名欄「i」基準で相対配置を判定すると、項目値欄「j」は「左」「左上」「上」の領域内にのみ存在し、対応関係はあり得ないと判定する。このため、関連付け許容判定部164は、判定結果を総合すると、項目名欄「i」は、項目値欄「j」の項目名欄にはなりえないと判断する。

[0036] また、例えば、図11に例示するように、関連付け許容判定部164は、項目値欄「j」基準で相対配置を判定すると、項目名欄「i」は「右」「右下」「下」の領域内にのみ存在し、対応関係はありえないと判定する。一方、関連付け許容判定部164は、項目名欄「i」基準で相対配置を判定すると、項目値欄「j」は「同位」の領域内にのみ存在し、「右」「下」の領域には広がっていないものの、「同位」でよくわからないため、対応関係は否定できないと判定

する。このため、関連付け許容判定部 164 は、判定結果を総合すると、項目名欄 i は、項目値欄 j の項目名欄にはなりえないと判断する。

[0037] ここで、上述した許容される／されない相対配置パターンを具体的に列挙し、そのそれぞれに対して許容される／されない、を指定しておく方法とは別の方法として、様々な様式の項目名欄と項目値欄の関連付けの事例から、相対配置パターンを取得して、それを基に許容する配置パターンを生成することも可能である。例えば、許容相対配置パターン生成部 169 は、任意の項目名欄と項目値欄が関連付くこと（または関連付かないこと）が既知のときに、その項目名欄と項目値欄の相対配置パターンがいずれかにより、許容する（または許容しない）相対配置パターンを生成する。

[0038] 許容相対配置パターン生成部 169 は、所定の様式の帳票における項目名欄と項目値欄との関連付けから、相対配置パターンを取得して、取得した相対配置パターンを基に許容する相対配置パターンのデータを作成する。つまり、許容相対配置パターン生成部 169 では、項目名欄と項目値欄の関連付けに関する情報を用いて、許容される（あるいは、許容されない）相対配置パターンを生成する。なお、項目名欄と項目値欄の関連付け情報は、人手等で関連付けられた結果が、許容相対配置パターン生成部の入力として与えられるものとする。

[0039] ここで、図 12 および図 13 の例を用いて、相対配置パターンを生成する処理について説明する。図 12 および図 13 は、相対配置パターンを生成する処理例を説明する図である。例えば、図 12 に例示するように、許容相対配置パターン生成部 169 は、項目名欄 i と項目値欄 j とが関連付くことが既知である場合に、項目名欄 i を基準とした相対配置パターン「000001000」を許容される相対配置パターンとして生成し、相対配置パターン保持部 171 に格納する。また許容相対配置パターン生成部 169 は、項目値欄 j を基準とした相対配置パターン「100100100」を許容される相対配置パターンとして生成し、相対配置パターン保持部 171 に格納する。

[0040] また、図 13 に例示するように、許容相対配置パターン生成部 169 は、

項目名欄 i と項目値欄 j とが関連付くことが既知である場合に、項目名欄 i を基準とした相対配置パターン「000010000」を許容される相対配置パターンとして生成し、相対配置パターン保持部 171 に格納する。また許容相対配置パターン生成部 169 は、項目値欄 j を基準とした相対配置パターン「110100100」を許容される相対配置パターンとして生成し、相対配置パターン保持部 171 に格納する。

[0041] また、関連付け許容判定部 164 は、登録済み様式情報利用時、すなわち、様式保持部 172 で保持されている様式情報を使って処理対象帳票中の項目値欄と項目名欄の関連付けを求める際に、予め設定された許容される分布パターンおよび許容されない分布パターンのうちのいずれか一つまたは両方を項目名別に定義する項目名個別ルールが様式情報として設定されている場合には、該項目名個別ルールに応じて、処理対象の帳票中の項目名欄と項目値欄との関連付く可能性の有無を判定するようにしてもよい。項目名個別ルールとして、例えば、項目名の文字列が「契約 ID」の欄は右方向にある項目値欄と関連付くことや、項目名の文字列が「顧客タイプ」の欄は下方向にある項目値と関連付かないこと等が設定されているものとする。

[0042] 評価関数生成部 165 は、評価関数で必要な、帳票に応じて決定される係数を求め、評価関数を生成し、適用する評価関数 170 f として作業データ蓄積部 170 に格納する。評価関数は、各項目値欄に対して、項目名欄と項目値欄の関連の度合いを評価する観点を反映した評価項目で構成される。評価関数には、項目値欄と項目名欄の関連付けを求める際の評価項目として、項目値欄と項目名欄の配置上の近さ、欄の幅／高さの差あるいはその割合の小ささ、項目値欄と項目名欄の関連付けの少なさ、といった評価項目のうちのいずれかまたはすべてを用いる。また、登録済み様式情報利用時には、これらの評価項目に加え、様式登録時の帳票中に存在していた項目のうち、処理対象帳票中にも存在する項目の多さ、といった評価項目を用いる。

[0043] 例えば、帳票は、項目名の情報から項目値の記載内容を理解できる媒体であることを前提とすると、可読性の観点から、項目名欄は項目値欄となるべ

く距離が近い位置に配置されている可能性が高い。従って、「項目名欄と項目値欄の距離が近いほうが、評価が高い」という、汎用的な評価項目を仮定できる。また、項目名欄と項目値欄の関連は、欄のサイズによっても表現されており、項目名欄の幅と項目値欄の幅が近い、つまり、帳票の縦方向または横方向に欄を射影した際に、重なる範囲が大きいほうが関連性が高い傾向があることから、汎用的な評価項目として「項目名欄の幅と項目値欄の幅の差が小さいほうが評価値が高い」を仮定できる。

[0044] また、汎用的な、あるいは項目名個別ルールの特許される相対配置パターンについて、相対配置パターンごとに予め評価項目値を用意しておき、項目名欄と項目値欄の関連付けが未知の帳票の項目名欄と項目値欄の組合せに対し、それらの相対配置パターンに応じてその評価項目値を適用し、評価関数に含めてもよい。相対配置パターンごとに予め用意する評価項目値としては、例えば、項目名欄と項目値欄の関連付けが既知の帳票において、関連付けられる事例が多い相対配置パターンほど高い評価項目値とすることが考えられる。

[0045] また、登録済み様式情報利用時には、項目名欄に関連付けられる項目値欄の、相対配置パターンや欄内に記載されている文字の種類と、様式登録時の帳票におけるそれらとの一致度合いに関する評価項目を、評価関数に含めてもよい。

[0046] あるいは、後述の制約条件を満たさない場合にペナルティとなる評価項目を含めてもよい。例えば、項目値欄が項目名欄と関連付かない場合も許容しつつも、なるべくそのような項目値欄が残らないように、「項目値欄に項目名欄が1個も割り当てられなければ評価が低い」という評価項目を評価関数に含めてもよい。

[0047] あるいは、評価関数生成部165は、登録済み様式情報利用時には、項目が追加された可能性のある場合、すなわち項目名欄、項目値欄が正しく判別できていない欄が含まれている可能性のある場合、各項目名欄に関連付けられる項目値欄の個数と、様式登録時の帳票におけるそれらとの一致度合いに関

する評価項目を、評価関数に含めてもよい。なお、この場合には、評価関数生成部165は、様式登録時の帳票において各項目名欄に関連付けられる項目値欄の個数を、様式保持部172で保持されている様式情報から算出する。

[0048] 評価関数生成部165は、項目値欄と項目名欄の関連付けを求める際的评价関数として、上述したような評価項目候補のうち、いずれか一つまたは複数を、予め指定しておいた重みにより重み付けし、関数として生成する。

[0049] ここで、図14の例を用いて評価関数と制約条件を数式として表現する際に用いる記号について説明する。図14は、評価関数と制約条件を数式として表現する際に用いる記号について説明するための図である。図14では、「集合と要素」、「定数」、「変数」および「評価関数」についてそれぞれ記号の定義が記載されている。

[0050] そして、ここで、図16、図18の例を用いて、評価関数の一例について説明する。図16は、様式登録時における評価関数の一例について説明する図である。図16の(1)式では、 x_{ij} を変数として、評価値の総和が最小となるように解を求める。また、 c_{ij} は、図16の(2)式により求められる。

[0051] 図18は、登録済み様式情報利用時の評価関数の一例について説明する図である。図18の(1)式、その中の c_{ij} は、図16と同じである。図18の(2)式は、様式情報から求められる様式登録時の帳票に含まれる項目名セットのうち、処理対象帳票中のいずれの項目値欄とも関連づかない項目名セットの少なさ、すなわち、様式登録時の帳票中に存在していた項目のうち、処理対象帳票中にも存在する項目の多さに関する評価項目を重み付けしたものである。

[0052] 制約条件付与部166は、所定の制約条件を付与し、適用する制約条件170gとして作業データ蓄積部に格納する。制約条件には、各項目値欄が関連付く項目名欄の組み合わせを仮定したときに、少なくとも満たすべき条件を定義する。なお、様式登録時と、登録済み様式情報利用時とでは、制約条

件が異なるものとする。なお、以下に例示する制約条件は、一例であり、その他制約条件を追加で付与してもよい。例えば、一部の項目について、項目名欄と項目値欄が関連付けられること、関連付けられないことが与えられていれば、制約条件に追加で付与してもよい。

[0053] 例えば、制約条件付与部166は、処理対象の帳票中の項目名欄と項目値欄とが判別できる、様式登録時であれば、制約条件「項目名欄と項目値欄が関連付けられるのは、関連付けが許容される相対配置パターンである場合に限られる」と、制約条件「各項目値欄は、少なくとも1個以上の項目名欄と関連付く」と、制約条件「文字列が同じ項目名欄が複数ある場合、その中で同じ項目値欄と関連付くのは最大1個」と、制約条件「項目値欄同士は、それぞれ関連付く項目名欄に記載されている項目名のセットが互いに異なる」とを付与する。

[0054] また、例えば、制約条件付与部166は、処理対象の帳票中の項目値欄の中に、欄の種別を判別できていない欄が含まれている可能性のある、登録済み様式情報利用時であれば、制約条件の付与に先立ち、まず、様式保持部172で保持している様式情報から、様式登録時の帳票の各項目値欄に対し、項目名セットを求める。ここで項目名セットとは、各項目値欄に関連付けられる項目名欄の文字列の集合のことをいう。次に、各項目名セットに対し、処理対象帳票中の項目名欄の中から、項目名セットに含まれる文字列をもつ項目名欄を選び、割り当てることにより、項目名欄セットを作成する。ここで項目名欄セットとは、項目名欄の集合のことをいう。その際、同じ文字列、つまり項目名をもつ項目名欄が処理対象帳票中に複数あり、複数通りの割り当て方が可能であれば、その割り当て方の場合の数だけ、項目名欄セットを列挙する。その上で、制約条件「項目名欄と項目値欄が関連付けられるのは、関連付けが許容される相対配置パターンである場合に限られる」と、制約条件「各項目名欄セットは最大1個の項目値欄と関連付く」と、制約条件「各項目値欄は、最大1個の項目名欄セットと関連付く」と、制約条件「同じ項目名セットに対して生成される項目名欄セットのうち、項目値欄と関連付

くのは最大1個」とを付与する。また、登録済みの様式情報に含まれる項目のうち、処理対象として必須の項目が指定されている場合には、それら項目の各項目名セットについて、制約条件「同じ項目名セットから生成される項目名欄セットのうち、項目値欄に関連付くものはちょうど1個」をさらに付与する。

[0055] また、制約条件については、図15、図17に例示する。図15は、様式登録時における制約条件の一例を示す図である。図15に例示するように、「各項目値欄は、少なくとも1個以上の項目名欄と関連付く」ことと（図15の（1）式参照）、「項目名欄と項目値欄が関連付けられるのは、関連付けが許容される相対配置パターンである場合に限られる」ことと（図15の（2）式参照）、「文字列が同じ項目名欄が複数ある場合、その中で同じ項目値欄と関連付くのは最大1個」であることと（図15の（3）式参照）、「項目値欄同士は、それぞれ関連付く項目名欄に記載されている項目名のセットが互いに異なる」こと（図15の（4）式参照）が様式登録時の制約条件として付与される。

[0056] また、本実施形態では、登録済み様式情報利用時には、項目値欄と項目名欄セットの関連の有無を追加の変数として置き、図17に例示するように、「各項目値欄は、最大1個の項目名欄セットと関連付く」ことと（図17の（1）式参照）、「各項目名欄セットは最大1個の項目値欄と関連付く」ことと「各項目名セットについて、同じ項目名セットから生成される項目名欄セットのうち、項目値欄に関連付くものは最大1個」であることと（図17の（2）式参照）、「項目名欄と項目値欄が関連付けられるのは、関連付けが許容される相対配置パターンである場合に限られる」ことと（図17の（4）式参照）、さらに項目値欄と項目名欄の関連の有無を表す変数と、項目値欄と項目名欄セットの関連の有無を表す変数の間に成立する関係（図17の（5）式参照）が制約条件として付与される。また、様式登録時の帳票の項目のうち、登録済み様式情報利用時において必須の項目がある場合には、「各項目名セットについて、同じ項目名セットから生成される項目名欄セット

のうち、項目値欄に関連付くものはちょうど1個」であること（図17の（3）式参照）が追加の制約条件として付与される。

[0057] 求解部167は、制約条件付与部によって付与された制約条件を満たし、かつ評価関数生成部によって生成された評価関数を最適化するように、処理対象の帳票に含まれる項目名欄と項目値欄との関連付けを特定する。

[0058] 例えば、求解部167では、項目値欄と項目名欄の関連の有無を変数として置き、制約条件を満たし、かつ評価関数を最小化する解の組み合わせを厳密または近似的に求める、あるいは、この問題を近似した数式で表現し解くことで、項目名欄と項目値欄の関連付けを求める。数式の表現方法は、想定する制約条件と評価関数の意味が変わらなければ、ソルバーの利用有無、使用するソルバーにあわせて更に近似した数式表現に変えてもよい。

[0059] 求解部167は、例えば、図19に例示するように、項目名欄と項目値欄との関連付けを特定する。図19は、第1の実施形態に係る情報処理装置による求解例を示す図である。図19の例を挙げて具体的に説明すると、例えば、求解部167は、文字列「契約ID」を含む項目名欄と文字列「12345678」を含む項目値欄とが関連付くものとして特定する。そして、求解部167は、項目名欄と項目値欄の関連情報170hを作業データ蓄積部170に格納する。

[0060] また、図19に例示するように、例えば、求解部167は、文字列「契約者名義」を含む項目名欄および文字列「フリガナ」を含む項目名欄と、文字列「デンデンタロウ」を含む項目値欄とが関連付くものとして特定する。そして、求解部167は、項目名欄と項目値欄の関連情報170hを作業データ蓄積部170に格納する。

[0061] 図20に例示するように、項目名欄と項目値欄の関連情報170hは、項目値欄を識別する「項目値欄ID」と、項目名欄を識別する「項目名欄ID」とが対応付けられている。図20は、項目名欄と項目値欄の関連情報を示す図である。

[0062] 修正部168bは、求解部167によって登録された項目名欄と項目値欄

の関連情報 170h を修正する。例えば、修正部 168b は、ユーザの指示に応じて、作業データ蓄積部 170 に記憶された項目名欄と項目値欄の関連情報 170h を修正する。あるいは、修正部 168b では、ユーザの指示に応じて、項目名個別ルールを追加、変更、削除する。

[0063] 登録部 168a は、作業データ蓄積部 170 で保持されている、様式登録時の処理対象帳票の項目名欄と項目値欄の関連付けを求めた結果とその過程で作成されたデータの一部を、帳票の様式を識別する識別子に対応付けて、様式情報として登録する。具体的には、登録部 168a は、作業データ蓄積部 170 に記憶された、欄種別データ 170b と、票別記載内容・記載位置データ 170c と、相対配置パターン 170d と、項目名欄と項目値欄の関連情報 170h を、様式保持部 172 にそれぞれ欄種別データ 172b、票別記載内容・記載位置データ 172d、相対配置パターン 172e、項目名欄と項目値欄の関連情報 172a として登録する。また、設定された項目名個別ルール 172c を様式保持部 172 に登録する。

[0064] なお、以上の説明では、登録部では、様式情報として、作業データ蓄積部 170 に記憶された、欄種別データ 170b と、票別記載内容・記載位置データ 170c と、相対配置パターン 170d と、項目名欄と項目値欄の関連情報 170h を、様式保持部 172 に登録し、登録済みの様式情報利用時に、様式保持部 172 に登録されたこれらのデータから、様式登録時の帳票の各項目値欄に対し、項目名セットを求めるようにしているが、項目名セットは、様式登録時に登録部で求めておき、様式情報の一部として様式保持部 172 に登録しておき、登録済みの様式情報利用時には、その項目名セットを利用することとしてもよい。

[0065] また、以上の説明では、記載内容・位置取得部、欄の種別判定部、票分割部、関連付け許容判定部での処理により得られたデータ、相対配置パターン保持部で保持するデータ、様式保持部で保持するデータを、評価関数生成部、制約条件付与部、求解部から成る機能部でそれぞれ処理することにより、登録部によって登録された項目名のセットに基づいて、処理対象の帳票に含

まれる項目名欄と項目値欄との関連付けを特定しているが、これら評価関数生成部、制約条件付与部、求解部を、ひとつの「求解部」としてまとめた構成とすることや、さらにそのまとめた「求解部」において、入力となるデータは同じとしつつ、内部的には制約条件や評価関数を明示的には用いない方法により、登録部によって登録された項目名のセットに基づいて、処理対象の帳票に含まれる項目名欄と項目値欄との関連付けを特定するようにしてもよい。

[0066] [情報処理装置の処理手順]

次に、図21Aおよび図21Bを用いて、第1の実施形態に係る情報処理装置10の処理手順の例を説明する。図21Aおよび図21Bは、第1の実施形態に係る情報処理装置における全体的な処理の流れの一例を示すフローチャートである。

[0067] 図21Aおよび図21Bに示すように、記載内容・位置取得部161が、帳票のデータを受け付け、帳票における欄の構成領域と欄に含まれる文字列を取得し、作業データ蓄積部170に、記載内容・位置データ170aを格納する（ステップS101）。

[0068] そして、欄の種別判別部162が、帳票のデータを受け付け、あるいは作業データ蓄積部170で保持されている記載内容・位置データ170aを参照し、帳票の各欄の種別を判別し、作業データ蓄積部170に、欄種別データ170bを格納する（ステップS102）。

[0069] 続いて、票分割部163が、帳票のデータを受け付け、かつ作業データ蓄積部170で保持されている記載内容・位置データ170aを参照し、帳票を票に分割し、作業データ蓄積部170に、票別記載内容・位置データ170cを格納する（ステップS103）。

[0070] そして、関連付け許容判定部164は、分割された票についてすべて後述の処理が済んでいるか判定し（ステップS104）、分割された票がすべて処理済みである場合には（ステップS104肯定）、処理を終了する。また、分割された票がすべて処理済みでない場合には（ステップS104否定）

、関連付け許容判定部164が、作業データ蓄積部170で保持されている票別記載内容・位置データ170cから、未処理の票(以下、単に「帳票」と記載)のものを1個選択する(ステップS105)。

[0071] 続いて、関連付け許容判定部164が、選択された票別記載内容・位置データ170cと欄種別データ170bを参照し、項目名欄と項目値欄の各組合せに対し、互いの相対配置パターンを調べ、作業データ蓄積部170に、相対配置パターン170dを格納する(ステップS106)。

[0072] そして、様式登録時の処理である場合には(ステップS107肯定)、関連付け許容判定部164が、作業データ蓄積部170で保持されている相対配置パターン170dと、相対配置パターン保持部171で保持されている許容される／されない相対配置パターン171aを参照し、また、修正部168bから項目名個別ルールを受け付け、帳票内の項目名欄と項目値欄の各組合せに対し、関連付く可能性の有無を判定し、作業データ蓄積部170に、関連付け許容情報170eを格納する(ステップS108)。

[0073] そして、評価関数生成部165が、作業データ蓄積部170で保持されている票別記載内容・位置データ170cと欄種別データ170bと相対配置パターン170dを参照し、評価関数を生成し、作業データ蓄積部170に、適用する評価関数170fを格納する(ステップS109)。

[0074] 続いて、制約条件付与部166が、作業データ蓄積部170で保持されている票別記載内容・位置データ170cと欄種別データ170bと関連付け許容情報170eを参照し、付与する制約条件を列挙し、作業データ蓄積部170に、適用する制約条件170gを格納する(ステップS110)。

[0075] その後、求解部167が、作業データ蓄積部170で保持されている適用する評価関数170fと適用する制約条件170gを参照し、制約条件を満たし評価関数を最適化する解を求め、作業データ蓄積部170に、項目名欄と項目値欄の関連情報170hを格納する(ステップS111)。

[0076] そして、修正部168bが、作業データ蓄積部170で保持されている票別記載内容・位置データ170cと項目名欄と項目値欄の関連情報170h

を参照し、人に提示し、人による修正内容を受け付け、修正内容に応じて、項目名欄と項目値欄の関連情報 170h を更新する、または項目名個別ルールを追加・変更・削除する（ステップ S 112）。

[0077] そして、項目名個別ルールの追加・変更・削除がある場合には（ステップ S 113 肯定）、ステップ S 108 の処理に戻る。また、項目名個別ルールの追加・変更・削除がない場合には（ステップ S 113 否定）、登録部 168a が、作業データ蓄積部 170 で保持されている票別記載内容・位置データ 170c と欄種別データ 170b と相対配置パターン 170d と項目名欄と項目値欄の関連情報 170h を参照し、また、修正部 168b から項目名個別ルールを受け付け、これらに対応づけ、ひとつの様式情報として、様式保持部 172 に登録する（ステップ S 114）。

[0078] また、ステップ S 107 において、様式登録時でない場合には（ステップ S 107 否定）、関連付け許容判定部 164 は、様式保持部 172 で保持されている様式情報をすべて適用済みであるか判定する（ステップ S 115）。

[0079] この結果、様式保持部 172 で保持されている様式情報をすべて適用済みではないと判定された場合には（ステップ S 115 否定）、関連付け許容判定部 164 が、様式保持部 172 で保持されている様式情報から、未適用の様式情報を 1 個選択する（ステップ S 116）。

[0080] 続いて、関連付け許容判定部 164 が、作業データ蓄積部 170 で保持されている相対配置パターン 170d と、相対配置パターン保持部 171 で保持されている許容される／されない相対配置パターン 171a と、様式情報に含まれる項目名別ルール 172c を参照し、帳票内の項目名欄と項目値欄の各組合せに対し、関連付く可能性の有無を判定し、作業データ蓄積部 170 に、関連付け許容情報 170e を格納する（ステップ S 117）。

[0081] 続いて、評価関数生成部 165 が、作業データ蓄積部 170 で保持されている票別記載内容・位置データ 170c と欄種別データ 170b と相対配置パターン 170d と、様式情報に含まれる票別記載内容・位置データ 172

d、欄種別データ172b、相対配置パターン172e、項目名欄と項目値欄の関連情報172aを参照し、評価関数を生成し、作業データ蓄積部170に、適用する評価関数170fを格納する（ステップS118）。

[0082] そして、制約条件付与部166が、作業データ蓄積部170で保持されている票別記載内容・位置データ170cと欄種別データ170bと関連付け許容情報170eと、様式情報に含まれる票別記載内容・位置データ172d、欄種別データ172b、項目名欄と項目値欄の関連情報172aを参照し、付与する制約条件を列挙し、作業データ蓄積部170に、適用する制約条件を格納する（ステップS119）。

[0083] その後、求解部167が、作業データ蓄積部170で保持されている適用する評価関数170fと適用する制約条件170gを参照し、制約条件を満たし評価関数を最適化する解を求め、作業データ蓄積部170に、項目名欄と項目値欄の関連情報を格納する。また、様式情報から求められる項目名セットのうち、項目値欄と関連付けられた項目名セットの数をカウントする（ステップS120）。

[0084] そして、項目値欄と関連付けられた項目名セットの数について、現在選択されている様式情報適用時の方が、これまでで最良の様式情報適用時より大きくない場合には（ステップS121否定）、ステップS115の処理に戻る。また、項目値欄と関連付けられた項目名セットの数について、現在選択されている様式情報適用時の方が、これまでで最良の様式情報適用時より大きい場合には（ステップS121肯定）、現在選択されている様式情報と、項目名欄と項目値欄の関連情報と、項目値欄と関連付けられた項目名セットの数を、これまでで最良の様式情報適用時のものとして、保持し（ステップS122）、ステップS115の処理に戻る。

[0085] また、ステップS115において、様式保持部172で保持されている様式情報をすべて適用済みであると判定された場合には（ステップS115肯定）、これまでで最良の様式情報適用時のものとして保持されている、項目名欄と項目値欄の関連情報を、選択された票における、項目名欄と項目値欄

の関連情報とし（ステップS 1 2 3）、ステップS 1 0 4に戻る。

[0086] 次に、図 2 2 を用いて、第 1 の実施形態に係る情報処理装置 1 0 の許容される／されない相対配置パターン登録時の処理手順の例を説明する。図 2 2 は、図 2 2 は、第 1 の実施形態に係る情報処理装置 1 0 の許容相対配置パターン生成部における処理の流れの一例を示すフローチャートである。

[0087] 図 2 2 に示すように、許容相対配置パターン生成部 1 6 9 が、帳票処理ツール外から、票別記載内容・位置データ、欄種別データ、項目名欄と項目値欄の関連情報に相当する入力データを受け付ける（ステップS 2 0 1）。そして、許容相対配置パターン生成部 1 6 9 が、未処理の関連付けられた項目名欄と項目値欄の組合せがあるか判定し（ステップS 2 0 2）、組合せがないと判定した場合には（ステップS 2 0 2 否定）、処理を終了する。

[0088] また、許容相対配置パターン生成部 1 6 9 が、未処理の関連付けられた項目名欄と項目値欄の組合せがあると判定した場合には（ステップS 2 0 2 肯定）、未処理の関連付けられた項目名欄と項目値欄の組合せを 1 個選択する（ステップS 2 0 3）。そして、許容相対配置パターン生成部 1 6 9 が、受け付けた、票別記載内容・位置データを参照し、選択された組合せについて、項目名欄を基準とした項目値欄の相対配置パターンと、項目値欄を基準とした項目名欄の相対配置パターンを調べる（ステップS 2 0 4）。

[0089] 続いて、許容相対配置パターン生成部 1 6 9 が、項目名欄を基準とした項目値欄の相対配置パターンと、項目値欄を基準とした項目名欄の相対配置パターンのそれぞれについて、相対配置パターン保持部 1 7 1 で保持されている、許容される／されない相対配置パターン 1 7 1 a の、該当の相対配置パターンの関連付け事例数を 1 加算する（ステップS 2 0 5）。

[0090] そして、許容相対配置パターン生成部 1 6 9 が、相対配置パターンの関連付け事例数が閾値を超えているか判定し（ステップS 2 0 6）、相対配置パターンの関連付け事例数が閾値を超えていないと判定した場合には（ステップS 2 0 6 否定）、ステップS 2 0 2 の処理に戻る。

[0091] また、許容相対配置パターン生成部 1 6 9 は、相対配置パターンの関連付

け事例数が閾値を超えていると判定した場合には（ステップS206肯定）、相対配置パターン保持部171で保持されている、許容される／されない相対配置パターン171aの、該当の相対配置パターンの許容可否を「○」に更新し（ステップS207）、ステップS202の処理に戻る。

[0092] 次に、図23～図25を用いて、第1の実施形態に係る情報処理装置10の記載内容・位置取得部161による処理手順の例を説明する。図23～図25は、第1の実施形態に係る情報処理装置の記載内容・位置取得部における処理の流れの一例を示すフローチャートである。

[0093] まず、図23を用いて、帳票の欄が罫線で囲われている場合の記載内容・位置取得部161による処理を説明する。図23に例示するように、記載内容・位置取得部161は、オブジェクト総数を n と設定し、 i の値を「1」と設定し、未探索集合を $[1 \dots n]$ と設定する（ステップS301）。そして、記載内容・位置取得部161は、 i の値が n の値よりも大きいか判定する（ステップS302）。

[0094] この結果、記載内容・位置取得部161は、 i の値が n の値よりも大きくない場合には（ステップS302否定）、 i 番目のオブジェクトの上と左に罫線があるか判定する（ステップS303）。この結果、記載内容・位置取得部161は、 i 番目のオブジェクトの上と左に罫線がない場合には（ステップS303否定）、ステップS308の処理に進む。また、記載内容・位置取得部161は、 i 番目のオブジェクトの上と左に罫線がある場合には（ステップS303肯定）、記載内容・位置取得部161は、 i 番目から罫線を時計回りに周回したオブジェクト集合を取得する（ステップS304）。

[0095] そして、記載内容・位置取得部161は、周回したオブジェクトを取得できたか判定する（ステップS305）。この結果、記載内容・位置取得部161は、周回したオブジェクトを取得できなかった場合には（ステップS305否定）、ステップS308の処理に進む。また、記載内容・位置取得部161は、周回したオブジェクトを取得できた場合には（ステップS305肯定）、周回したオブジェクト集合を1つの欄として、文字列や領域の情報

を、作業データ蓄積部170に記載内容・位置データ170aとして保存する（ステップS306）。

[0096] 続いて、記載内容・位置取得部161は、周回したオブジェクト集合を未探索集合から削除し（ステップS307）、 i の値に「1」を追加し（ステップS308）、ステップS302の処理に戻る。

[0097] ステップS302において、記載内容・位置取得部161は、 i の値が n の値よりも大きい場合には（ステップS302肯定）、未探索集合にオブジェクトがないか判定し（ステップS309）、未探索集合にオブジェクトがある場合には（ステップS309否定）、未探索集合から任意のオブジェクトを選択する（ステップS310）。

[0098] そして、記載内容・位置取得部161は、オブジェクトを1つの欄として、作業データ蓄積部170に記載内容・位置データ170aを保存し（ステップS311）、未探索集合から、選択したオブジェクトのインデックスを削除し（ステップS312）、ステップS309の処理に戻る。また、記載内容・位置取得部161は、未探索集合にオブジェクトがない場合には（ステップS309肯定）、処理を終了する。

[0099] 次に、図24を用いて、帳票の欄が罫線で囲われていない場合の記載内容・位置取得部161による処理を説明する。図24に例示するように、記載内容・位置取得部161は、オブジェクト総数を n と設定し、 i の値を「1」と設定し、未探索集合を $[1 \dots n]$ と設定する（ステップS401）。そして、記載内容・位置取得部161は、未探索集合にオブジェクトがないか判定し（ステップS402）、未探索集合にオブジェクトがある場合には（ステップS402否定）、未探索集合から任意のオブジェクトを選択する（ステップS403）。

[0100] そして、記載内容・位置取得部161は、オブジェクトを1つの欄として、作業データ蓄積部170に記載内容・位置データ170aを保存し（ステップS404）、未探索集合から、選択したオブジェクトのインデックスを削除し（ステップS405）、ステップS402の処理に戻る。また、記載

内容・位置取得部161は、未探索集合にオブジェクトがない場合には（ステップS402肯定）、処理を終了する。

[0101] 次に、図25を用いて、帳票が画像データである場合の記載内容・位置取得部161による処理を説明する。記載内容・位置取得部161は、画像から罫線を抽出し（ステップS501）、罫線から画像を分割する（ステップS502）。そして、記載内容・位置取得部161は、画像の総数を n と設定し、 i の値を「1」と設定する（ステップS503）。

[0102] そして、記載内容・位置取得部161は、 i の値が n の値よりも大きいか判定する（ステップS504）。この結果、記載内容・位置取得部161は、 i の値が n の値よりも大きくない場合には（ステップS504否定）、画像上の文字を認識し（ステップS505）、認識した文字と画像の領域に関する情報を作業データ蓄積部170に記載内容・位置データ170aとして保存する（ステップS506）。

[0103] 続いて、記載内容・位置取得部161は、 i の値に「1」を追加し（ステップS507）、ステップS504の処理に戻る。また、ステップS504において、記載内容・位置取得部161は、 i の値が n の値よりも大きいと判定した場合には（ステップS504肯定）、処理を終了する。

[0104] [第1の実施形態の効果]

このように、第1の実施形態に係る情報処理装置10は、様式情報を登録し、登録済み様式情報利用時には、登録した様式情報から求められる項目名のセットに基づいて、処理対象の帳票に含まれる項目名欄と項目値欄との関連付けを特定する。このため、帳票の欄の配置に変更があった場合であっても、帳票に含まれる項目名欄と項目値欄との関連付けを特定すること可能である。

[0105] また、第1の実施形態に係る情報処理装置10は、欄の形状や、欄同士が隣接しているかどうかに関わらず、欄を基準にした他の欄の存在領域の分布パターン（相対配置パターン）により、項目名欄と項目値欄の関連付く可能性の有無を判断する。このため、情報処理装置10では、従来技術で仮定して

いた項目名欄や項目値欄の形状やそれらの配置パターンに関する条件を満たさない帳票に対しても、項目名欄と項目値欄の関連付けが可能になる。

[0106] 例えば、従来技術では、各項目値欄に対して隣接する項目名欄のみとの関連の有無を判断し、かつ2つの項目名欄同士の配置から項目名欄同士の親子関係を判断し、さらに、これらを組み合わせて関連の有無を間接的に決定していたため、項目値欄と隣接しない項目名欄との関連付けが不適切になる場合があった。これに対して、第1の実施形態に係る情報処理装置10は、任意の項目名欄、項目値欄の組合せに関して、関連付く可能性の有無を判断するので、項目値欄と隣接しない項目名欄との関連付けが不適切になる場合をなくすことが可能である。

[0107] また、第1の実施形態に係る情報処理装置10は、様式登録時、関連付く可能性のある組合せの中で、各項目値欄がいずれかの項目名欄と対応づく制約条件を付与することにより、様式登録時は、どの項目名欄とも関連付けられない項目値欄が残らず、すべての項目値欄に対して項目名欄が関連付けられるよう、項目名欄と項目値との関連付けを求めることができる。

[0108] また、第1の実施形態に係る情報処理装置10は、許容される相対配置パターンを項目名別に指定したうえで、項目名欄と項目値欄の関連付く可能性の有無を判断し、他の制約条件とともに解を探索できるようにすることにより、許容される／されない相対配置パターンが汎用的な項目とは異なる項目であっても、項目名欄と項目値との関連付けを適切に求めることができる。

[0109] また、第1の実施形態に係る情報処理装置10は、登録済み様式情報利用時の処理対象帳票に、様式登録時の帳票に存在しない項目が含まれていても、再度様式登録を行うことなく、様式登録時の帳票に存在していた全項目について、項目名欄と項目値との関連付けを求めることができる。

[0110] また、第1の実施形態に係る情報処理装置10は、例えば、様式登録時の帳票と、登録済み様式情報利用時の処理対象帳票で、項目名が縦と横の両方に配置されている場合に縦と横が入れ替わった場合や、項目名の親子関係の順序性が変わった場合でも、項目値欄の位置を特定できる。

[0111] (システム構成等)

また、図示した各装置の各構成要素は機能概念的なものであり、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、各装置の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷や使用状況などに応じて、任意の単位で機能的または物理的に分散・統合して構成することができる。さらに、各装置にて行なわれる各処理機能は、その全部または任意の一部が、CPUおよび当該CPUにて解析実行されるプログラムにて実現され、あるいは、ワイヤードロジックによるハードウェアとして実現され得る。

[0112] また、本実施の形態において説明した各処理のうち、自動的におこなわれるものとして説明した処理の全部または一部を手動的におこなうこともでき、あるいは、手動的におこなわれるものとして説明した処理の全部または一部を公知の方法で自動的におこなうこともできる。この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

[0113] (プログラム)

また、上記実施形態において説明した情報処理装置が実行する処理をコンピュータが実行可能な言語で記述したプログラムを作成することもできる。例えば、実施形態に係る情報処理装置10が実行する処理をコンピュータが実行可能な言語で記述した関連付けプログラムを作成することもできる。この場合、コンピュータが関連付けプログラムを実行することにより、上記実施形態と同様の効果を得ることができる。さらに、かかる関連付けプログラムをコンピュータ読み取り可能な記録媒体に記録して、この記録媒体に記録された関連付けプログラムをコンピュータに読み込ませて実行することにより上記実施形態と同様の処理を実現してもよい。

[0114] 図26は、関連付けプログラムを実行するコンピュータを示す図である。図26に例示するように、コンピュータ1000は、例えば、メモリ101

0と、CPU1020と、ハードディスクドライブインタフェース1030と、ディスクドライブインタフェース1040と、シリアルポートインタフェース1050と、ビデオアダプタ1060と、ネットワークインタフェース1070とを有し、これらの各部はバス1080によって接続される。

[0115] メモリ1010は、図26に例示するように、ROM (Read Only Memory) 1011及びRAM1012を含む。ROM1011は、例えば、BIOS (Basic Input Output System) 等のブートプログラムを記憶する。ハードディスクドライブインタフェース1030は、図26に例示するように、ハードディスクドライブ1090に接続される。ディスクドライブインタフェース1040は、図26に例示するように、ディスクドライブ1100に接続される。例えば磁気ディスクや光ディスク等の着脱可能な記憶媒体が、ディスクドライブ1100に挿入される。シリアルポートインタフェース1050は、図26に例示するように、例えばマウス1110、キーボード1120に接続される。ビデオアダプタ1060は、図26に例示するように、例えばディスプレイ1130に接続される。

[0116] ここで、図26に例示するように、ハードディスクドライブ1090は、例えば、OS1091、アプリケーションプログラム1092、プログラムモジュール1093、プログラムデータ1094を記憶する。すなわち、上記の、関連付けプログラムは、コンピュータ1000によって実行される指令が記述されたプログラムモジュールとして、例えばハードディスクドライブ1090に記憶される。

[0117] また、上記実施形態で説明した各種データは、プログラムデータとして、例えばメモリ1010やハードディスクドライブ1090に記憶される。そして、CPU1020が、メモリ1010やハードディスクドライブ1090に記憶されたプログラムモジュール1093やプログラムデータ1094を必要に応じてRAM1012に読み出し、各種処理手順を実行する。

[0118] なお、関連付けプログラムに係るプログラムモジュール1093やプログラムデータ1094は、ハードディスクドライブ1090に記憶される場合

に限られず、例えば着脱可能な記憶媒体に記憶され、ディスクドライブ等を介してCPU 1020によって読み出されてもよい。あるいは、関連付けプログラムに係るプログラムモジュール1093やプログラムデータ1094は、ネットワーク（LAN (Local Area Network)、WAN (Wide Area Network) 等）を介して接続された他のコンピュータに記憶され、ネットワークインタフェース1070を介してCPU 1020によって読み出されてもよい。

符号の説明

- [0119] 10 情報処理装置
 - 11 a 操作入力部
 - 11 b 帳票入力部
 - 12 表示部
 - 13 OS
 - 14 記憶部
 - 15 エディタ
 - 16 帳票処理ツール
 - 161 記載内容・位置取得部
 - 162 欄の種別判別部
 - 163 票分割部
 - 164 関連付け許容判定部
 - 165 評価関数生成部
 - 166 制約条件付与部
 - 167 求解部
 - 168 a 登録部
 - 168 b 修正部
 - 169 許容相対配置パターン生成部
 - 170 作業データ蓄積部
 - 170 a 記載内容・位置データ

- 170 b、172 b 欄種別データ
- 170 c、172 d 票別記載内容・位置データ
- 170 d、172 e 相対配置パターン
- 170 e 関連付け許容情報
- 170 f 適用する評価関数
- 170 g 適用する制約条件
- 170 h、172 a 項目名欄と項目値欄の関連情報
- 171 相対配置パターン保持部
- 171 a 許容される／されない相対配置パターン
- 172 様式保持部
- 172 c 項目名別ルール

請求の範囲

- [請求項1] 所定の帳票における項目値欄と関連付けられた項目名欄に対応付けられた項目名のセットを登録する登録部と、
- 前記登録部によって登録された項目名のセットに基づいて、処理対象の帳票に含まれる項目名欄と項目値欄との関連付けを特定する求解部と
- を有することを特徴とする情報処理装置。
- [請求項2] 前記処理対象の帳票中の各欄を基準にして他の欄の存在領域の分布パターンをそれぞれ求め、該分布パターンと、予め設定された許容される分布パターンおよび許容されない分布パターンのうちのいずれか一つまたは両方と比較し、前記処理対象の帳票中の項目名欄と項目値欄との関連付く可能性の有無を判定する関連付け許容判定部をさらに有することを特徴とする請求項1に記載の情報処理装置。
- [請求項3] 前記処理対象の帳票中の項目名欄と項目値欄とが判別できる場合に、項目名欄と項目値欄が関連付けられるのは、欄の相対的な配置関係の観点から関連付けが許容される場合に限られるという制約条件と、各項目値欄は、少なくとも1個以上の項目名欄と関連付くという制約条件と、文字列が同じ項目名欄が複数ある場合、その中で同じ項目値欄と関連付くのは最大1個という制約条件と、項目値欄同士は、それぞれ関連付く項目名欄に記載されている項目名のセットが互いに異なるという制約条件とを付与する制約条件付与部をさらに有し、
- 前記求解部は、前記制約条件付与部によって付与された制約条件を満たすように、処理対象の帳票に含まれる項目名欄と項目値欄との関連付けを特定することを特徴とする請求項1に記載の情報処理装置。
- [請求項4] あらかじめ項目値欄と項目名欄の関連付けが特定されている帳票の様式情報を用いて別の帳票を処理する場合に、登録されている様式情報で保持されている項目名セットそれぞれについて、処理対象の帳票中の項目名欄セットを生成し、項目名欄と項目値欄が関連付けられる

のは、欄の相対的な配置関係の観点から関連付けが許容される場合に限られるという制約条件と、各項目値欄は、最大1個の項目名欄セットと関連付くという制約条件と、各項目名欄セットは最大1個の項目値欄と関連付くという制約条件と、同じ項目名セットに対して生成される項目名欄セットのうち、項目値欄と関連付くのは最大1個という制約条件とを付与する制約条件付与部をさらに有し、

前記求解部は、前記制約条件付与部によって付与された制約条件を満たすように、処理対象の帳票に含まれる項目値欄と項目名欄セットの関連付けと、項目名欄と項目値欄との関連付けを特定することを特徴とする請求項1に記載の情報処理装置。

[請求項5] 前記関連付け許容判定部は、予め設定された許容される分布パターンおよび許容されない分布パターンのうちのいずれか一つまたは両方を項目名別に定義する項目名個別ルールが設定されている場合には、該項目名個別ルールに応じて、前記処理対象の帳票中の項目名欄と項目値欄との関連付く可能性の有無を判定することを特徴とする請求項2に記載の情報処理装置。

[請求項6] 所定の様式の帳票における項目名欄と項目値欄との関連付けから、相対配置パターンを取得して、取得した相対配置パターンを基に許容する配置パターンのデータを作成する許容相対配置パターン生成部をさらに有することを特徴とする請求項1に記載の情報処理装置。

[請求項7] 情報処理装置によって実行される関連付け方法であって、
所定の帳票における項目値欄と関連付けられた項目名欄に対応付けられた項目名のセットを登録する登録工程と、
前記登録工程によって登録された項目名のセットに基づいて、処理対象の帳票に含まれる項目名欄と項目値欄との関連付けを特定する求解工程と
を含むことを特徴とする関連付け方法。

[請求項8] 所定の帳票における項目値欄と関連付けられた項目名欄に対応付け

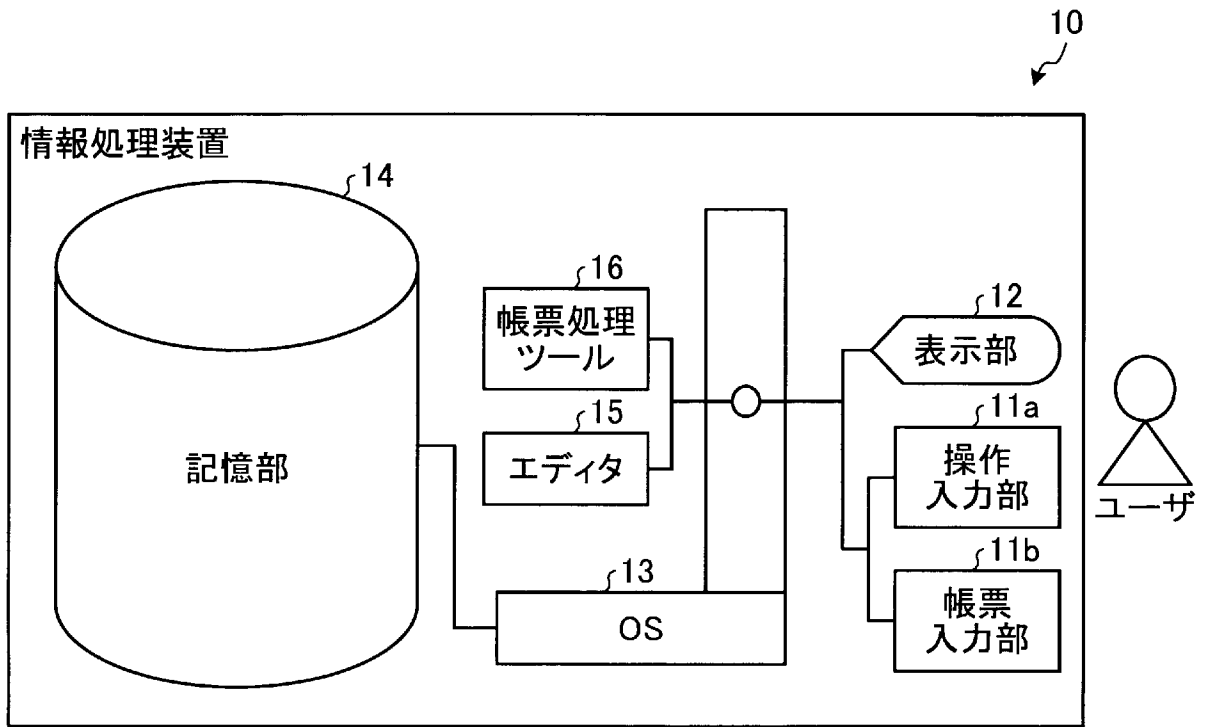
られた項目名のセットを登録する登録ステップと、

前記登録ステップによって登録された項目名のセットに基づいて、
処理対象の帳票に含まれる項目名欄と項目値欄との関連付けを特定する
求解ステップと

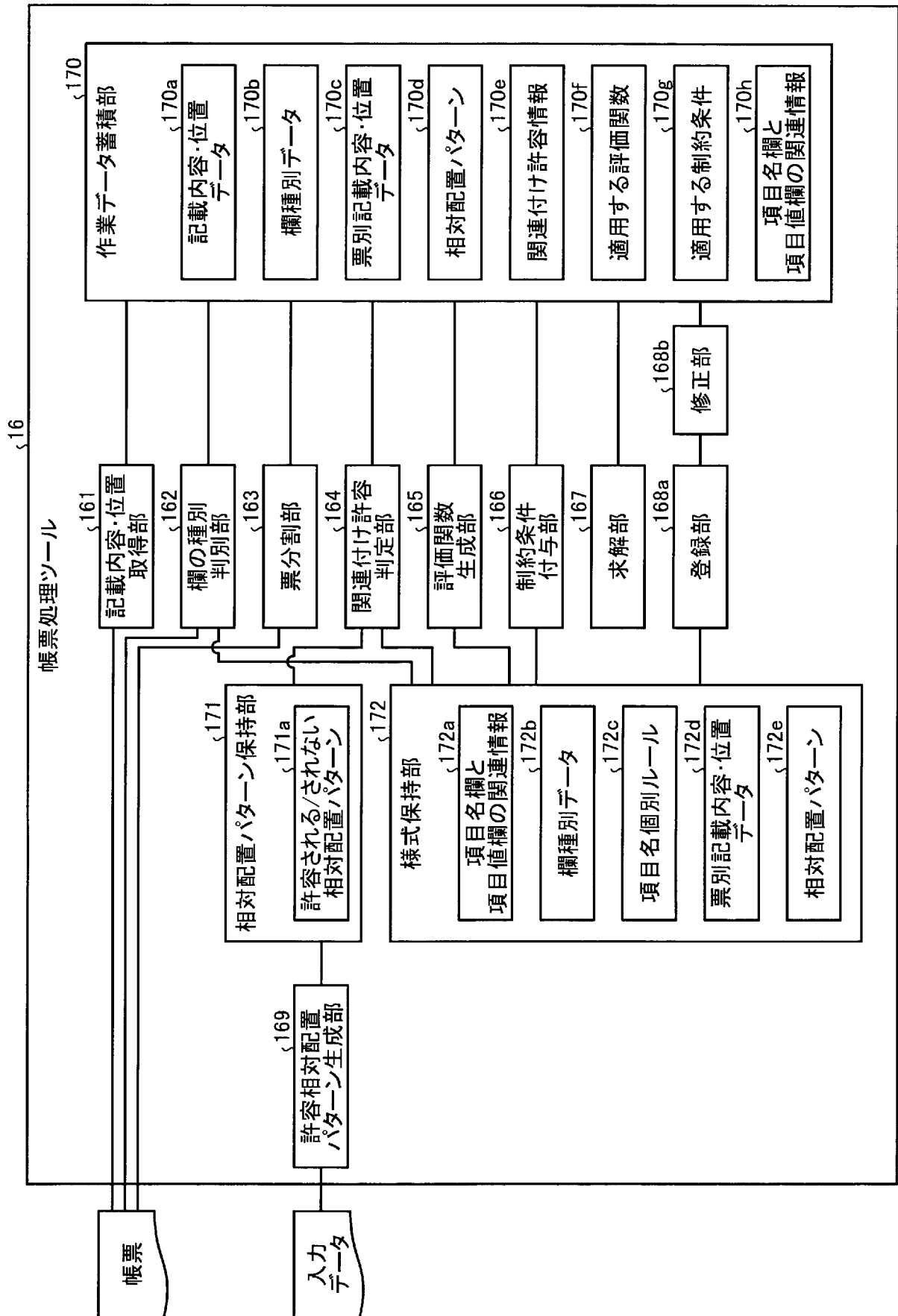
をコンピュータに実行させることを特徴とする関連付けプログラム

。

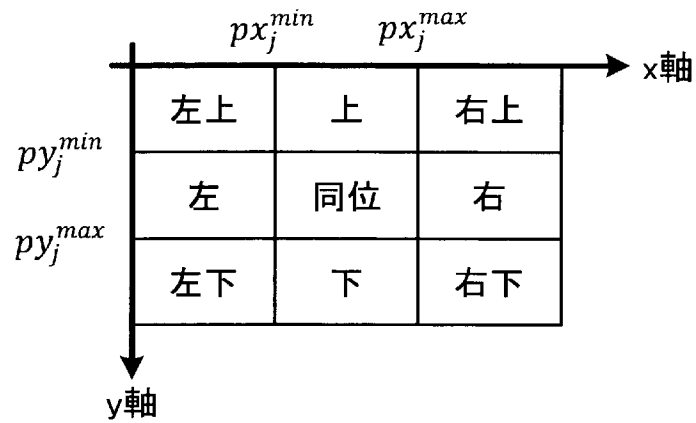
[図1]



[図2]



[図3]



[図4]

欄ID	文字列	文字の種類	欄の構成領域
1	基本情報	漢字	...
2	契約ID	漢字、アルファベット	...
3	12345678	数字	...
4	契約者名義	漢字	...
5	フリガナ	片仮名	...
6	デンデンタロウ	片仮名	...
7	氏名	漢字	...
8	電電太郎	漢字	...
9	申込情報	漢字	...
10	顧客タイプ	漢字、片仮名	...
11	新規	漢字	...

[図5]

欄ID	種別
1	TRUE
2	TRUE
3	FALSE
4	TRUE
5	TRUE
6	FALSE
7	TRUE
8	FALSE
9	TRUE
10	TRUE
11	FALSE

[6]

様式 v1.0 請求書 会社名: 株式会社XXXX 部署名: 通販会社販売担当 担当者名: 鈴木太郎 請求日: 2018年10月15日 次の通り、商品の発送を完了しましたので、ご報告申し上げます。 期限内に代金の支払いをお願いいたします。 合計金額: 6,500円 お支払期限: 2018年10月30日	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>商品名</th> <th>数量</th> <th>単位</th> <th>単価</th> <th>金額</th> </tr> </thead> <tbody> <tr> <td>キャットハウス CH-001</td> <td>1</td> <td>個</td> <td>4,000</td> <td>4,000</td> </tr> <tr> <td>猫砂(7kg) CS-007</td> <td>5</td> <td>個</td> <td>300</td> <td>1,500</td> </tr> <tr> <td>猫用おもちゃセット CTS-077</td> <td>1</td> <td>個</td> <td>1,000</td> <td>1,000</td> </tr> </tbody> </table>	商品名	数量	単位	単価	金額	キャットハウス CH-001	1	個	4,000	4,000	猫砂(7kg) CS-007	5	個	300	1,500	猫用おもちゃセット CTS-077	1	個	1,000	1,000	10月30日 abc銀行def支店(普) 1234567 口座番号: 株式会社XXXX 口座名義:	契約者情報 氏名 フリガナ エステイティ タロウ 漢字 NIT 太郎 〒111-1111 住所 NTT県NTT市NTT町1-2-3 NTTビル5F 連絡先 TEL: 111-2222-3333 e-mail: ntt_taro@ntt.co.jp	配送先情報 氏名 フリガナ エステイティ タロウ 漢字 NIT 太郎 〒111-1111 住所 NTT県NTT市NTT町1-2-3 NTTビル5F 連絡先 TEL: 111-2222-3333 e-mail: ntt_taro@ntt.co.jp お届け日 2018年 10月 3日 (時間帯: 16:00)	備考
商品名	数量	単位	単価	金額																					
キャットハウス CH-001	1	個	4,000	4,000																					
猫砂(7kg) CS-007	5	個	300	1,500																					
猫用おもちゃセット CTS-077	1	個	1,000	1,000																					



様式 v1.0 請求書 会社名: 株式会社XXXX 部署名: 通販会社販売担当 担当者名: 鈴木太郎 請求日: 2018年10月15日 次の通り、商品の発送を完了しましたので、ご報告申し上げます。 期限内に代金の支払いをお願いいたします。 合計金額: 6,500円 お支払期限: 2018年10月30日	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>商品名</th> <th>数量</th> <th>単位</th> <th>単価</th> <th>金額</th> </tr> </thead> <tbody> <tr> <td>キャットハウス CH-001</td> <td>1</td> <td>個</td> <td>4,000</td> <td>4,000</td> </tr> <tr> <td>猫砂(7kg) CS-007</td> <td>5</td> <td>個</td> <td>300</td> <td>1,500</td> </tr> <tr> <td>猫用おもちゃセット CTS-077</td> <td>1</td> <td>個</td> <td>1,000</td> <td>1,000</td> </tr> </tbody> </table>	商品名	数量	単位	単価	金額	キャットハウス CH-001	1	個	4,000	4,000	猫砂(7kg) CS-007	5	個	300	1,500	猫用おもちゃセット CTS-077	1	個	1,000	1,000	10月30日 abc銀行def支店(普) 1234567 口座番号: 株式会社XXXX 口座名義:	契約者情報 氏名 フリガナ エステイティ タロウ 漢字 NIT 太郎 〒111-1111 住所 NTT県NTT市NTT町1-2-3 NTTビル5F 連絡先 TEL: 111-2222-3333 e-mail: ntt_taro@ntt.co.jp	配送先情報 氏名 フリガナ エステイティ タロウ 漢字 NIT 太郎 〒111-1111 住所 NTT県NTT市NTT町1-2-3 NTTビル5F 連絡先 TEL: 111-2222-3333 e-mail: ntt_taro@ntt.co.jp お届け日 2018年 10月 3日 (時間帯: 16:00)	備考
商品名	数量	単位	単価	金額																					
キャットハウス CH-001	1	個	4,000	4,000																					
猫砂(7kg) CS-007	5	個	300	1,500																					
猫用おもちゃセット CTS-077	1	個	1,000	1,000																					

[図7]

項目名欄ID	項目値欄ID	基準	相対配置パターンID
1	3	項目名	8
1	3	項目値	548
2	3	項目名	8
2	3	項目値	544
4	3	項目名	64
4	3	項目値	516
5	3	項目名	192
5	3	項目値	514
7	3	項目名	192
7	3	項目値	514
9	3	項目名	64
9	3	項目値	516
10	3	項目名	64
10	3	項目値	516
1	6	項目名	8
1	6	項目値	804
2	6	項目名	1
2	6	項目値	768

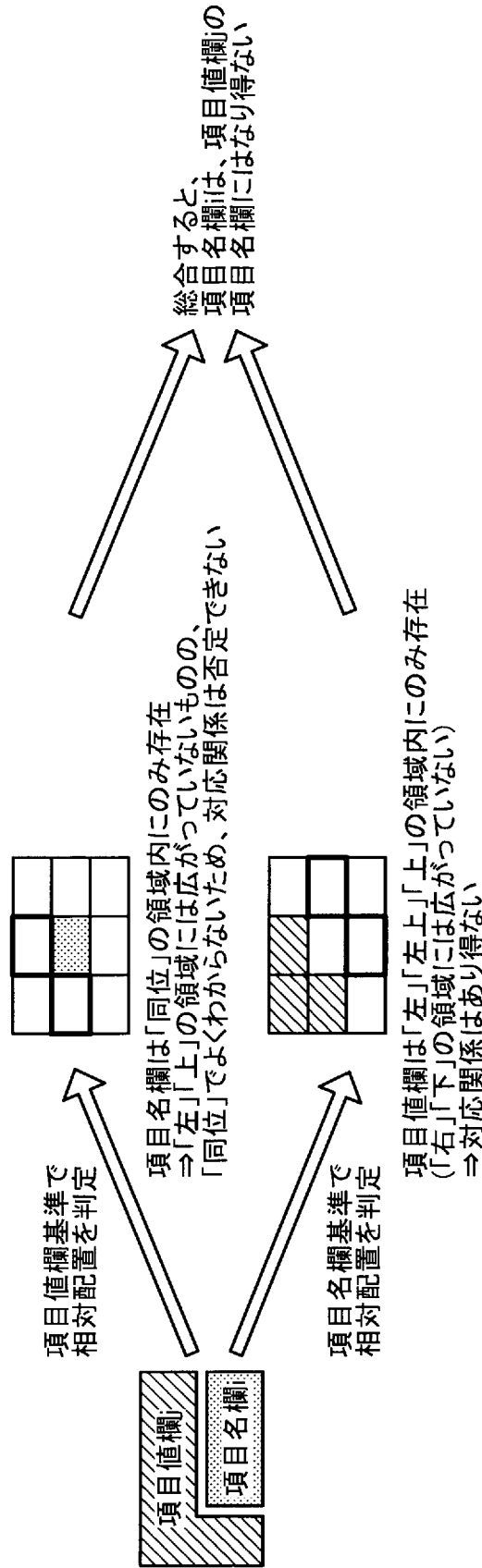
[図8]

ID	基準	相対配置パターン	許容可否
0	項目名	000000000	-
1	項目名	000000001	×
2	項目名	000000010	○
3	項目名	000000011	○
4	項目名	000000100	×
5	項目名	000000101	-
6	項目名	000000110	○
7	項目名	000000111	○
8	項目名	000001000	○
9	項目名	000001001	○
10	項目名	000001010	-
11	項目名	000001011	-
12	項目名	000001100	-
13	項目名	000001101	-
14	項目名	000001110	-
15	項目名	000001111	○
511	項目名	111111111	×
512	項目値	000000000	-
513	項目値	000000001	×
514	項目値	000000010	×
515	項目値	000000011	×
516	項目値	000000100	×
517	項目値	000000101	-
518	項目値	000000110	×
519	項目値	000000111	×
544	項目名	000100000	○
545	項目名	000100001	-
546	項目値	000100010	-
547	項目値	000100011	-
548	項目値	000100100	○
549	項目値	000100101	-
550	項目値	000100110	○
551	項目値	000100111	○

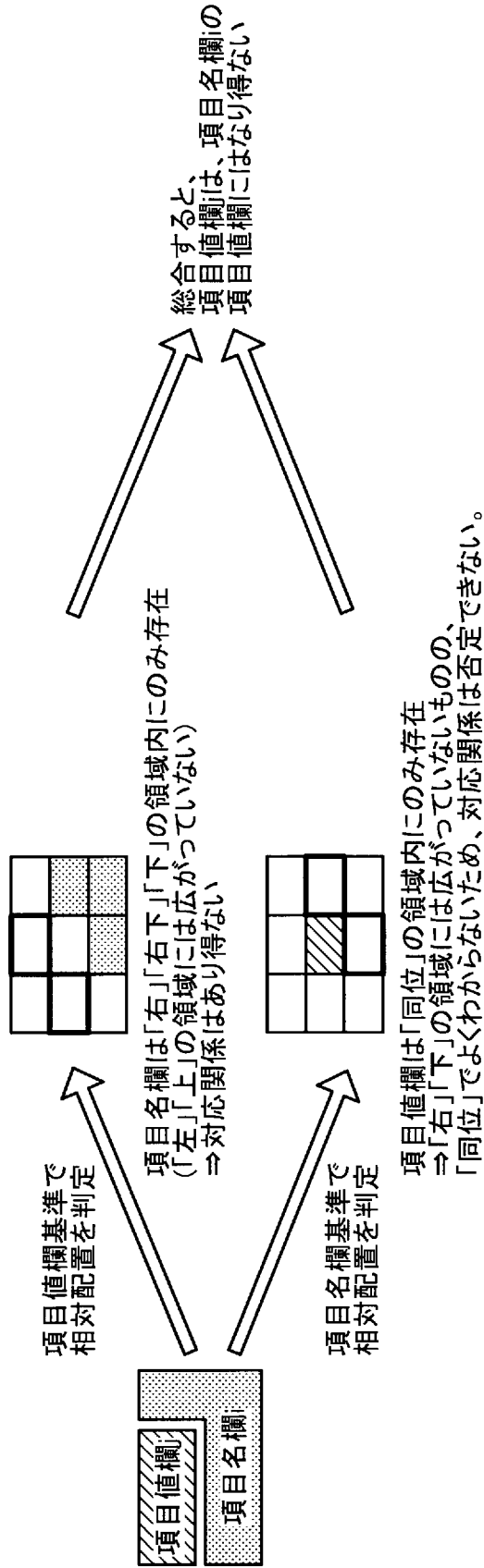
[図9]

項目名欄ID	項目値欄ID	許容可能
1	3	1
2	3	1
4	3	0
5	3	0
7	3	0
9	3	0
10	3	0
1	6	1
2	6	0

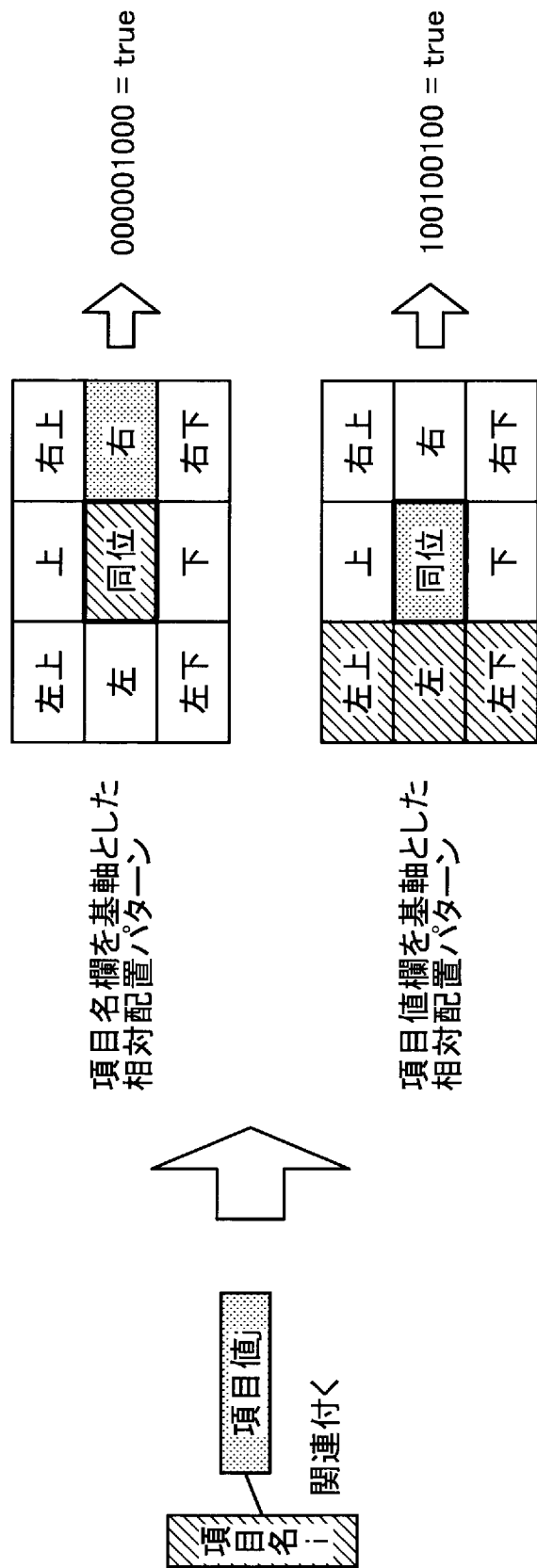
[図10]



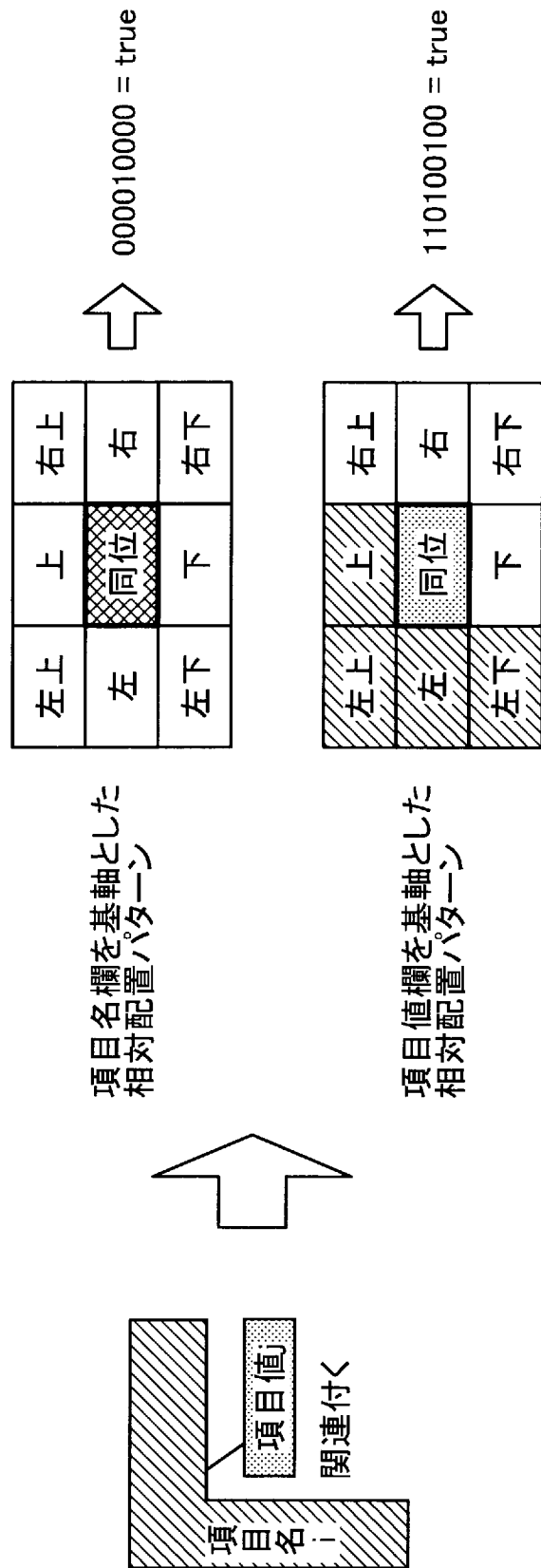
[図11]



[図12]



[図13]



[図14]

<集合と要素>

N_I : 項目名欄集合 $n = |N_I|$: 項目名欄集合の総数 v_i : 項目名欄(N_I の要素)
 N_V : 項目値欄集合 (厳密には、欄の種別が項目名と判定されなかった欄集合なので、項目名欄を含む可能性あり) v_j : 項目値欄(N_V の要素)
 $m = |N_V|$: 項目値欄集合の総数

S : 項目名欄に含まれる文字列(項目名)の集合 s : 項目名欄に含まれる文字列(S の要素)

R : 項目名セットの集合 r : 項目名セット(R, R^* の要素)

R^* : 処理対象として必須の項目の項目名セットの集合

K : 項目名欄セットの集合 k : 項目名欄セット(K, K_r の要素)

K_r : 項目名セット r に対する項目名欄セットの集合($K_r \subseteq K$)

<定数>

$$a_{ij} = \begin{cases} 1: \text{項目名欄}i\text{が項目値欄}j\text{と関連する可能性がある} \\ 0: \text{項目名欄}i\text{が項目値欄}j\text{と関連する可能性がない} \end{cases}$$

c_{ij} : 項目名欄 i と項目値欄 j が関連付くことに対する評価値

u_{si} : 項目名欄 i の文字列が文字列 s と同一かどうか(0/1)

$$p_{ki} = \begin{cases} 1: \text{項目名欄}i\text{が項目名欄セット}k\text{の要素である} \\ 0: \text{項目名欄}i\text{が項目名欄セット}k\text{の要素でない} \end{cases}$$

<変数>

$$x_{ij} = \begin{cases} 1: \text{項目名欄}i\text{が項目値欄}j\text{と関連づく} \\ 0: \text{項目名欄}i\text{が項目値欄}j\text{と関連づかない} \end{cases} \quad z_{kj} = \begin{cases} 1: \text{項目名欄セット}k\text{が項目値欄}j\text{と関連づく} \\ 0: \text{項目名欄セット}k\text{が項目値欄}j\text{と関連づかない} \end{cases}$$

[図15]

$$\begin{aligned}
 & \text{s. t.} \quad \sum_{v_i \in N_I} x_{ij} \geq 1 \quad (v_j \in N_V) \\
 & x_{ij} \leq a_{ij} \quad (v_i \in N_I, v_j \in N_V) \\
 & \sum_{v_i \in N_I} u_{si} x_{ij} \leq 1 \quad (s \in S, v_j \in N_V) \\
 & \sum_{s \in S} \sum_{v_i \in N_I} \{ u_{si} (x_{ij} - x_{il}) \}^2 > 0 \quad (v_j, v_l \in \{N_V \mid j \neq l\}) \\
 & x_{ij} \in \{0,1\}
 \end{aligned}$$

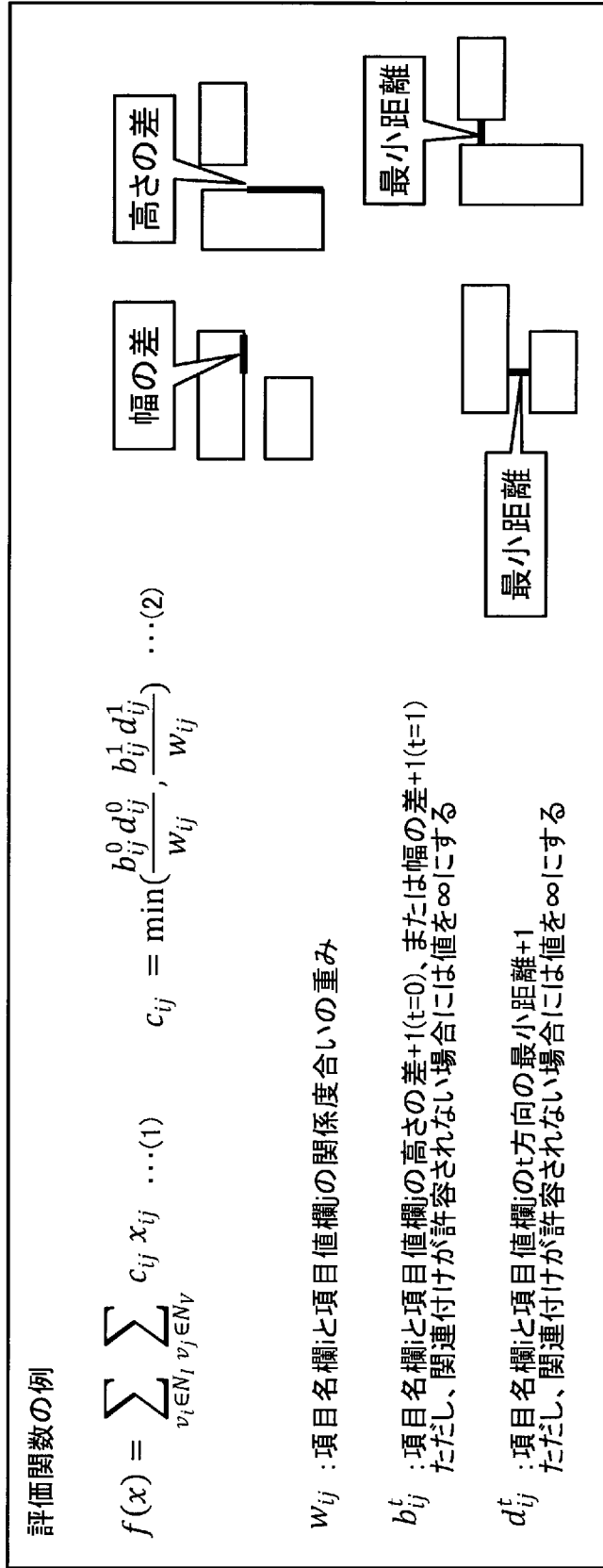
…(1) 各項目値欄は、少なくとも1個以上の項目名欄と関連付く

…(2) 項目名欄と項目値欄が関連付けられるのは、関連付けが許容される相対配置パターンである場合に限られる

…(3) 文字列が同じ項目名欄が複数ある場合、その中で同じ項目値欄と関連付くのは最大1個

…(4) 項目値欄同士は、それぞれ関連付く項目名欄に記載されている項目名のセットが互いに異なる

[図16]



[図17]

- $$\text{s. t. } \sum_{k \in K} z_{kj} \leq 1 \quad (v_j \in N_V)$$
 …(1) 各項目値欄は、最大1個の項目名欄セットと関連付く
- $$\sum_{k \in K_r} \sum_{v_j \in N_V} z_{kj} \leq 1 \quad (r \in R)$$
 …(2) 各項目名欄セットは最大1個の項目値欄と関連付く
 また、各項目名セットについて、同じ項目名セットから生成される項目名欄セットのうち、項目値欄に関連付くものは最大1個
- $$\sum_{k \in K_r} \sum_{v_j \in N_V} z_{kj} = 1 \quad (r \in R^*)$$
 …(3) 各項目名欄セットは最大1個の項目値欄と関連付く
 また、処理対象として必須の項目については、各項目名セットについて、同じ項目名セットから生成される項目名欄セットのうち、項目値欄に関連付くものはちょうど1個
- $$x_{ij} \leq a_{ij} \quad (v_i \in N_I, v_j \in N_V)$$
 …(4) 項目名欄と項目値欄が関連付けられるのは、
 関連付けが許容される相対配置パターンである場合に限られる
- $$x_{ij} = \sum_{k \in K} p_{ik} z_{kj} \quad (v_i \in N_I, v_j \in N_V)$$
 …(5)
- $$x_{ij} \in \{0,1\}$$
- $$z_{kj} \in \{0,1\}$$

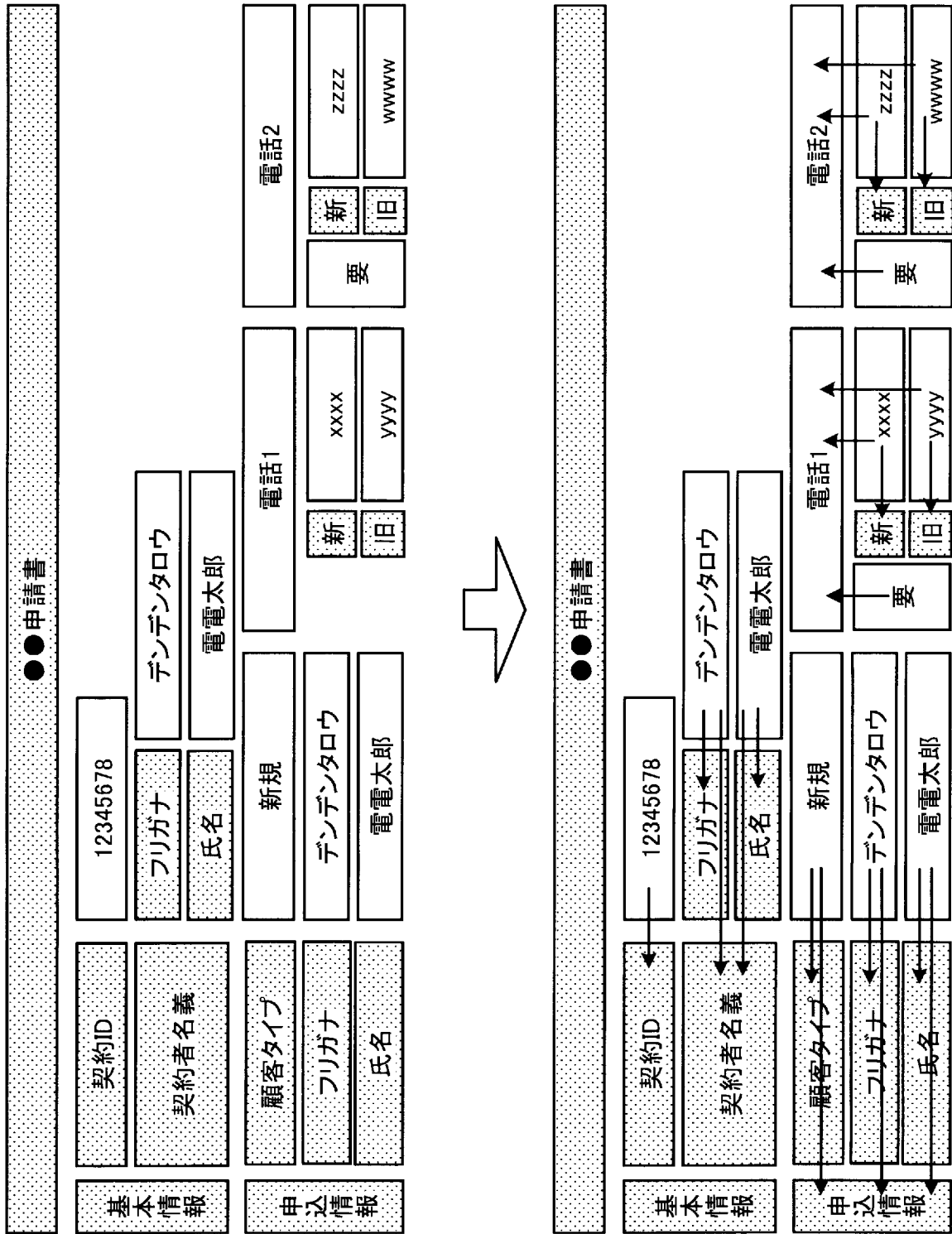
[図18]

$$f(x) = \sum_{v_i \in N_I} \sum_{v_j \in N_V} c_{ij} x_{ij} \quad \dots(1)$$

$$+ \sigma \sum_{r \in R} (1 - \sum_{k \in K_r} \sum_{v_j \in N_V} z_{kj}) \quad \dots(2)$$

σ : 重み(0以上)

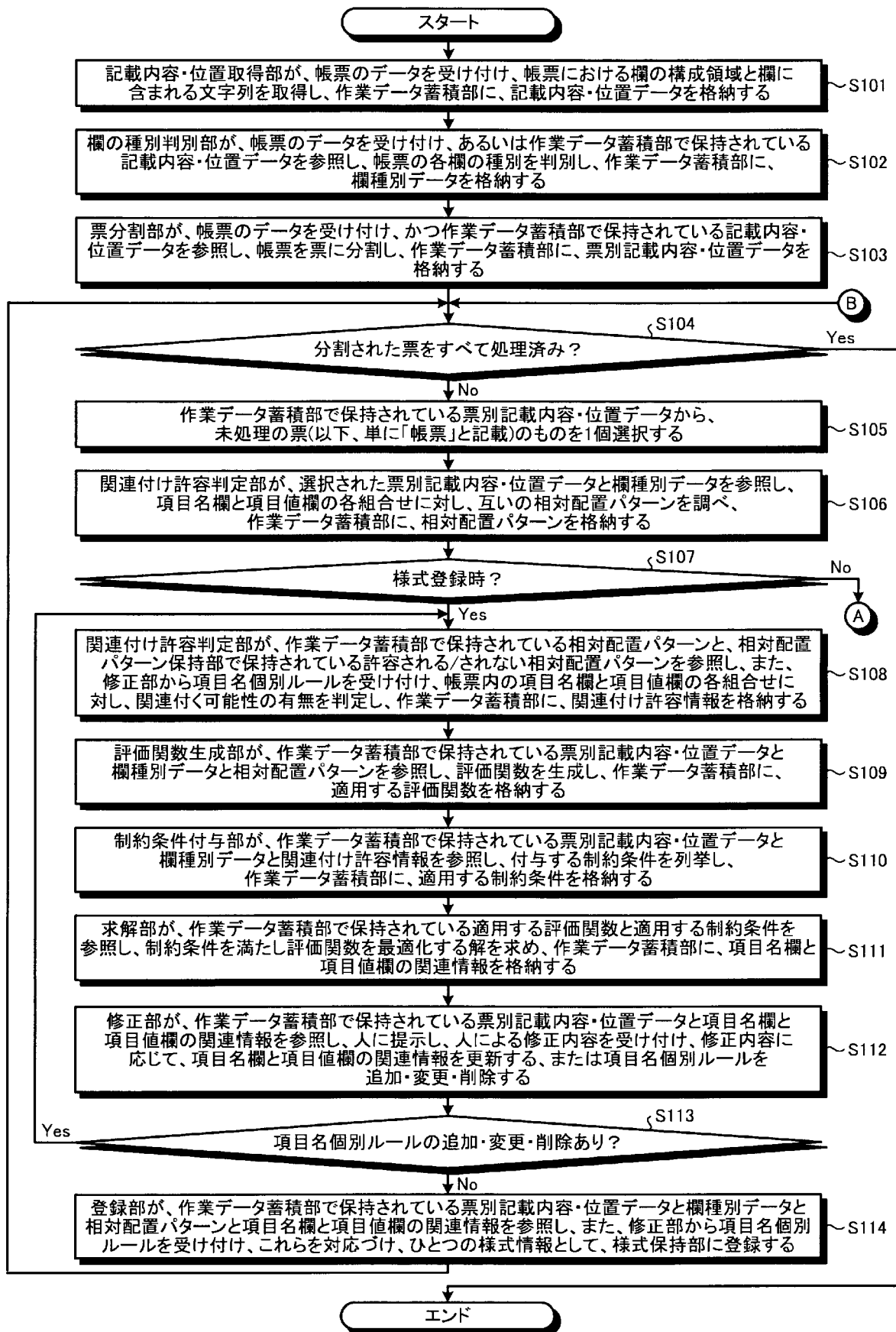
[図19]



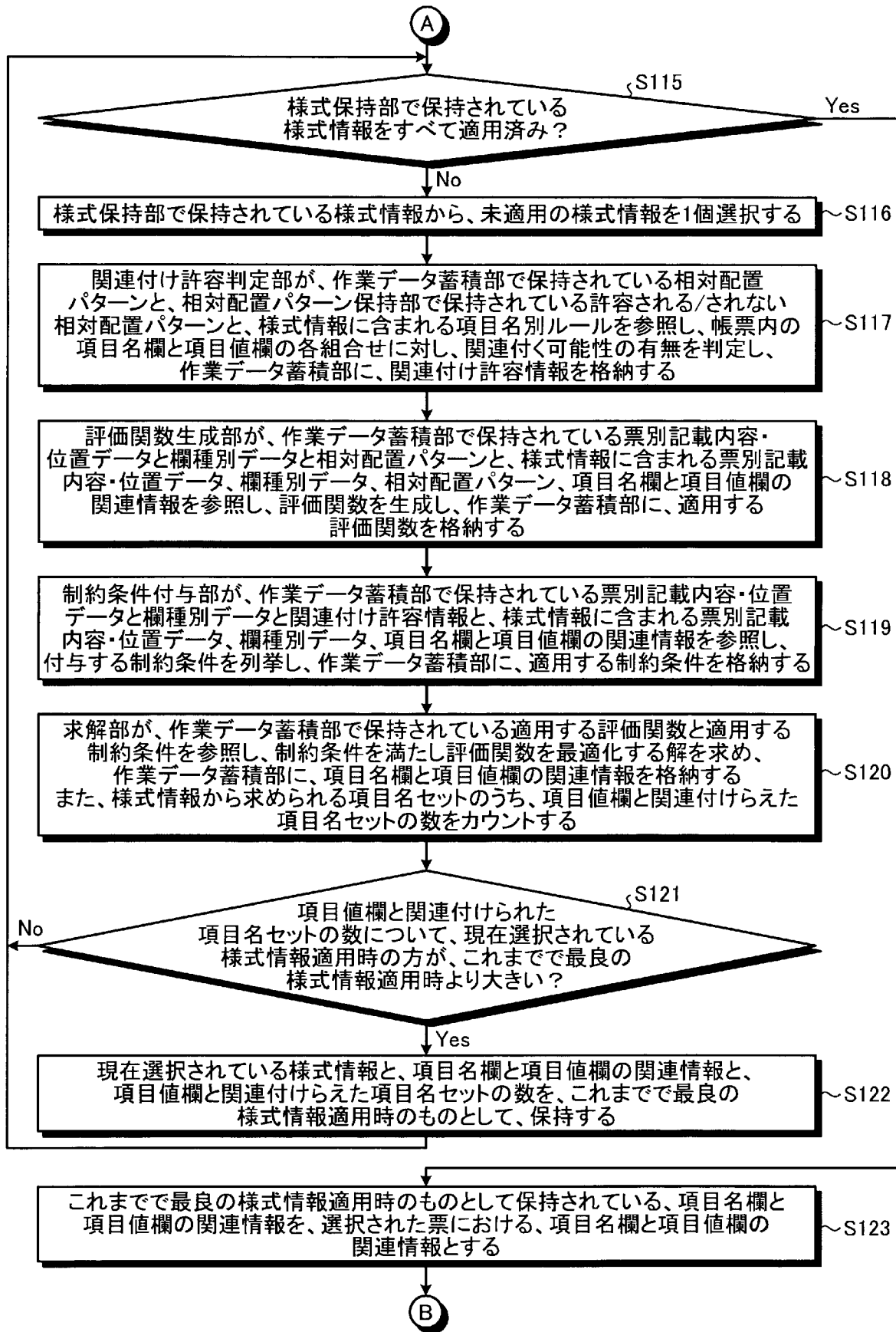
[図20]

項目値欄ID	項目名欄ID
3	2
6	4
6	5
8	4
8	7
11	10

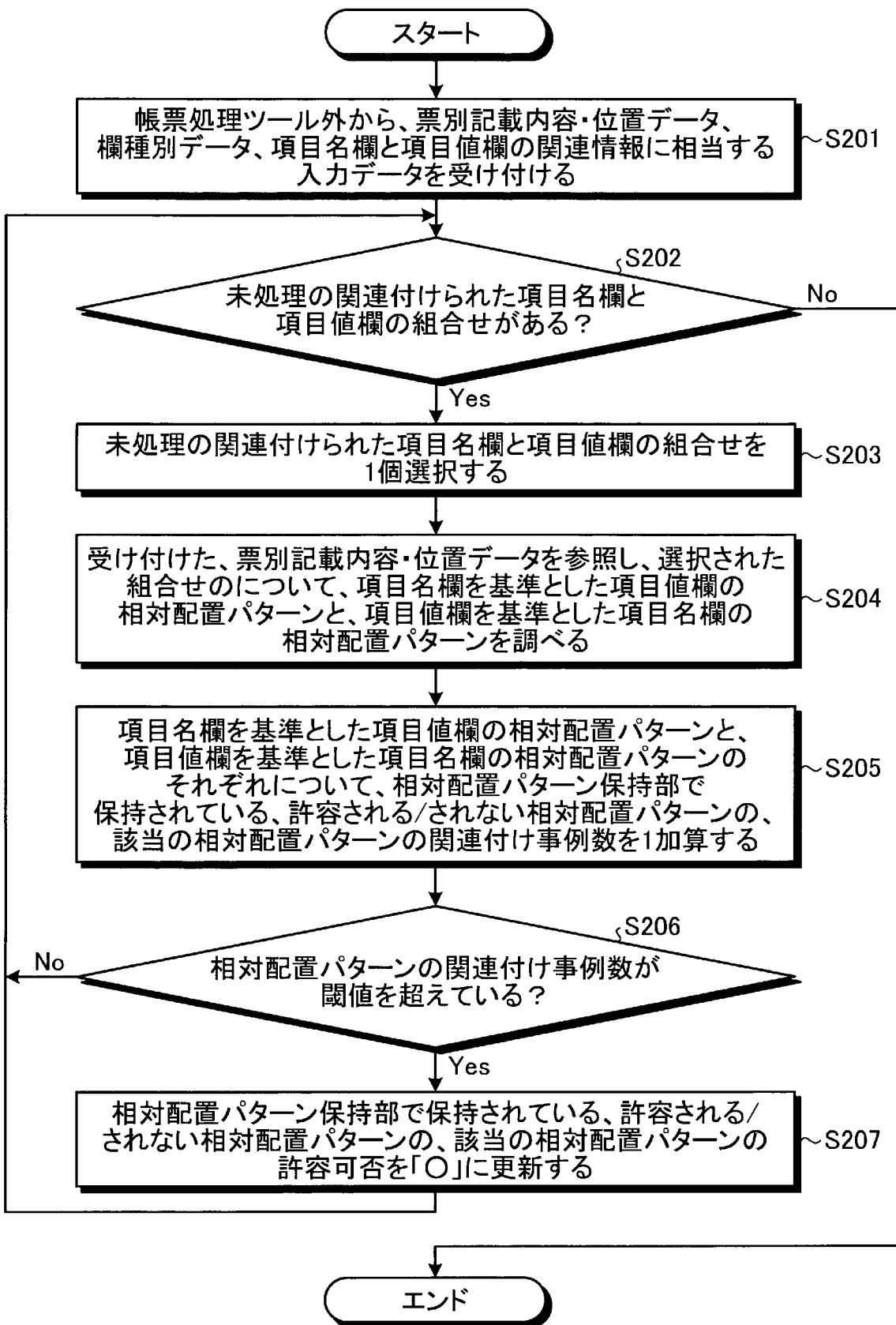
[図21A]



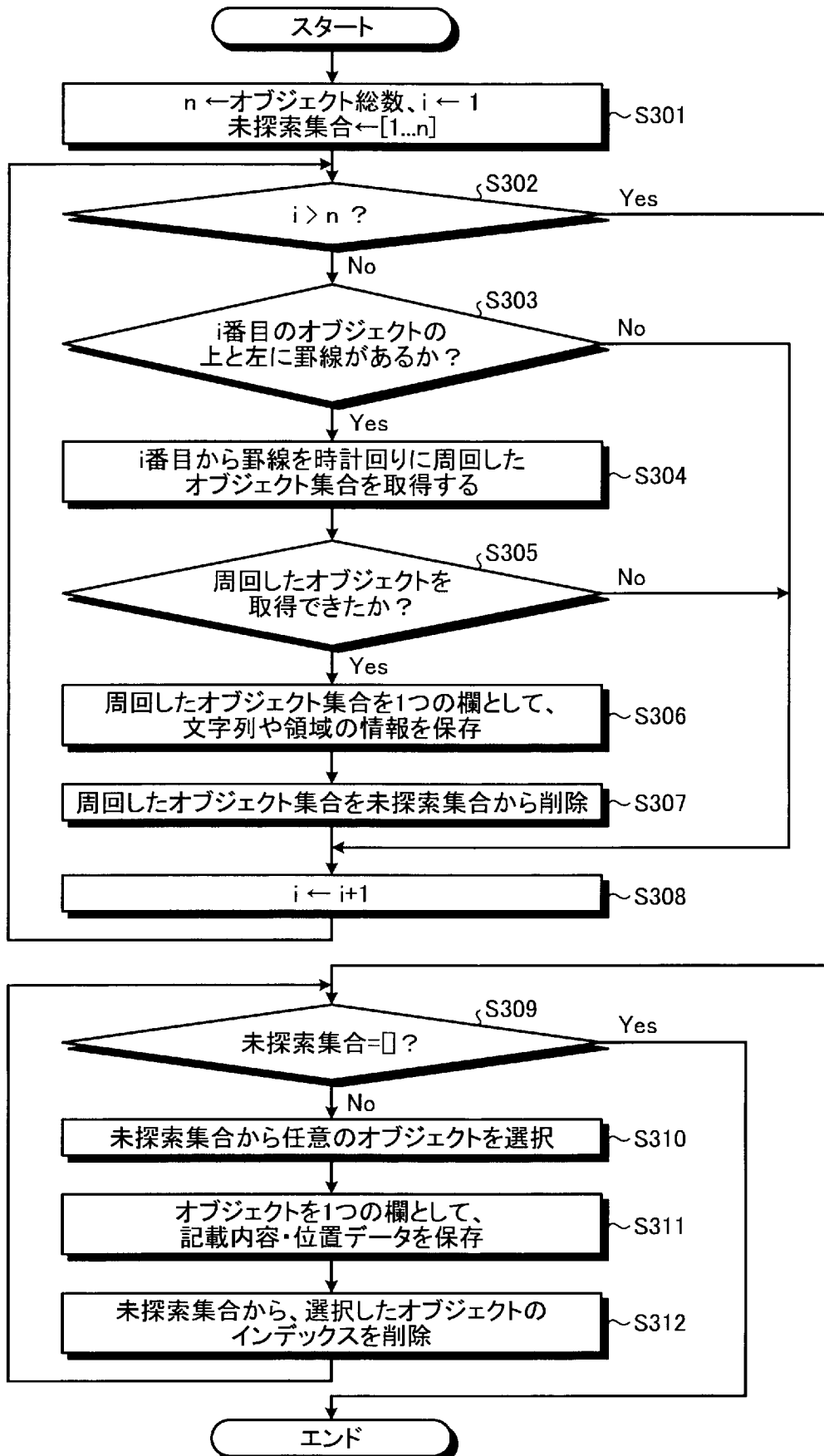
[図21B]



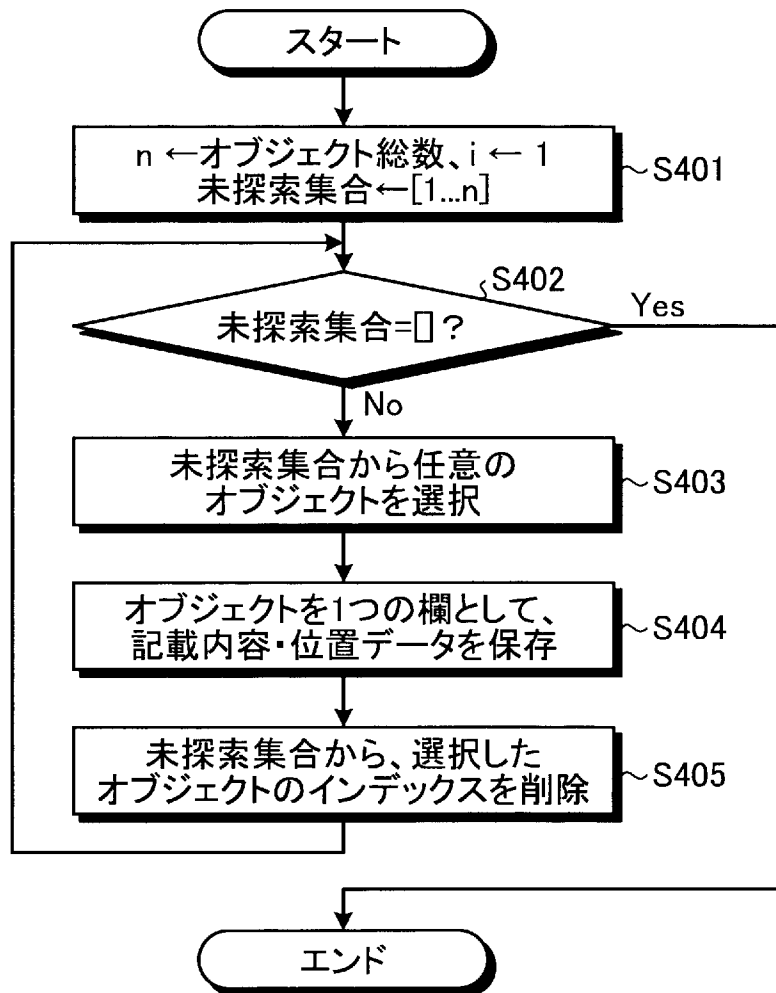
[図22]



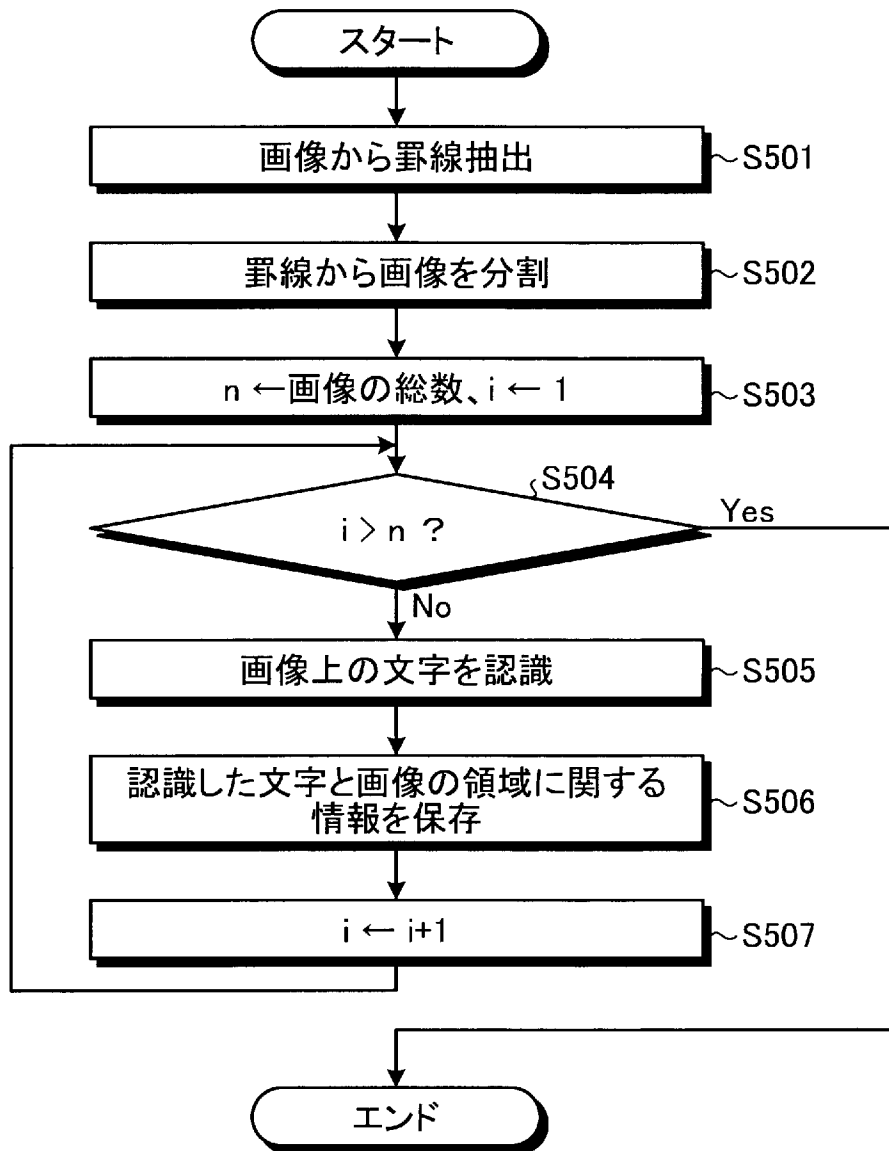
[図23]



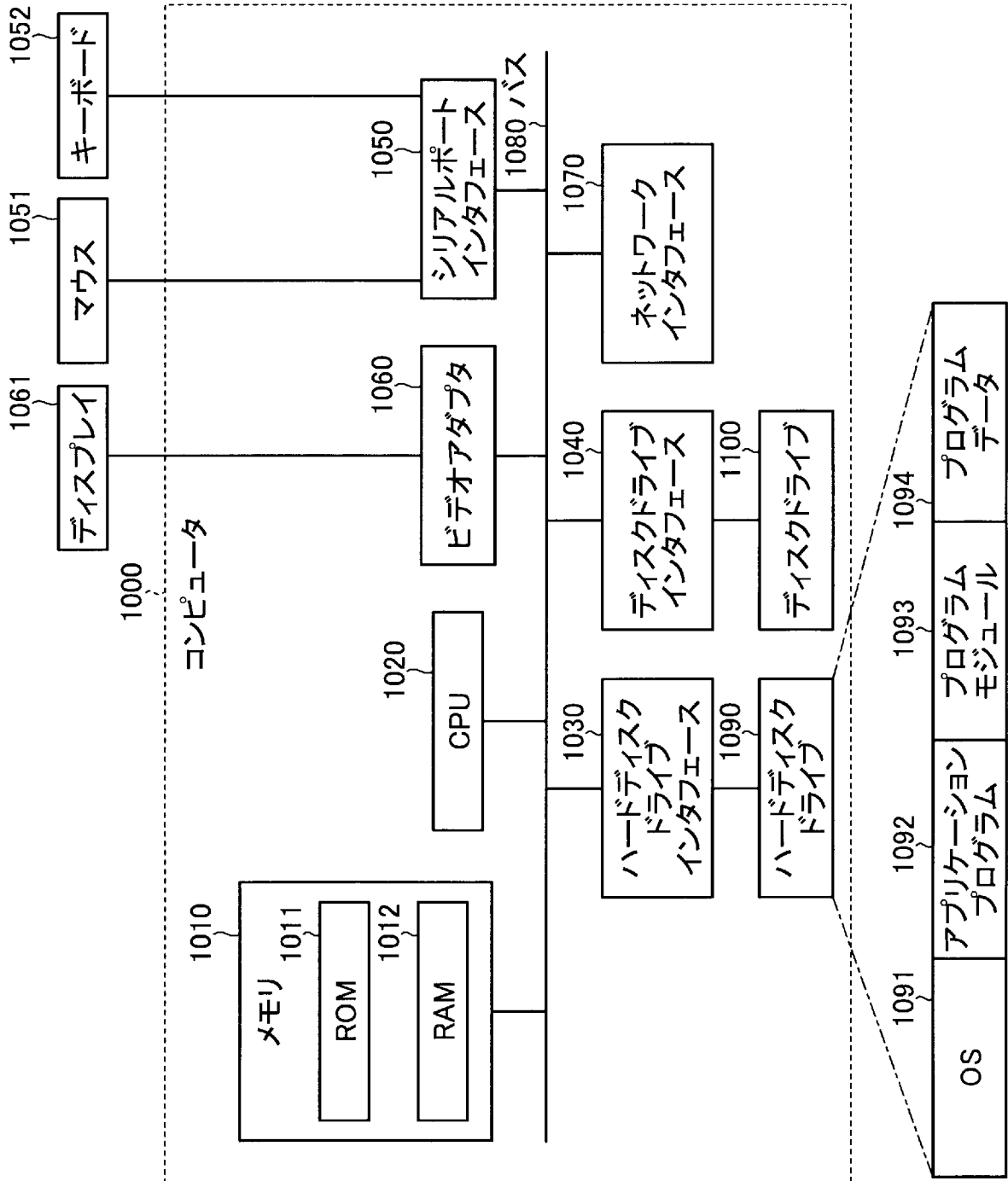
[図24]



[図25]



[図26]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2020/005631

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl. G06K9/20 (2006.01) i
FI: G06K9/20340C

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)

Int.Cl. G06K9/20

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

Published examined utility model applications of Japan	1922-1996
Published unexamined utility model applications of Japan	1971-2020
Registered utility model specifications of Japan	1996-2020
Published registered utility model applications of Japan	1994-2020

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 A	WO 2019/026147 A1 (PFU LTD.) 07.02.2019 (2019-02-07), claims 1-11, paragraphs [0040]-[0083], fig. 5, 6	1-2, 5-8 3-4
A	JP 2018-10489 A (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 18.01.2018 (2018-01-18), claims 1-8	1-8

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
15.04.2020

Date of mailing of the international search report
28.04.2020

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

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/JP2020/005631

WO 2019/026147 A1 07.02.2019 (Family: none)

JP 2018-10489 A 18.01.2018 (Family: none)

A. 発明の属する分野の分類（国際特許分類（IPC）） G06K 9/20(2006.01)i FI: G06K9/20 340C		
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）） G06K9/20 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 1971-2020年 日本国実用新案登録公報 1996-2020年 日本国登録実用新案公報 1994-2020年		
国際調査でを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
X A	WO 2019/026147 A1 (株式会社P F U) 07.02.2019 (2019-02-07) [請求項1] - [請求項11], 段落 [0040] - [0083], [図5] - [図6]	1-2, 5-8 3-4
A	JP 2018-10489 A (日本電信電話株式会社) 18.01.2018 (2018-01-18) [請求項1] - [請求項8]	1-8
<input type="checkbox"/> C欄の続きにも文献が列挙されている。 <input checked="" type="checkbox"/> 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抵触するものではなく、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 15.04.2020	国際調査報告の発送日 28.04.2020	
名称及びあて先 日本国特許庁(ISA/JP) 〒100-8915 日本国 東京都千代田区霞が関三丁目4番3号	権限のある職員（特許庁審査官） 新井 則和 5H 8937 電話番号 03-3581-1101 内線 3531	

国際調査報告
パテントファミリーに関する情報

国際出願番号

PCT/JP2020/005631

引用文献	公表日	パテントファミリー文献	公表日
WO 2019/026147 A1	07.02.2019	(ファミリーなし)	
JP 2018-10489 A	18.01.2018	(ファミリーなし)	