(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第4687089号
(P4687089)
(45)発行日　平成23年5月25日(2011.5.25)　(24)登録日　平成23年2月25日(2011.2.25)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 180D
G06F 17/30 350C
G06F 17/30 320D

請求項の数 14　(全 27 頁)

(21)出願番号　特願2004-355789(P2004-355789)
(22)出願日　平成16年12月8日(2004.12.8)
(65)公開番号　特開2006-163941(P2006-163941A)
(43)公開日　平成18年6月22日(2006.6.22)
審査請求日　平成19年11月12日(2007.11.12)

(73)特許権者　000004237
日本電気株式会社
東京都港区芝五丁目７番１号
(74)代理人　100103090
弁理士　岩壁　冬樹
(74)代理人　100124501
弁理士　塩川　誠人
(72)発明者　久寿居　大
東京都港区芝五丁目７番１号　日本電気株式会社内
(72)発明者　立石　健二
東京都港区芝五丁目７番１号　日本電気株式会社内

最終頁に続く

(54)【発明の名称】重複レコード検出システム、および重複レコード検出プログラム

(57)【特許請求の範囲】
【請求項１】
語の変換に用いられる辞書であって、当該語に対応する代表的な語である代表語を対応付けた辞書である代表語辞書と、相互に省略可能な前記代表語である省略可能語をグループ化した辞書である省略語辞書とを記憶する変換語記憶部と、
複数の情報からなる複数のレコードを保持するデータベースの各レコード間の表記の類似度を計算する類似度計算部と、
前記類似度計算部が計算した前記類似度が所定の値以上であるレコードの組み合わせである重複レコード候補を抽出する重複候補抽出部とを備え、
前記類似度計算部は、前記各レコードに含まれる語のうち、前記代表語辞書に含まれる語を対応する代表語に変換し、当該代表語に隣接する位置に前記省略語辞書において当該代表語と同一のグループに含まれる代表語を追加し、代表語が追加された各レコード間の表記の類似度を計算する
ことを特徴とする重複レコード検出システム。
【請求項２】
変換語記憶部は、同義語を代表語として記憶し、
類似度計算部は、データベースに登録されている各レコードに含まれる語を、対応する同義語に変換してレコード間の表記の類似度を計算する
請求項１記載の重複レコード検出システム。
【請求項３】

データベースのレコードを構成し、前記データベースのレコードに登録されている情報が区切られる単位であるフィールドの情報を入力するデータベース情報入力部を含み、
類似度計算部は、各レコード間の表記の類似度を前記フィールドごとに算出し、入力されたフィールドの情報に基づいて、フィールドごとに算出された類似度から、各レコード間の表記の類似度を計算する
請求項１または請求項２記載の重複レコード検出システム。
【請求項４】
重複候補抽出部が抽出した重複レコード候補が互いに同一の内容の情報のレコードの組み合わせである重複レコードであるか否かを類似度に応じて規定したルールである重複判定ルールを記憶する重複判定ルール記憶部と、
前記重複判定ルール記憶部が記憶している前記重複判定ルールに規定された類似度と、各レコード間の類似度または各レコード間のフィールドごとの類似度との関係にもとづいて、前記重複レコード候補が前記重複レコードであるか否かを判定する重複レコード判定部とを含む
請求項１から請求項３のうちいずれか１項に記載の重複レコード検出システム。
【請求項５】
重複レコード判定部が重複レコードであると判定したレコード間で、異なる部分から導出される語の組を代表語候補として抽出し、抽出した代表語候補を代表語辞書に含めて変換語記憶部に記憶させる代表語候補抽出部を含む
請求項４記載の重複レコード検出システム。
【請求項６】
代表語候補抽出部は、重複レコードと判定された２つのレコードのうち、一のレコードの文字列が他のレコードの文字列に含まれる場合、２つのレコードで異なる部分の文字列と、共通する部分の文字列との組を、省略可能語候補として抽出し、抽出した省略可能語候補を省略語辞書に含めて変換語記憶部に記憶させる
請求項５記載の重複レコード検出システム。
【請求項７】
代表語候補抽出部は、データベース内のレコードのうちのいずれかのレコードにおいて、抽出された代表語候補に含まれる語を全て含むレコードが存在する場合、当該代表語候補を、省略可能語候補とし、当該省略可能語候補を省略語辞書に含めて変換語記憶部に記憶させる
請求項５記載の重複レコード検出システム。
【請求項８】
代表語候補抽出部は、抽出した代表語候補のうち、当該代表語候補に含まれる語を組み合わせた文字列が、他の代表語候補に含まれる語と一致する場合、当該代表語候補を、変換語記憶部に記憶させる対象の代表語候補から除外する
請求項７記載の重複レコード検出システム。
【請求項９】
代表語候補抽出部は、抽出された省略可能語候補がいずれかの代表語候補に含まれる組の語をいずれも含んでいる場合、当該省略可能語候補を、変換語記憶部に記憶させる対象の代表語から除外する
請求項８記載の重複レコード検出システム。
【請求項１０】
重複レコード候補を出力し、使用者が、重複レコード候補が重複レコードであるか否かの判定を入力する入出力部を含み、
重複レコード判定部は、重複候補抽出部が抽出した重複レコード候補のうち、重複判定ルールにより重複レコードでないと判定された重複レコード候補を、当該重複レコード候補に含まれる語の数の最も多い重複レコード候補から順に前記入出力部に出力する
請求項４から請求項９のうちいずれか１項記載の重複レコード検出システム。
【請求項１１】

重複レコード判定部は、重複判定ルールにより重複レコードでないと判定された重複レコード候補のうち、重複すると判定されるレコードの組合せの数が多い順に、前記重複レコード候補を入出力部に出力する

請求項１０記載の重複レコード検出システム。

【請求項１２】

重複レコードであると重複レコード判定部、または入出力部を介して使用者に判定されたレコードの組み合わせを記憶する重複レコードデータベースと、

前記重複レコードデータベースが記憶しているレコードの組み合わせを構成するレコードのうち、一のレコード以外のレコードをデータベースから削除する重複レコード削除部とを含む

請求項１０または請求項１１記載の重複レコード検出システム。

【請求項１３】

データベースに登録すべく使用者が入力した情報の語を、変換語記憶部が記憶している語に変換して、変換した語、または入力された前記情報の語と合致する語からなる情報を含むレコードをデータベースから抽出するデータベース登録部と、

前記データベース登録部が抽出したレコードを表示する表示部とを含む

請求項１から請求項１２のうちいずれか１項記載の重複レコード検出システム。

【請求項１４】

語の変換に用いられる辞書であって、当該語に対応する代表的な語である代表語を対応付けた辞書である代表語辞書と、相互に省略可能な前記代表語である省略可能語をグループ化した辞書である省略語辞書とを記憶する変換語記憶部を備えたコンピュータに適用される重複レコード検出プログラムであって、

前記コンピュータに、

複数の情報からなる複数のレコードを保持するデータベースの各レコード間の表記の類似度を計算する類似度計算処理と、

前記類似度計算処理で計算された前記類似度が、所定の値以上であるレコードの組み合わせである重複レコード候補を抽出する重複候補抽出処理とを実行させ、

前記類似度計算処理で、前記各レコードに含まれる語のうち、前記代表語辞書に含まれる語を対応する代表語に変換させ、当該代表語に隣接する位置に前記省略語辞書において当該代表語と同一のグループに含まれる代表語を追加させ、代表語が追加された各レコード間の表記の類似度を計算させる

ための重複レコード検出プログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、店舗等の情報が登録されたデータベースに重複して登録されている情報を検出する重複レコード検出システム、および重複レコード検出プログラムに関する。

【背景技術】

【０００２】

店舗や、人物、書物等の情報によって構成されるデータベースに、重複する情報が登録されている場合がある。

【０００３】

具体的には、例えば、同じ情報が異なる書式でデータベースに登録されていたり、同義であるが異なる語によってデータベースに登録されていたりする。同じ情報が重複してデータベースに登録されていると、データベースの容量が大きくなってしまったりするという問題がある。

【０００４】

データベースの容量を削減するために、特許文献１には、多数の人物の情報が登録されているデータベースから、同一人物の情報の重複登録を検出するシステムが記載されている。

【０００５】
また、特許文献２には、書物の情報が登録されているデータベースから一の書物を検索対象として抽出する際に、異なる書式によって同一の書物が複数重複してデータベースに登録されていても、検索対象となる書物をすべて抽出する装置が記載されている。
【０００６】
【特許文献１】特開平１１－１８４８８４号公報　（段落００１７～００４９、図１）
【特許文献２】特開２００４－２９９６９号公報　（段落００２２～００７５、図２）
【発明の開示】
【発明が解決しようとする課題】
【０００７】
特許文献１に記載されているシステムは、例えば、カナ氏名、漢字氏名、カナ住所、漢字住所、生年月日などの書式を、統一した表記法による書式に正規化することによって、同一人物の情報の重複登録を検出する。
【０００８】
しかし、特許文献１に記載されているシステムは、同義であるが異なる語によってデータベースに登録されている情報の重複登録を検出することができないという問題がある。具体的には、例えば、同一人物の生年月日が西暦と和暦とで重複して登録されていると、重複登録を検出することができない。また、同一人物の住所の情報が、「東京都千代田区・・」という表記と、「都内千代田区・・」という表記とで重複して登録されていると、重複登録を検出することができない。
【０００９】
また、特許文献２に記載されている装置は、入力された検索対象の書物の情報と、データベースに登録されている書物の情報との類似度を算出して類似度の高い書物を検索結果として抽出するが、同義であるが異なる語によってデータベースに登録されている情報の類似度は低くなるため、そのような書物は抽出されにくいという問題がある。
【００１０】
具体的には、例えば、検索対象の書物の名称として「にほん」と入力された場合、「にっぽん」という名称の書物の類似度は低く算出されるため、「にっぽん」という名称の書物は抽出されにくくなってしまう。
【００１１】
そこで、本発明は、同義語や省略可能語による表記の差異があっても、重複する情報をデータベースから検出する重複レコード検出システム、および重複レコード検出プログラムを提供することを目的とする。
【課題を解決するための手段】
【００１２】
本発明による重複レコード検出システムは、語の変換に用いられる辞書であって、その語に対応する代表的な語である代表語を対応付けた辞書である代表語辞書と、相互に省略可能な代表語である省略可能語をグループ化した辞書である省略語辞書とを記憶する変換語記憶部と、複数の情報からなる複数のレコードを保持するデータベースの各レコード間の表記の類似度を計算する類似度計算部と、類似度計算部が計算した類似度が所定の値以上であるレコードの組み合わせである重複レコード候補を抽出する重複候補抽出部とを備え、類似度計算部が、各レコードに含まれる語のうち、代表語辞書に含まれる語を対応する代表語に変換し、当該代表語に隣接する位置に省略語辞書において当該代表語と同一のグループに含まれる代表語を追加し、代表語が追加された各レコード間の表記の類似度を計算することを特徴とする。
【００１３】
変換語記憶部は、同義語を代表語として記憶してもよく、類似度計算部は、データベースに登録されている各レコードに含まれる語を、対応する同義語に変換してレコード間の表記の類似度を計算してもよい。
【００１５】

データベースのレコードを構成し、データベースのレコードに登録されている情報が区切られる単位であるフィールドの情報を入力するデータベース情報入力部を含んでもよく、類似度計算部は、各レコード間の表記の類似度をフィールドごとに算出し、入力されたフィールドの情報に基づいて、フィールドごとに算出された類似度から、各レコード間の表記の類似度を計算してもよい。そのような構成によれば、フィールド間の類似度を用いて、レコード間の類似度を計算することができる。

【００１６】

重複候補抽出部が抽出した重複レコード候補が、互いに同一の内容の情報のレコードの組み合わせである重複レコードであるか否かを類似度に応じて規定したルールである重複判定ルールを記憶する重複判定ルール記憶部と、重複判定ルール記憶部が記憶している重複判定ルールに規定された類似度と、各レコード間の類似度または各レコード間のフィールドごとの類似度との関係にもとづいて、重複レコード候補が重複レコードであるか否かを判定する重複レコード判定部とを含んでもよい。そのような構成によれば、重複レコード候補が重複レコードであるか否かを、自動的に判定することができる。

【００１７】

重複レコード判定部が重複レコードであると判定したレコード間で、異なる部分から導出される語の組を代表語候補として抽出し、抽出した代表語候補を代表語辞書に含めて変換語記憶部に記憶させる代表語候補抽出部を含んでもよい。そのような構成によれば、重複レコードから、変換語候補を抽出することができる。

【００１８】

代表語候補抽出部は、重複レコードと判定された２つのレコードのうち、一のレコードの文字列が他のレコードの文字列に含まれる場合、２つのレコードで異なる部分の文字列と、共通する部分の文字列との組を、省略可能語候補として抽出し、抽出した省略可能語候補を省略語辞書に含めて変換語記憶部に記憶させてもよい。

【００１９】

代表語候補抽出部は、データベース内のレコードのうちのいずれかのレコードにおいて、抽出された代表語候補に含まれる語を全て含むレコードが存在する場合、当該代表語候補を、省略可能語候補とし、当該省略可能語候補を省略語辞書に含めて変換語記憶部に記憶させてもよい。また、代表語候補抽出部は、抽出した代表語候補のうち、当該代表語候補に含まれる語を組み合わせた文字列が、他の代表語候補に含まれる語と一致する場合、当該代表語候補を、変換語記憶部に記憶させる対象の代表語候補から除外してもよい。また、代表語候補抽出部は、抽出された省略可能語候補がいずれかの代表語候補に含まれる組の語をいずれも含んでいる場合、当該省略可能語候補を、変換語記憶部に記憶させる対象の代表語から除外してもよい。

【００２１】

重複レコード候補を出力し、使用者が、重複レコード候補が重複レコードであるか否かの判定を入力する入出力部を含んでもよく、重複レコード判定部は、重複候補抽出部が抽出した重複レコード候補のうち、重複判定ルールにより重複レコードでないと判定された重複レコード候補を、当該重複レコード候補に含まれる語の数の最も多い重複レコード候補から順に前記入出力部に出力してもよい。そのような構成によれば、使用者が、入出力部を介して判定を入力する回数を減らすことができる。

【００２２】

重複レコード判定部は、重複判定ルールにより重複レコードでないと判定された重複レコード候補のうち、重複すると判定されるレコードの組合せの数が多い順に、重複レコード候補を入出力部に出力してもよい。

【００２４】

重複レコードであると重複レコード判定部、または入出力部を介して使用者に判定されたレコードの組み合わせを記憶する重複レコードデータベースと、重複レコードデータベースが記憶しているレコードの組み合わせを構成するレコードのうち、一のレコード以外のレコードをデータベースから削除する重複レコード削除部とを含んでもよい。そのよう

な構成によれば、重複レコードをデータベースから削除することができる。

【００２５】

データベースに登録すべく使用者が入力した情報の語を、変換語記憶部が記憶している語に変換して、変換した語、または入力された情報の語と合致する語からなる情報を含むレコードをデータベースから抽出するデータベース登録部と、データベース登録部が抽出したレコードを表示する表示部とを含んでもよい。そのような構成によれば、重複する情報のデータベースへの登録を防ぐことができる。

【００２６】

本発明による重複レコード検出プログラムは、語の変換に用いられる辞書であって、当該語に対応する代表的な語である代表語を対応付けた辞書である代表語辞書と、相互に省略可能な前記代表語である省略可能語をグループ化した辞書である省略語辞書とを記憶する変換語記憶部を備えたコンピュータに適用される重複レコード検出プログラムであって、コンピュータに、複数の情報からなる複数のレコードを保持するデータベースの各レコード間の表記の類似度を計算する類似度計算処理と、類似度計算処理で計算した類似度が、所定の値以上であるレコードの組み合わせである重複レコード候補を抽出する重複候補抽出処理とを実行させ、類似度計算処理で、各レコードに含まれる語のうち、代表語辞書に含まれる語を対応する代表語に変換させ、その代表語に隣接する位置に省略語辞書においてその代表語と同一のグループに含まれる代表語を追加させ、代表語が追加された各レコード間の表記の類似度を計算させることを特徴とする。

【発明の効果】

【００２７】

本発明によれば、同義語や省略可能語による表記の差異があっても、重複する情報をデータベースから検出することができる。

【発明を実施するための最良の形態】

【００２８】

実施の形態１．

本発明の第１の実施の形態について、図面を参照して説明する。図１は、本発明の第１の実施の形態の一構成例を説明するブロック図である。

【００２９】

本発明の第１の実施の形態による重複レコード検出システム２０は、同義語が登録されている同義語辞書と、省略可能な語が登録されている省略可能語辞書とによって構成される変換語辞書（変換語記憶部）５、変換語辞書５を用いて、店舗の情報が登録されているデータベース２に登録されている複数のレコード間の類似度を計算する類似度計算部３、類似度計算部３が計算した類似度が所定の閾値以上であった情報を抽出する重複候補抽出部６、および各部の動作を制御するプログラムを記憶する記憶部１を含む。

【００３０】

重複レコード検出システム２０は、プログラムによって処理を実行するサーバ等のコンピュータによって実現される。なお、重複レコード検出システム２０は、外部の記憶媒体が記憶しているプログラムに従って処理を実行してもよい。また、変換語辞書５は、予め同義語および省略可能語が登録されているものとする。

【００３１】

図２は、データベース２に登録されている情報の例を示す説明図である。データベース２には、例えば、店舗の名称や住所、電話番号が登録されている。なお、データベース２には、各店舗の情報が、レコードに区切られて登録され、店舗の各情報は、登録されているレコードの各情報の属性に応じたフィールドに区切られて登録されているものとする。具体的には、図２に示す例では、レコードＩＤが「００１」のレコードには「エヌイーシー奈良支店」の各情報が登録され、レコードＩＤが「００２」のレコードには「日電奈良支店」の各情報が登録され、レコードＩＤが「００３」のレコードには「ＮＥＣ奈良支店」の各情報が登録されている。

【００３２】

また、データベース２の各レコードの名称のフィールドには、「エヌイーシー奈良支店」、「日電奈良支店」、および「ＮＥＣ奈良支店」が登録されており、住所のフィールドには、「○○○１の１」、「○○○１－１」、および「○○○１－１」が登録されており、電話番号のフィールドには、「０００－１１１－１２３４」、「０００－１１１－１２３５」、および「０００－１１１－１２３４」が登録されているものとする。

【００３３】

なお、データベース２における各レコードのフィールドの数等の情報を入力し、入力されたデータベース２の情報を類似度計算部３に出力するデータベース情報入力部４を含んでもよい。データベース情報入力部４には、使用者がキーボード等の入力手段を用いてデータベース２の情報を入力してもよい。また、データベース情報入力部４は、記憶部１や外部の記憶媒体が記憶しているデータベース２の情報を読み込んでもよい。

【００３４】

データベース情報入力部４には、例えば、どのフィールドは何を表しているのか（例えば、各レコードの先頭のフィールドはＩＤである等）、どのフィールドとどのフィールドとを結合して１つのフィールドとして扱う（例えば、住所が「都道府県」、「市町村」、および「番地とビル名」に分かれている各フィールドを１つのフィールドとして扱う等）のか、どのフィールドの類似度計算に変換語辞書５を用いるのか、およびレコード間の類似度を算出する際の各フィールドの重み（名称フィールド、住所フィールド、および電話番号フィールドの重みの比を、１：１：１とする）等の情報を入力する。

【００３５】

図３は、変換語辞書５に登録されている情報の例を示す説明図である。変換語辞書５を構成する同義語辞書には、語と、その語の同義語のうち代表的な語である代表語とが対応づけられて登録されている。また、変換語辞書５を構成する省略可能語辞書には、代表語のうち、相互に省略可能な代表語に同じグループＩＤが付されて登録されている。

【００３６】

図３の例によれば、「エヌイーシー」と「日本電気」との代表語は「ＮＥＣ」であり、「日本電気株式会社」の代表語は「日電」である。また、「ＮＥＣ」と「日電」とは相互に省略可能な省略可能語である。

【００３７】

なお、同義語辞書において、同義語の欄の語は、代表語として用いられることはないものとする。また、省略可能語辞書において、省略可能語として登録されている語は、代表語であってもよいが、同義語ではないものとする。

【００３８】

類似度計算部３は、例えば、形態素解析等の方法を用いて、データベース２に登録されている各情報を語の単位に分解する。なお、情報を語の単位に分解する他の方法として、例えば、スペースの前後で語の単位に分解したり、文字種が切り替わる位置（例えば、カタカナから漢字に切り替わる位置等）で語の単位に分解したりする方法がある。類似度計算部３は、同義語辞書を検索して、分解した語が同義語辞書に同義語として登録されていると、その同義語に対応づけられている代表語に変換する。

【００３９】

類似度計算部３は、省略可能語辞書を検索して、代表語に変換された語が省略可能語として登録されていると、同じグループＩＤが付されている省略可能語を、データベース２に登録されているレコードの語に追加する。そして、類似度計算部３は、代表語に変換され、省略可能語が追加された各レコード間の類似度を計算する。

【００４０】

類似度計算部３が各レコード間の類似度を計算する方法は、例えば、各情報の対応するフィールドの語の文字を先頭から１文字ずつ比較していき、合致すればその文字の類似度を１とし、合致しなければその文字の類似度を０とする。そして、例えば、各文字の類似度を合計した数を、語の文字数で割った商（すなわち、０から１の間で正規化した値）をそのフィールドの類似度とする。なお、各情報のフィールドの語の文字数が異なっている

場合は、各フィールドを構成する語のうち最も多い文字数で、各文字の類似度を合計した数を割った商をそのフィールドの類似度とする。

【００４１】

そして、類似度計算部３は、各フィールドの類似度に、各フィールドごとの所定の重みの値を乗じた積を合計した数を、フィールドの数で割った商（すなわち、０から１の間で正規化した値）を、レコードの類似度として計算する。

【００４２】

本発明の類似度の計算方法は、上述した方法に限定されるものではなく、編集距離を用いる方法等の、他の方法を用いてもよい。なお、類似度計算部３は、データベース情報入力部４に入力された情報にもとづいて、各フィールドおよび各レコードの類似度を計算してもよい。

【００４３】

重複候補抽出部６は、類似度計算部３が計算した類似度が、所定の閾値以上であるレコードの組を、重複レコード候補として抽出する。なお、重複候補抽出部６は、他の方法を用いて、重複レコード候補を抽出してもよい。

【００４４】

重複レコード検出システム２０は、コンピュータに、複数の情報からなる複数のレコードを保持するデータベースに登録されている情報に用いられている語を、語に対応する変換語を記憶する変換語辞書５が記憶している変換語に変換して、レコード間の類似度を計算させる類似度計算処理と、類似度計算処理で計算した類似度が、所定の値以上であるレコードの組み合わせである重複レコード候補を抽出させる重複候補抽出処理とを実行させるための重複レコード検出プログラムを搭載している。

【００４５】

次に、本発明の第１の実施の形態の動作を、具体例を挙げて図面を参照して説明する。図２の例に示すデータベース２に登録されている情報から、重複レコード候補を抽出する。図４は、本発明の第１の実施の形態の動作を説明するフローチャートである。

【００４６】

まず、類似度計算部３が、データベース２に登録されている情報を読み込む（ステップＳ１０１）。レコードＩＤが「００１」のレコードの名称のフィールドは、「エヌイーシー奈良支店」である。類似度計算部３は、「エヌイーシー奈良支店」に形態素解析等を行って語に分解する（ステップＳ１０２）。具体的には、「エヌイーシー」と「奈良」と「支店」とに分解する。

【００４７】

なお、ここでは、データベース情報入力部４に入力された情報が、名称フィールド、および住所フィールドの類似度計算に、変換語辞書５を用いることを示していたものとする。すると、類似度計算部３は、変換語辞書５を参照して、分解した語が同義語であれば代表語に変換する（ステップＳ１０３）。図３を参照すると、「エヌイーシー」が同義語であるので、「エヌイーシー」を代表語である「ＮＥＣ」に変換する。「奈良」および「支店」は同義語辞書に登録されていないので変換を行わない。すると、レコードＩＤ「００１」の語は、「ＮＥＣ」、「奈良」、および「支店」である。

【００４８】

次に、類似度計算部３は、変換語辞書５を参照して、省略可能語があれば、同じグループＩＤの省略可能語を追加する（ステップＳ１０４）。図３を参照すると、「ＮＥＣ」と「日電」とが同じグループＩＤの省略可能語であるので、「日電」を追加する。すると、レコードＩＤ「００１」の語は、「ＮＥＣ」、「日電」、「奈良」、および「支店」である。

【００４９】

類似度計算部３は、分解した語を結合する（ステップＳ１０５）。すると、レコードＩＤ「００１」のレコードの名称のフィールドは、「ＮＥＣ日電奈良支店」および「日電ＮＥＣ奈良支店」に変換される。

【００５０】
類似度計算部３は、上述したステップＳ１０１からステップＳ１０５の動作を、レコードＩＤ「００２」およびレコードＩＤ「００３」に対しても行う。
【００５１】
具体的には、レコードＩＤが「００２」のレコードの名称のフィールドは、「日電奈良支店」である。類似度計算部３は、「日電奈良支店」に形態素解析を行って語に分解する。具体的には、「日電」と「奈良」と「支店」とに分解する。
【００５２】
類似度計算部３は、変換語辞書５を参照して、分解した各語が同義語であれば代表語に変換する。図３を参照すると、「日電」は代表語であり、「奈良」および「支店」は同義語辞書に登録されていないので変換を行わない。
【００５３】
次に、類似度計算部３は、変換語辞書５を参照して、省略可能語があれば、同じグループＩＤの省略可能語を追加する。図３を参照すると、「ＮＥＣ」と「日電」とが同じグループＩＤの省略可能語であるので、「ＮＥＣ」を追加する。すると、レコードＩＤ「００２」の語は、「ＮＥＣ」、「日電」、「奈良」、および「支店」である。
【００５４】
類似度計算部３は、分解した語を結合する。すると、レコードＩＤ「００２」のレコードの名称のフィールドは、「ＮＥＣ日電奈良支店」および「日電ＮＥＣ奈良支店」に変換される。
【００５５】
同様に、レコードＩＤが「００３」のレコードの名称のフィールドは、「ＮＥＣ奈良支店」である。類似度計算部３は、「ＮＥＣ奈良支店」に形態素解析を行って語に分解する。具体的には、「ＮＥＣ」と「奈良」と「支店」とに分解する。
【００５６】
類似度計算部３は、変換語辞書５を参照して、分解した各語が同義語であれば代表語に変換する。図３を参照すると、「ＮＥＣ」が代表語であり、「奈良」および「支店」は同義語辞書に登録されていないので変換を行わない。
【００５７】
次に、類似度計算部３は、変換語辞書５を参照して、省略可能語があれば、同じグループＩＤの省略可能語を追加する。図３を参照すると、「ＮＥＣ」と「日電」とが同じグループＩＤの省略可能語であるので、「日電」を追加する。すると、レコードＩＤ「００６」の語は、「ＮＥＣ」、「日電」、「奈良」、および「支店」である。
【００５８】
類似度計算部３は、分解した語を結合する。すると、レコードＩＤ「００３」のレコードの名称のフィールドは、「ＮＥＣ日電奈良支店」および「日電ＮＥＣ奈良支店」に変換される。
【００５９】
次に、類似度計算部３は、変換したレコードＩＤ「００１」、「００２」および「００３」の名称のフィールドの相互の類似度を計算する（ステップＳ１０６）。
【００６０】
まず、レコードＩＤ「００１」の名称のフィールドと、レコードＩＤ「００２」の名称のフィールドとの類似度を計算する。レコードＩＤ「００１」の変換後の名称のフィールドは、「ＮＥＣ日電奈良支店」と、「日電ＮＥＣ奈良支店」とであり、レコードＩＤ「００２」の変換後の名称のフィールドは、「ＮＥＣ日電奈良支店」と、「日電ＮＥＣ奈良支店」とである。レコードＩＤ「００１」の「ＮＥＣ日電奈良支店」と、レコードＩＤ「００２」の「ＮＥＣ日電奈良支店」とは、９文字中９文字が合致するので、９×１÷９＝１となり、類似度は１である。
【００６１】
同様に、レコードＩＤ「００１」の「日電ＮＥＣ奈良支店」と、レコードＩＤ「００２

」の「日電ＮＥＣ奈良支店」とは、９文字中９文字が合致するので、９×１÷９＝１となり、類似度は１である。

【００６２】

また、レコードＩＤ「００１」の「ＮＥＣ日電奈良支店」と、レコードＩＤ「００２」の「日電ＮＥＣ奈良支店」とは、９文字中４文字が合致するので、４×１÷９＝０．４４（小数点３桁目四捨五入）となり、類似度は０．４４である。

【００６３】

同様に、また、レコードＩＤ「００１」の「日電ＮＥＣ奈良支店」と、レコードＩＤ「００２」の「ＮＥＣ日電奈良支店」とは、９文字中４文字が合致するので、４×１÷９＝０．４４（小数点３桁目四捨五入）となり、類似度は０．４４である。

【００６４】

ここで、類似度計算部３は、最も類似度が高い値を採用することとする。すると、レコードＩＤ「００１」の名称のフィールドと、レコードＩＤ「００２」の名称のフィールドとの類似度は１である。

【００６５】

同様に、レコードＩＤ「００１」の名称のフィールドと、レコードＩＤ「００３」の名称のフィールドとの類似度を計算すると、類似度は１となる。また、レコードＩＤ「００２」の名称のフィールドと、レコードＩＤ「００３」の名称のフィールドとの類似度を計算すると、類似度は１となる。

【００６６】

次に、類似度計算部３は、レコードＩＤ「００１」、「００２」および「００３」の住所のフィールドの相互の類似度を計算する（ステップＳ１０７）。

【００６７】

レコードＩＤ「００１」の住所のフィールドは「〇〇〇１の１」であり、レコードＩＤ「００２」の住所のフィールドは「〇〇〇１－１」であり、レコードＩＤ「００３」の住所のフィールドは「〇〇〇１－１」である。

【００６８】

レコードＩＤ「００１」の住所のフィールド「〇〇〇１の１」と、レコードＩＤ「００２」の住所のフィールド「〇〇〇１－１」とは、６文字中５文字が合致するので、５×１÷６＝０．８３（小数点３桁目四捨五入）となり、類似度は０．８３である。

【００６９】

レコードＩＤ「００１」の住所のフィールド「〇〇〇１の１」と、レコードＩＤ「００３」の住所のフィールド「〇〇〇１－１」とは、６文字中５文字が合致するので、５×１÷６＝０．８３（小数点３桁目四捨五入）となり、類似度は０．８３である。

【００７０】

レコードＩＤ「００２」の住所のフィールド「〇〇〇１－１」と、レコードＩＤ「００３」の住所のフィールド「〇〇〇１－１」とは、６文字中６文字が合致するので、６×１÷６＝１となり、類似度は１である。

【００７１】

次に、類似度計算部３は、レコードＩＤ「００１」、「００２」および「００３」の電話番号のフィールドの相互の類似度を計算する（ステップＳ１０８）。

【００７２】

レコードＩＤ「００１」の電話番号のフィールドは「０００－１１１－１２３４」であり、レコードＩＤ「００２」の電話番号のフィールドは「０００－１１１－１２３５」であり、レコードＩＤ「００３」の電話番号のフィールドは「０００－１１１－１２３４」である。

【００７３】

レコードＩＤ「００１」の電話番号のフィールド「０００－１１１－１２３４」と、レコードＩＤ「００２」の電話番号のフィールド「０００－１１１－１２３５」とは、１２文字中１１文字が合致するので、１１×１÷１２＝０．９２（小数点３桁目四捨五入）と

なり、類似度は０．９２である。
【００７４】
レコードＩＤ「００１」の電話番号のフィールド「０００－１１１－１２３４」と、レコードＩＤ「００３」の電話番号のフィールド「０００－１１１－１２３４」とは、１２文字中１２文字が合致するので、１２×１÷１２＝１となり、類似度は１である。
【００７５】
レコードＩＤ「００２」の電話番号のフィールド「０００－１１１－１２３５」と、レコードＩＤ「００３」の電話番号のフィールド「０００－１１１－１２３４」とは、１２文字中１１文字が合致するので、１１×１÷１２＝０．９２（小数点３桁目四捨五入）となり、類似度は０．９２である。
【００７６】
類似度計算部３は、レコードＩＤ「００１」、「００２」および「００３」の各フィールドの相互の類似度を、各フィールドの類似度に重みの値を乗じた積を合計した数を、フィールドの数で割った商を、各レコード間の類似度として計算する（ステップＳ１０９）。なお、ここでは、データベース情報入力部４に入力された情報が、名称フィールド、住所フィールド、および電話番号フィールドの重みの比が、１：１：１であることを示していたものとする。
【００７７】
すると、レコードＩＤ「００１」とレコードＩＤ「００２」との類似度は、（１×１＋０．８３×１＋０．９２×１）÷３＝０．９２（小数点３桁目四捨五入）となる。
【００７８】
また、レコードＩＤ「００１」とレコードＩＤ「００３」との類似度は、（１×１＋０．８３×１＋１×１）÷３＝０．９４（小数点３桁目四捨五入）となる。
【００７９】
レコードＩＤ「００２」とレコードＩＤ「００３」との類似度は、（１×１＋１×１＋０．９２×１）÷３＝０．９７（小数点３桁目四捨五入）となる。
【００８０】
類似度計算部３は、計算した各レコードの組の類似度と、類似度を計算したレコードの組とを重複候補抽出部６に出力する。重複候補抽出部６は、類似度計算部３が計算した類似度が、所定の閾値以上である各レコードを、重複レコード候補として抽出する（ステップＳ１１０）。ここで、所定の閾値を０．９０とすると、重複候補抽出部６は、レコードＩＤ「００１」、レコードＩＤ「００２」、およびレコードＩＤ「００３」を重複レコード候補として抽出する。
【００８１】
表示部（図示せず）は、重複候補抽出部６が抽出した各レコードを表示する（ステップＳ１１１）。
【００８２】
以上に述べたように、この実施の形態によれば、同義語や省略可能語による表記の差異があっても、重複する情報をデータベース２から抽出することができる。
【００８３】
実施の形態２．
本発明の第２の実施の形態について、図面を参照して説明する。図５は、本発明の第２の実施の形態の一構成例を示すブロック図である。
【００８４】
本発明の第２の実施の形態の構成は、第１の実施の形態の構成に、重複候補抽出部６が抽出した重複レコード候補が、重複レコードであるか否かを判定するルールである重複判定ルールを記憶する重複判定ルール記憶部８、重複判定ルール記憶部８が記憶しているルールにもとづいて、重複候補抽出部６が抽出した重複レコード候補が重複レコードであるか否かを判定する重複レコード判定部７、および重複レコード判定部７が重複レコードであると判定したレコードを記憶する重複レコードデータベース９を加えたものであり、そ

の他の構成要素は第１の実施の形態と同様なため、その他の構成要素には図１と同じ符号を付し、説明を省略する。

【００８５】

図６は、重複判定ルールの例を示す説明図である。図６に示した例によると、重複判定ルールは、例えば、レコード相互の類似度が特定の値を超えているならば、それらを重複レコードとみなす、というルールや、レコード相互の類似度が特定の値以下であれば、それらを重複レコードではないとみなす、というルールや、いずれかのフィールドの類似度が特定の値以下であれば、それらを重複レコードではないとみなす、というルールや、あるフィールドの類似度が所定の値以上であり、かつ、他のあるフィールドの類似度が所定の値以上であれば、それらを重複レコードとみなす、等である。

【００８６】

図７は、重複判定ルール記憶部８が記憶している重複判定ルールの例を示す説明図である。図７の例に示すように、重複判定ルールは、それぞれ条件部分（図７におけるＩＦ以下の部分）と、結論部分（図７におけるＴＨＥＮ以下の部分）とで構成される。

【００８７】

そして、条件部分には、レコードの組の類似度の値や、フィールドの組の類似度の値が、ある値よりも大きい、小さい、以上、または以下等の条件を、ＡＮＤ、ＯＲ、およびＮＯＴで組み合わせて記述する。

【００８８】

また、結論部分には、条件部分のに記述されている条件に合致するレコードの組を、重複レコードであると記述したり、重複レコードではないと記述したりする。また、結論部分において、条件部分のネスト（入れ子）を記述してもよい。

【００８９】

図７の（１）式に示す例では、レコードの組の類似度の値が１であれば、重複レコードであるというルールを記述している。また、図７の（２）式に示す例では、住所フィールドの組の類似度が０．９を超えていて、かつ、電話番号フィールドの組の類似度が０．９を超えていた場合、名称フィールドの組の類似度が０．９を超えていれば、レコードの組は重複レコードであり、名称フィールドの組の類似度が０．９以下であれば、レコードの組を重複レコードではないというルールを記述している。

【００９０】

類似度計算部３は、計算した各フィールドの組の類似度と、各レコードの組の類似度とを重複候補抽出部６に出力する。重複候補抽出部６は、類似度計算部３が計算した類似度が、所定の閾値以上である各レコードを、重複レコード候補として抽出し、重複レコード候補の各フィールドの組の類似度と、各レコードの組の類似度とを重複レコード判定部７に出力する。

【００９１】

本発明の第２の実施の形態の動作を、具体例を挙げて説明する。まず、図２の例に示したレコードＩＤ「００１」、レコードＩＤ「００２」およびレコードＩＤ「００３」が重複レコードであるか否かを判定する場合を例に説明する。

【００９２】

類似度計算部３が、各フィールドの組の類似度と、各レコードの組の類似度とを計算するまでの動作は、第１の実施の形態における動作と同様なため、説明を省略する。

【００９３】

類似度計算部３は、計算した各フィールドの組の類似度と、各レコードの組の類似度とを重複候補抽出部６に出力する。重複候補抽出部６は、類似度計算部３が計算した類似度が、所定の閾値以上である各レコードを重複レコード候補として抽出し、重複レコード候補の各フィールドの組の類似度と、各レコードの組の類似度とを重複レコード判定部７に出力する。ここで、所定の閾値を０．９とすると、重複候補抽出部６は、レコードＩＤ「００１」、レコードＩＤ「００２」、およびレコードＩＤ「００３」を重複レコード候補として抽出する。

【００９４】
重複レコード判定部７は、レコードＩＤ「００１」とレコードＩＤ「００２」との類似度が０．９２であるので、図７の例に示す式（１）の条件部分（レコードの組の類似度の値が１）に合致しないので、レコードＩＤ「００１」とレコードＩＤ「００２」との重複レコードの判定に、式（１）を適用しない。
【００９５】
重複レコード判定部７は、レコードＩＤ「００１」の住所フィールドと、レコードＩＤ「００２」の住所フィールドとの類似度が１であるが、レコードＩＤ「００１」の電話番号フィールドと、レコードＩＤ「００２」の電話番号フィールドとの類似度が０．８３であるので、図７の例に示す式（２）の条件部分（住所フィールドの組の類似度が０．９を超えていて、かつ、電話番号フィールドの組の類似度が０．９を超えていた場合）に合致しないので、レコードＩＤ「００１」とレコードＩＤ「００２」との重複レコードの判定に、式（２）を適用しない。
【００９６】
また、重複レコード判定部７は、レコードＩＤ「００１」とレコードＩＤ「００３」との類似度が０．９４であるので、図７の例に示す式（１）の条件部分（レコードの組の類似度の値が１）に合致しないので、レコードＩＤ「００１」とレコードＩＤ「００３」との重複レコードの判定に、式（１）を適用しない。
【００９７】
重複レコード判定部７は、レコードＩＤ「００１」の住所フィールドと、レコードＩＤ「００３」の住所フィールドとの類似度が１であるが、レコードＩＤ「００１」の電話番号フィールドと、レコードＩＤ「００３」の電話番号フィールドとの類似度が０．８３であるので、図７の例に示す式（２）の条件部分（住所フィールドの組の類似度が０．９を超えていて、かつ、電話番号フィールドの組の類似度が０．９を超えていた場合）に合致しないので、レコードＩＤ「００１」とレコードＩＤ「００３」との重複レコードの判定に、式（２）を適用しない。
【００９８】
重複レコード判定部７は、レコードＩＤ「００２」とレコードＩＤ「００３」との類似度が０．９７であるので、図７の例に示す式（１）の条件部分（レコードの組の類似度の値が１）に合致しないので、レコードＩＤ「００２」とレコードＩＤ「００３」との重複レコードの判定に、式（１）を適用しない。
【００９９】
重複レコード判定部７は、レコードＩＤ「００２」の住所フィールドと、レコードＩＤ「００３」の住所フィールドとの類似度が１であって、レコードＩＤ「００２」の電話番号フィールドと、レコードＩＤ「００３」の電話番号フィールドとの類似度が１であるので、図７の例に示す式（２）の条件部分（住所フィールドの組の類似度が０．９を超えていて、かつ、電話番号フィールドの組の類似度が０．９を超えていた場合）に合致する。また、レコードＩＤ「００２」の名称フィールドと、レコードＩＤ「００３」の名称フィールドとの類似度が０．９２であるので、式（２）の結果部分における条件部分（名称フィールドの組の類似度が０．９を超えている）に合致するので、レコードＩＤ「００２」とレコードＩＤ「００３」とが重複レコードであると判定する。
【０１００】
重複レコード判定部７は、重複レコードであると判定した各レコードを、重複レコードデータベース９に記憶させる。
【０１０１】
なお、重複レコード判定部７は、重複判定ルール記憶部８が記憶している重複判定ルールを適用しなかったレコードの組を、表示部に表示させてもよい。すると、使用者が重複レコードであるか否かを判定することができる。
【０１０２】
以上に述べたように、この実施の形態によれば、重複レコード判定部７が、予め重複判

定ルール記憶部８が記憶している重複判定ルールにもとづいて、各レコードの組が重複レコードであるか否かを自動的に判定することができる。

【０１０３】

また、重複レコードデータベース９が、重複レコード判定部７が重複レコードであると判定したレコードの組を記憶するため、使用者は、重複レコード判定部７が重複レコードであると判定したレコードの組を確認することができる。

【０１０４】

実施の形態３．

本発明の第３の実施の形態を、図面を参照して説明する。図８は、本発明の第３の実施の形態の一構成例を示すブロック図である。

【０１０５】

本発明の第３の実施の形態の構成は、第２の実施の形態の構成に、重複レコード判定部７が重複レコードであると判定したレコードの組から変換語の候補を抽出して変換語辞書５に登録する変換語候補抽出部１０を加えた点が第２の実施の形態の構成と異なり、その他の点は第２の実施の形態の構成と同様である。そのため、第２の実施の形態と同様な構成要素には、図５と同じ符号を付し、説明を省略する。

【０１０６】

変換語候補抽出部１０は、重複レコード判定部７が重複レコードであると判定したレコードの組を比較して、異なる部分に、例えば、形態素解析等を行って、重複レコードの組における異なる部分の語の組を同義語候補の組として抽出する。

【０１０７】

なお、変換語候補抽出部１０は、重複レコード判定部７が重複レコードであると判定したレコードの組を比較して、一方のレコードが、他方のレコードに含まれる場合には、一方のレコードと他方のレコードとの異なる部分と、共通する部分との組を省略可能語候補の組として抽出する。

【０１０８】

また、変換語候補抽出部１０は、抽出した同義語候補の組が、他の一のレコードに含まれる場合は、抽出した同義語候補の組を省略可能語候補の組とする。

【０１０９】

変換語候補抽出部１０は、抽出した同義語候補の組のうち、他の同義語候補や省略可能語候補の組み合わせで構成される同義語候補の組を、同義語候補の組から除外する。

【０１１０】

また、変換語候補抽出部１０は、省略可能語候補の組のうち、他の同義語候補や省略可能語候補に含まれる省略可能語候補の組を、省略可能語候補の組から除外する。

【０１１１】

変換語候補抽出部１０は、変換語辞書５を参照して、変換語候補、および省略可能語候補の組のうち、変換語辞書５に登録されている語以外の語を変換語辞書５に登録する。

【０１１２】

次に、この実施の形態において、重複レコードから同義語候補および省略可能語候補を名称フィールドから抽出する際の動作を、具体例を挙げて図面を参照して説明する。図９は、本発明の第３の実施の形態の動作を説明するフローチャートである。図１０は、重複レコード判定部７が重複レコードであると判定したレコードの組の例を示す説明図である。

【０１１３】

変換語候補抽出部１０は、重複レコード判定部７が重複レコードであると判定したレコードの組を比較して、異なる部分に、例えば、形態素解析等を行って、重複レコードの組における異なる部分の語を同義語候補として抽出し、一方のレコードが、他方のレコードに含まれる場合には、一方のレコードと他方のレコードとの異なる部分と、共通する部分とを省略可能語候補として抽出する（ステップＳ３０１）。

【０１１４】

具体的には、変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００１」とレコードＩＤ「００２」とを比較して、レコードＩＤ「００１」と、レコードＩＤ「００２」とで異なる部分である「日電ＮＥＣ」と「エヌイーシー」とを同義語候補の組として抽出する。

【０１１５】

変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００１」とレコードＩＤ「００３」とを比較して、レコードＩＤ「００１」と、レコードＩＤ「００３」とで異なる部分である「ＮＥＣ」と「エヌイーシー」とを同義語候補の組として抽出する。

【０１１６】

変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００１」とレコードＩＤ「００４」とを比較すると、レコードＩＤ「００４」が、レコードＩＤ「００１」に含まれるので、異なる部分である「日電」と、共通する部分である「ＮＥＣ奈良支店」とを省略可能語候補の組として抽出する。

【０１１７】

変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００１」とレコードＩＤ「００５」とを比較すると、レコードＩＤ「００５」が、レコードＩＤ「００１」に含まれるので、異なる部分である「ＮＥＣ」と、共通する部分である「日電奈良支店」とを省略可能語候補の組として抽出する。

【０１１８】

変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００２」とレコードＩＤ「００３」とを比較すると、レコードＩＤ「００２」が、レコードＩＤ「００３」に含まれるので、異なる部分である「日電」と、共通する部分である「エヌイーシー奈良支店」とを省略可能語候補の組として抽出する。

【０１１９】

変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００２」とレコードＩＤ「００４」とを比較して、レコードＩＤ「００２」と、レコードＩＤ「００４」とで異なる部分である「エヌイーシー」と「ＮＥＣ」とを同義語候補の組として抽出する。

【０１２０】

変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００２」とレコードＩＤ「００５」とを比較して、レコードＩＤ「００２」と、レコードＩＤ「００５」とで異なる部分である「エヌイーシー」と「日電」とを同義語候補の組として抽出する。

【０１２１】

変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００３」とレコードＩＤ「００４」とを比較して、レコードＩＤ「００３」と、レコードＩＤ「００４」とで異なる部分である「日電エヌイーシー」と「ＮＥＣ」とを同義語候補の組として抽出する。

【０１２２】

変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００３」とレコードＩＤ「００５」とを比較すると、レコードＩＤ「００５」が、レコードＩＤ「００３」に含まれるので、異なる部分である「エヌイーシー」と、共通する部分である「日電奈良支店」とを省略可能語候補の組として抽出する。

【０１２３】

変換語候補抽出部１０は、図１０の例に示したレコードＩＤ「００４」とレコードＩＤ「００５」とを比較して、レコードＩＤ「００４」と、レコードＩＤ「００５」とで異なる部分である「ＮＥＣ」と「日電」とを同義語候補の組として抽出する。

【０１２４】

次に、変換語候補抽出部１０は、抽出した同義語候補の組が、他の一のレコードに含まれる場合は、抽出した同義語候補の組を省略可能語候補の組とする（ステップＳ３０２）。

【０１２５】

具体的には、変換語候補抽出部１０は、レコードＩＤ「００２」とレコードＩＤ「００

５」とを比較して抽出した同義語候補の組である「エヌイーシー」と「日電」とが、レコードＩＤ「００３」の「日電エヌイーシー奈良支店」に含まれるので、同義語候補の組である「エヌイーシー」と「日電」とを省略可能語候補の組とする。

【０１２６】

また、変換語候補抽出部１０は、レコードＩＤ「００４」とレコードＩＤ「００５」とを比較して抽出した同義語候補の組である「ＮＥＣ」と「日電」とが、レコードＩＤ「００１」の「日電ＮＥＣ奈良支店」に含まれるので、同義語候補の組である「ＮＥＣ」と「日電」と省略可能語候補の組とする。

【０１２７】

次に、変換語候補抽出部１０は、抽出した同義語候補の組のうち、他の同義語候補や省略可能語候補の組み合わせで構成される同義語候補の組を、同義語候補の組から除外する（ステップＳ３０３）。

【０１２８】

具体的には、変換語候補抽出部１０は、レコードＩＤ「００４」とレコードＩＤ「００５」とを比較して抽出した同義語候補の組である「ＮＥＣ」と「日電」とを組み合わせると、レコードＩＤ「００１」とレコードＩＤ「００２」とを比較して同義語候補として抽出した「日電ＮＥＣ」を構成するので、レコードＩＤ「００４」とレコードＩＤ「００５」とを比較して抽出した同義語候補の組である「ＮＥＣ」と「日電」とを、同義語候補の組から除外する。

【０１２９】

変換語候補抽出部１０は、レコードＩＤ「００２」とレコードＩＤ「００５」とを比較して抽出した同義語候補の組である「エヌイーシー」と「日電」とを組み合わせると、レコードＩＤ「００３」とレコードＩＤ「００４」とを比較して同義語候補として抽出した「日電エヌイーシー」を構成するので、レコードＩＤ「００２」とレコードＩＤ「００５」とを比較して抽出した同義語候補の組である「エヌイーシー」と「日電」とを、同義語候補の組から除外する。

【０１３０】

変換語候補抽出部１０は、省略可能語候補の組のうち、他の同義語候補や省略可能語候補に含まれる省略可能語候補の組を、省略可能語候補の組から除外する（ステップＳ３０４）。

【０１３１】

具体的には、変換語候補抽出部１０は、レコードＩＤ「００１」とレコードＩＤ「００４」とを比較して抽出した省略可能語候補の組である「日電」と「ＮＥＣ奈良支店」とは、レコードＩＤ「００４」とレコードＩＤ「００５」とを比較して抽出した同義語候補の組である「ＮＥＣ」と「日電」とを含むので省略可能語候補から除外する。

【０１３２】

変換語候補抽出部１０は、レコードＩＤ「００１」とレコードＩＤ「００５」とを比較して抽出した省略可能語候補の組である「ＮＥＣ」と「日電奈良支店」とは、レコードＩＤ「００４」とレコードＩＤ「００５」とを比較して抽出した同義語候補の組である「ＮＥＣ」と「日電」とを含むので省略可能語候補から除外する。

【０１３３】

変換語候補抽出部１０は、レコードＩＤ「００２」とレコードＩＤ「００３」とを比較して抽出した省略可能語候補の組である「日電」と「エヌイーシー奈良支店」とは、レコードＩＤ「００２」とレコードＩＤ「００５」とを比較して抽出した同義語候補の組である「エヌイーシー」と「日電」とを含むので省略可能語候補から除外する。

【０１３４】

変換語候補抽出部１０は、レコードＩＤ「００３」とレコードＩＤ「００５」とを比較して抽出した省略可能語候補の組である「エヌイーシー」と「日電奈良支店」とは、レコードＩＤ「００２」とレコードＩＤ「００５」とを比較して抽出した同義語候補の組である「エヌイーシー」と「日電」とを含むので省略可能語候補から除外する。

【０１３５】
　すると、レコードＩＤ「００１」とレコードＩＤ「００３」とを比較して抽出された同義語候補の組、およびレコードＩＤ「００２」とレコードＩＤ「００４」とを比較して抽出された同義語候補の組である「ＮＥＣ」と「エヌイーシー」とが同義語候補の組となる。
【０１３６】
　また、レコードＩＤ「００２」とレコードＩＤ「００５」とを比較して抽出された同義語候補の組、およびレコードＩＤ「００４」とレコードＩＤ「００５」とを比較して抽出された同義語候補の組である「エヌイーシー」と「日電」とが、省略可能語候補に変更されて省略可能語候補の組となる。
【０１３７】
　変換語候補抽出部１０は、変換語辞書５を参照して、変換語候補、および省略可能語候補の組のうち、変換語辞書５に登録されている語以外の語を変換語辞書５に登録する（ステップＳ３０５）。
【０１３８】
　なお、変換語候補抽出部１０は、同義語候補の組である「ＮＥＣ」と「エヌイーシー」とのいずれかを代表語として、変換語辞書５に登録する。変換語候補抽出部１０は、例えば、５０音順のや、アルファベット順の早い方の語や、文字数の少ない方の語を代表語として、同義語候補の組の語を変換語辞書５に登録する。
【０１３９】
　このとき、同義語候補の組の語のいずれかが既に代表語として変換語辞書５に登録されていた場合、変換語候補抽出部１０は、同義語候補の組の他の語を、その代表語の同義語として変換語辞書５に登録する。
【０１４０】
　なお、同義語候補の組の語のすべてが既に代表語として変換語辞書５に登録されていた場合、変換語候補抽出部１０は、いずれか１つの語を代表語として変換語辞書５に登録し、同義語候補の組の他の語を、その代表語の同義語として変換語辞書５に登録する。このとき、変換語候補抽出部１０は、既に代表語として変換語辞書５に登録されていた語の同義語を、代表語として変換語辞書５に登録した語の同義語として、変換語辞書５に登録する。
【０１４１】
　変換語候補抽出部１０は、省略可能語候補の組が変換語辞書５に登録されていなかった場合、新たにグループＩＤを決定して省略可能語候補として変換語辞書５に登録する。なお、新たなグループＩＤは、例えば、既に登録済みのグループＩＤの最大値に１を加えた値とする。
【０１４２】
　変換語候補抽出部１０は、省略可能語候補の組のうち、いずれかが既に省略可能語候補として変換語辞書５に登録されていた場合、他の省略可能語候補を、既に変換語辞書５に登録されている省略可能語候補と同じグループＩＤで変換語辞書５に登録する。
【０１４３】
　変換語候補抽出部１０は、省略可能語候補の組がすべて既に省略可能語候補として同じグループＩＤで変換語辞書５に登録されていた場合は、変換語辞書５に登録する動作を行わない。変換語候補抽出部１０は、省略可能語候補の組を構成する省略可能語候補のそれぞれが、異なるグループＩＤで既に省略可能語候補として変換語辞書５に登録されていた場合、グループＩＤの値を比較して、グループＩＤの値が大きい方の省略可能語候補のグループＩＤの値を、グループＩＤが小さい方の値に変更する。
【０１４４】
　なお、変換語候補抽出部１０は、同義語候補または省略可能語候補を表示部に表示して、同義語候補または省略可能語候補として変換語辞書５に登録するか否かを使用者に選択させてもよい。

【０１４５】
以上に述べたように、この実施の形態によれば、重複レコード判定部７が重複レコードとして判定したレコードに含まれる語のうち、所定の条件に合致する語を、同義語候補または省略可能語候補として変換語辞書５に登録することができる。
【０１４６】
実施の形態４.
本発明の第４の実施の形態を、図面を参照して説明する。図１１は、本発明の第４の実施の形態の一構成例を示すブロック図である。
【０１４７】
本発明の第４の実施の形態の構成は、第３の実施の形態の構成に、同義語候補、および省略可能語候補を変換語辞書５に登録するか否か、および重複レコード候補を重複レコードデータベース９に登録するか否かを使用者に確認する入出力部１１を加えたものであり、その他の構成要素は第３の実施の形態と同様なため、その他の構成要素には図８と同じ符号を付し、説明を省略する。
【０１４８】
入出力部１１は、表示部である液晶ディスプレイ等と、入力手段であるキーボード等によって実現される。
【０１４９】
次に、本発明の第４の実施の形態の動作を、図面を参照して説明する。図１２は、本発明の第４の実施の形態の動作を説明するフローチャートである。
【０１５０】
まず、類似度計算部３が、データベース２に登録されている情報を読み込む（ステップＳ４０１）。
【０１５１】
類似度計算部３は、第１の実施の形態における動作と同様の動作を行い、各レコードの組の類似度を計算し、計算した各レコードの組の類似度と、類似度を計算したレコードの組とを重複候補抽出部６に出力する（ステップＳ４０２）。
【０１５２】
重複候補抽出部６は、例えば、第１の実施の形態における動作と同様の動作を行い、重複レコード候補をデータベース２から抽出する（ステップＳ４０３）。
【０１５３】
重複レコード判定部７は、第２の実施の形態における動作と同様の動作を行い、重複候補抽出部６が抽出した重複レコード候補が重複レコードであるか否かを判定し（ステップＳ４０４）、重複レコードであると判定した重複レコード候補を重複レコードデータベース９に記憶させる。
【０１５４】
変換語候補抽出部１０は、第３の実施の形態における動作と同様の動作を行い、重複レコード判定部７が重複レコードであると判定したレコードから、同義語候補および省略可能語候補（以下、単に変換語候補という）を抽出する（ステップＳ４０５）。
【０１５５】
変換語候補抽出部１０は、変換語候補を抽出すると、抽出した変換語候補を入出力部１１に出力する。入出力部１１は、変換語候補抽出部１０が入力した変換語候補を表示し、使用者に変換語候補が変換語であるか否かを判定させる。
【０１５６】
使用者が、入出力部１１を操作して変換語候補が変換語であると判定すると、変換語候補抽出部１０は、変換語候補を変換語辞書５に登録する（ステップＳ４０６）。
【０１５７】
変換語候補抽出部１０が、変換語候補を変換語辞書５に登録すると、重複候補抽出部６は、第１の実施の形態における動作と同様の動作を行い、重複レコード候補をデータベース２から抽出する（ステップＳ４０３）。新たな変換語が変換語辞書５に登録されると、

新たに重複レコードと判定されるレコードが発生する可能性があるからである。

【０１５８】

変換語候補抽出部１０が変換語候補の抽出を終了したり、使用者が、入出力部１１を操作して変換語候補が変換語であるか否かの判定を拒否したりすると、重複レコード判定部７は、重複レコードであると判定しなかった重複レコード候補を入出力部１１に出力する。入出力部１１は、重複レコード判定部７が入力した重複レコード候補を表示し、使用者に重複レコード候補が重複レコードであるか否かを判定させる（ステップＳ４０７）。なお、重複レコード判定部７は、重複レコードであると判定しなかった重複レコード候補のうち、重複する可能性のあるレコードの組み合わせの数が多い順番で、重複レコード候補を入出力部１１に出力してもよい。

【０１５９】

使用者が、入出力部１１を操作して重複レコード候補が重複レコードであると判定すると、重複レコード判定部７は、重複レコード候補を重複レコードであると判定し、重複レコードデータベース９に登録する（ステップＳ４０８）。

【０１６０】

重複レコード判定部７が、重複レコード候補を重複レコードデータベース９に登録すると、変換語候補抽出部１０は、第３の実施の形態における動作と同様の動作を行い、重複レコード判定部７が重複レコードであると判定したレコードから、変換語候補を抽出する（ステップＳ４０５）。重複レコードが増加すると、増加した重複レコードから新たに変換語候補が抽出される可能性があるからである。

【０１６１】

重複レコード判定部７が、重複レコードであると判定しなかった重複レコード候補をすべて入出力部１１に出力したり、使用者が、入出力部１１を操作して重複レコード候補の判定を拒否したりすると、変換語候補抽出部１０は、第３の実施の形態における動作と同様の動作を行い、重複レコード判定部７が重複レコードであると判定したレコードから変換語候補を抽出する（ステップＳ４０５）。

【０１６２】

変換語候補抽出部１０が、第３の実施の形態における動作と同様の動作を行い、重複レコード判定部７が重複レコードであると判定したレコードから変換語候補を抽出する動作を終了すると、すべての構成要素は動作を終了する。

【０１６３】

以上に述べたように、この実施の形態によれば、重複レコードを検出するために、使用者に確認する回数を減らすことができる。

【０１６４】

また、使用者が、変換語候補が変換語であると判定した場合に増加した新たな重複レコードから変換語候補を抽出することができる。

【０１６５】

さらに、使用者が、重複レコード候補が重複レコードであると判定した場合に増加した新たな重複レコードから変換語候補を抽出することができる。

【０１６６】

なお、重複レコード判定部７は、重複レコードであると判定しなかった重複レコード候補が複数存在した場合、最も語の数の多い重複レコード候補から順に入出力部１１に出力してもよい。すると、変換語候補抽出部１０が変換語候補を抽出する可能性が高い順に重複レコード候補が入出力部１１に出力されて使用者が重複レコードであるか否かを判定するため、変換語候補抽出部１０が重複レコード候補から多くの変換語候補を抽出すると、語の数の少ない重複レコード候補からは変換語候補抽出部１０が変換語候補を抽出する可能性が低くなり、変換語候補や、重複レコードを検出するために使用者に確認する回数を減らすことができる。

【０１６７】

また、変換語候補抽出部１０は、使用者が複数の変換語候補を変換語であると判定する

と、使用者が判定した複数の変換語を類似度計算部３に出力し、類似度計算部３は、使用者が判定した複数の変換語に応じてデータベース２の該当するレコードを変換し、重複候補抽出部６は、変換されたレコードにもとづいて重複レコード候補の組を抽出してもよい。そして、重複レコード判定部７は、重複候補抽出部６が抽出した重複レコード候補の組のうち、重複レコード候補の組を構成する重複レコード候補の数が多い順に、重複レコード候補を入出力部１１に出力してもよい。

【０１６８】

重複レコード判定部７は、使用者が複数の重複レコード候補を重複レコードであると判定すると、使用者が重複レコードであると判定した重複レコード候補を変換語候補抽出部１０に出力し、変換語候補抽出部１０は、抽出した変換語候補の数が多い重複レコード候補から抽出した変換語候補を、順に入出力部１１に出力してもよい。

【０１６９】

実施の形態５．

本発明の第５の実施の形態を、図面を参照して説明する。図１３は、本発明の第５の実施の形態の一構成例を示すブロック図である。

【０１７０】

本発明の第５の実施の形態の構成は、本発明の第４の実施の形態の構成に、重複レコードデータベース９に登録されている重複レコードの組を構成するレコードのうち、一のレコードを除いて、他のレコードをデータベース２から削除する重複レコード削除部１２を加えたものであり、その他の構成要素は第４の実施の形態と同様なため、その他の構成要素には図１１と同じ符号を付し、説明を省略する。

【０１７１】

重複レコード削除部１２は、重複レコードデータベース９に登録された重複レコードの組を入出力部１１に出力して、使用者に削除するレコードを選択させ、使用者が選択したレコードをデータベース２から削除してもよい。また、重複レコード削除部１２は、重複レコードデータベース９に登録された重複レコードの組のうち、最もレコードＩＤの値の小さいレコード以外のレコードをデータベース２から削除してもよい。

【０１７２】

また、重複レコード削除部１２は、削除したレコードの記録を記憶してもよい。

【０１７３】

以上に述べたように、この実施の形態によれば、重複レコード判定部７が、重複レコード判定ルールにもとづいて重複レコードであると判定した重複レコードを、重複レコードデータベース９に登録してから、重複レコード削除部１２を介してデータベース２から削除するため、使用者が削除するレコードを確認したり、削除したレコードの記録を記憶させておいたりすることができる。

【０１７４】

実施の形態６．

本発明の第６の実施の形態を、図面を参照して説明する。図１４は、本発明の第６の実施の形態の一構成例を示すブロック図である。

【０１７５】

本発明の第６の実施の形態の構成は、第３の実施の形態の構成に、使用者が新たにデータベース２に追加する情報を入力するデータベース登録部１３と、データベース２に登録されている情報を検索する検索部１４と、使用者に請求する料金を算出する検索料金算出部１５とを加えたものであり、その他の構成要素は第３の実施の形態と同様なため、その他の構成要素には図８と同じ符号を付し、説明を省略する。

【０１７６】

データベース登録部１３は、使用者が新たにデータベース２に追加する情報を入力すると、変換語辞書５に登録されている同義語と省略可能語とにもとづいて、使用者が入力した情報と重複する情報である可能性のあるレコードを表示部に表示させる。

【０１７７】

例えば、図２の例に示す情報がデータベース２に登録され、図３の例に示す情報が変換語辞書５に登録されている場合に、使用者が、データベース登録部１３に、名称が「日本電気奈良支店」である情報を入力する。

【０１７８】

すると、データベース登録部１３は、入力された「日本電気奈良支店」に形態素解析等の方法を用いて、「日本電気奈良支店」を、「日本電気」と「奈良支店」との語に分解する。

【０１７９】

そして、データベース登録部１３は、同義語辞書５を参照して、「日本電気」および「奈良支店」の同義語と省略可能語とを抽出する。「日本電気」の代表語である同義語は、「ＮＥＣ」であるため、データベース登録部１３は、「ＮＥＣ」を抽出する。また、データベース登録部１３は、「ＮＥＣ」を代表語とする同義語である「エヌイーシー」を抽出する。

【０１８０】

さらに、データベース登録部１３は、同義語辞書５を参照して、「日本電気」、「ＮＥＣ」、および「エヌイーシー」のいずれかの省略可能語を抽出する。具体的には、「ＮＥＣ」の省略可能語として「日電」を抽出する。

【０１８１】

そして、データベース登録部１３は、同義語辞書５から抽出した語や、使用者が入力した情報の語を組み合わせて、使用者が入力した情報から変形した可能性がある情報を生成する。具体的には、「ＮＥＣ奈良支店」、「エヌイーシー奈良支店」、「日電ＮＥＣ奈良支店」、「日電エヌイーシー奈良支店」、「ＮＥＣ日電奈良支店」、「エヌイーシー日電奈良支店」、「日電奈良支店」等を生成する。

【０１８２】

データベース登録部１３は、生成した情報と合致する情報がデータベース２に登録されているか否かを検索して、データベース２から合致する情報を抽出する。すると、図２の例に示すレコードＩＤ「００１」、レコードＩＤ「００２」、およびレコードＩＤ「００３」が抽出される。

【０１８３】

データベース登録部１３は、抽出した各レコードを、重複可能性のあるレコードとして表示部に表示させる。

【０１８４】

検索部１４は、上述したデータベース登録部１３の動作と同様な動作を行って、データベース２から、使用者が検索部１４に検索キーとして入力した情報、およびその情報から変形した可能性がある情報に合致する情報を、検索結果として表示部に表示させる。

【０１８５】

検索料金算出部１５は、検索部１４が検索結果を表示部に表示させると、使用者に請求する所定の料金を算出する。なお、使用者が、データベース２の所有者または管理者に、データベース２の使用料金を支払っている場合は、検索料金算出部１５は、データベース２の所有者または管理者に請求する料金を算出してもよい。

【０１８６】

以上に述べたように、この実施の形態によれば、使用者がデータベース２に情報を登録する際に、重複する可能性のある情報を使用者に提示するため、新しく登録する情報が、重複レコードとなることを防ぐことができる。

【０１８７】

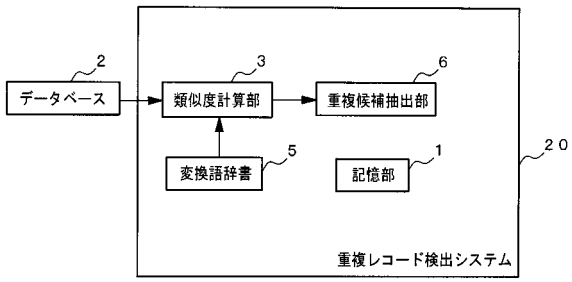
また、検索部１４がデータベース２に登録されている情報を検索し、検索料金算出部１５が、検索部１４が行った情報の検索に応じた料金を算出するため、重複レコード検出システム２０の所有者は、料金を使用者またはデータベース２の所有者または管理者に請求することができる。

【０１８８】

実施の形態７．
本発明の第７の実施の形態を、図面を参照して説明する。図１５は、本発明の第７の実施の形態の一構成例を示すブロック図である。
【０１８９】
本発明の第７の実施の形態の構成は、第５の実施の形態の構成に、重複レコード削除部１２がデータベース２から削除した情報に応じて、データベース２の所有者等に請求する料金を算出する削除料金算出部１６を加えたものであり、その他の構成要素は第５の実施の形態と同様なため、その他の構成要素には図１３と同じ符号を付し、説明を省略する。
【０１９０】
重複レコード削除部１２は、第５の実施の形態における動作と同様の動作を行い、重複レコードをデータベース２から削除する。
【０１９１】
削除料金算出部１６は、重複レコード削除部１２がデータベース２から削除した情報に応じて、データベース２の所有者または管理者に請求する料金を算出する。
【０１９２】
以上に述べたように、この実施の形態によれば、データベース２に登録されている重複レコードの削除に応じた料金を、データベース２の所有者または管理者に請求することができる。
【０１９３】
なお、重複レコード検出システム２０の所有者等は、重複レコード検出システム２０に他のデータベースに登録されている情報を入力して、重複レコード削除部１２に情報を削除させてもよい。すると、変換語候補抽出部１０が、変換語を変換語辞書５に登録するため、重複レコード判定部７による重複レコードの判定精度を向上させることができる。そのため、例えば、第４の実施の形態で、使用者が、重複レコード候補が重複レコードであるか否かの判定を行う回数を減らすことができる。
【産業上の利用可能性】
【０１９４】
本発明は、データベースに重複して登録されている情報を抽出するシステムに適用することができる。
【図面の簡単な説明】
【０１９５】
【図１】本発明の第１の実施の形態の一構成例を説明するブロック図である。
【図２】データベースに登録されている情報の例を示す説明図である。
【図３】変換語辞書に登録されている情報の例を示す説明図である。
【図４】本発明の第１の実施の形態の動作を説明するフローチャートである。
【図５】本発明の第２の実施の形態の一構成例を示すブロック図である。
【図６】重複判定ルールの例を示す説明図である。
【図７】重複判定ルール記憶部が記憶している重複判定ルールの例を示す説明図である。
【図８】本発明の第３の実施の形態の一構成例を示すブロック図である。
【図９】本発明の第３の実施の形態の動作を説明するフローチャートである。
【図１０】重複レコード判定部が重複レコードであると判定したレコードの組の例を示す説明図である。
【図１１】本発明の第４の実施の形態の一構成例を示すブロック図である。
【図１２】本発明の第４の実施の形態の動作を説明するフローチャートである。
【図１３】本発明の第５の実施の形態の一構成例を示すブロック図である。
【図１４】本発明の第６の実施の形態の一構成例を示すブロック図である。
【図１５】本発明の第７の実施の形態の一構成例を示すブロック図である。
【符号の説明】
【０１９６】
１　　記憶部

２　データベース
３　類似度計算部
４　データベース情報入力部
５　変換語辞書
６　重複候補抽出部
７　重複レコード判定部
８　重複判定ルール記憶部
９　重複レコードデータベース
１０　変換語候補抽出部
１１　入出力部
１２　重複レコード削除部
１３　データベース登録部
１４　検索部
１５　検索料金算出部
１６　削除料金算出部
２０　重複レコード検出システム

【図１】



【図２】

ID	名称	住所	電話番号
001	エヌイーシー奈良支店	○○○1の1	000−111−1234
002	日電奈良支店	○○○1−1	000−111−1235
003	NEC奈良支店	○○○1−1	000−111−1234
…	…	…	…

【図３】

変換語辞書

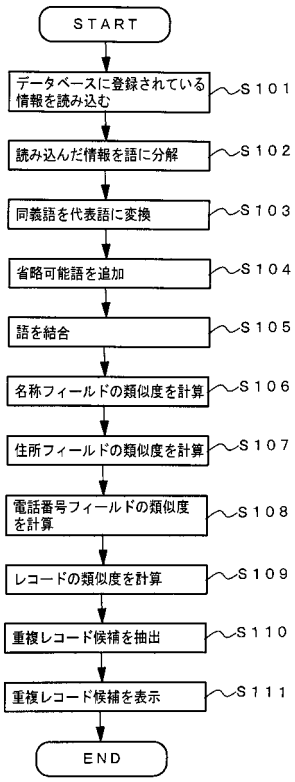
同義語辞書

代表語	同義語
NEC	エヌイーシー
NEC	日本電気
日電	日本電気株式会社

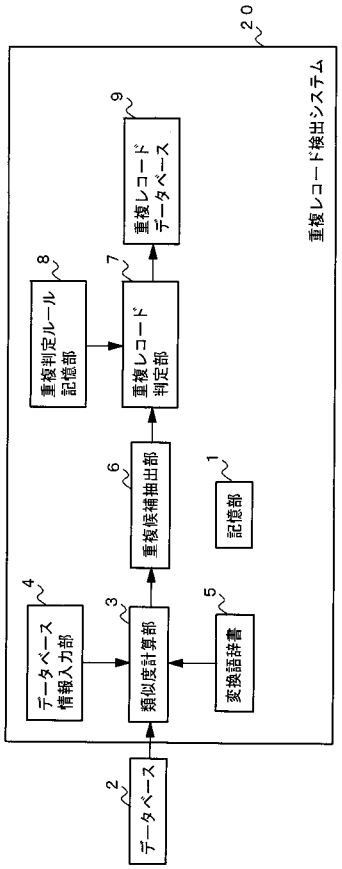
省略可能語辞書

省略可能語	グループID
NEC	1
日電	1

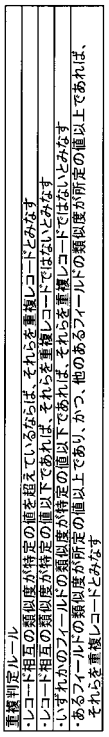
【図４】



【図５】



【図６】

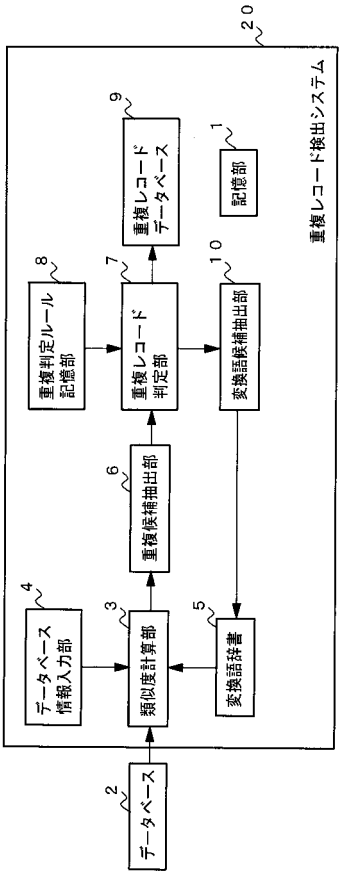


【図７】

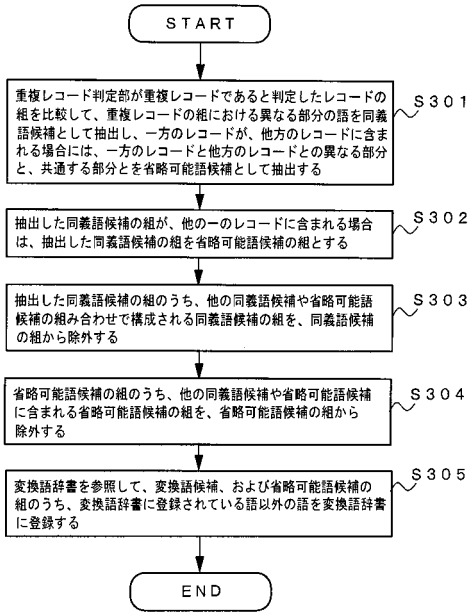
```
IF 〔類似度＝1.0〕 THEN 〔重複レコードである〕      ……(1)

IF 〔 (住所フィールドの類似度＞0.9)
      AND(電話番号フィールドの類似度＞0.9)〕
THEN 〔
    IF 〔名称フィールドの類似度＞0.9〕
        THEN〔重複レコードである〕
    IF 〔名称フィールドの類似度≦0.9〕
        THEN〔重複レコードではない〕
〕      ……(2)
```

【図８】



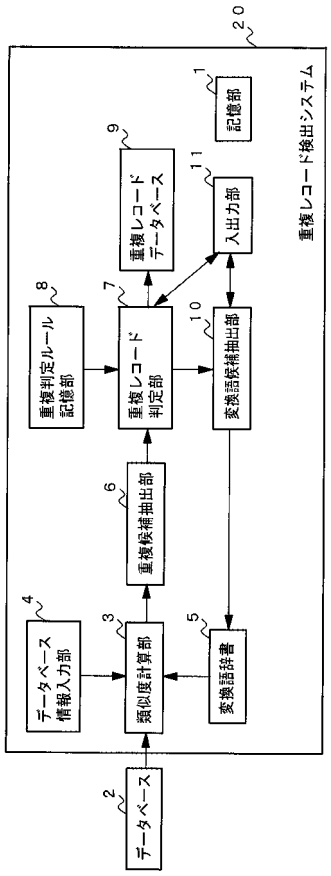
【図９】



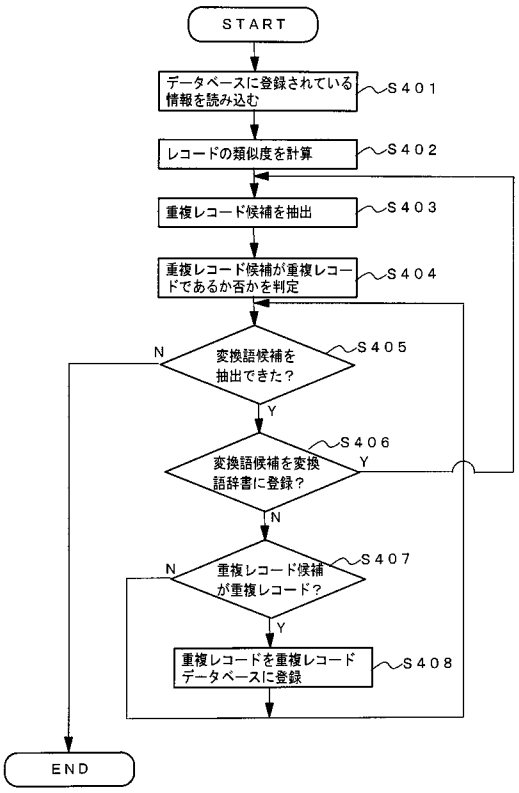
【図１０】

ID	名称	住所	電話番号
001	日電NEC奈良支店	○○○1−1	000−111−1234
002	エヌイーシー奈良支店	○○○1−1	000−111−1234
003	日電エヌイーシー奈良支店	○○○1−1	000−111−1234
004	NEC奈良支店	○○○1−1	000−111−1234
005	日電奈良支店	○○○1−1	000−111−1234

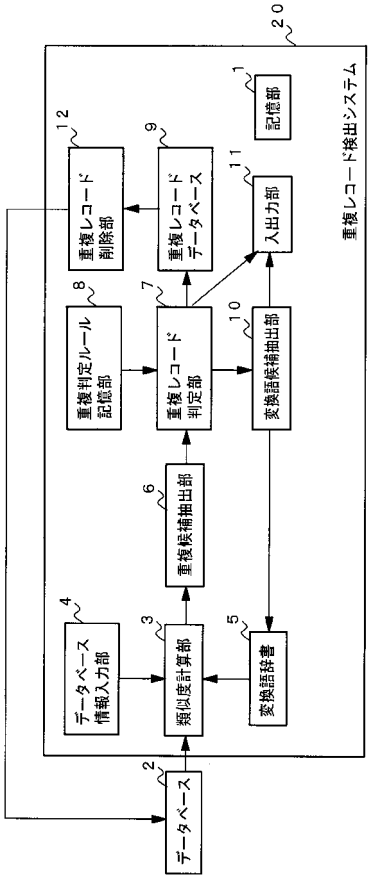
【図１１】



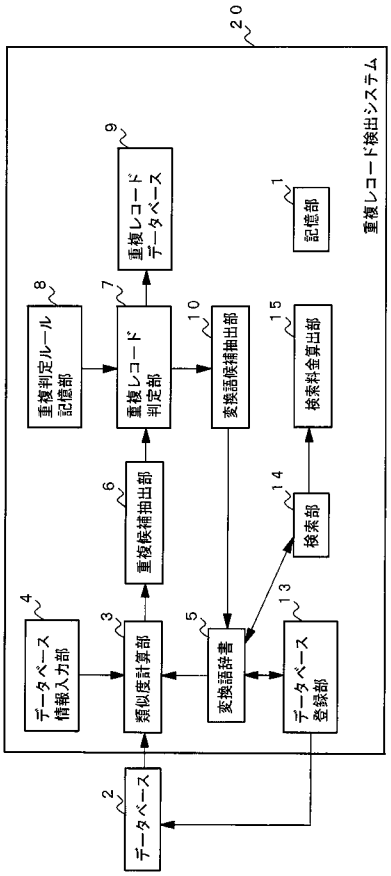
【図１２】



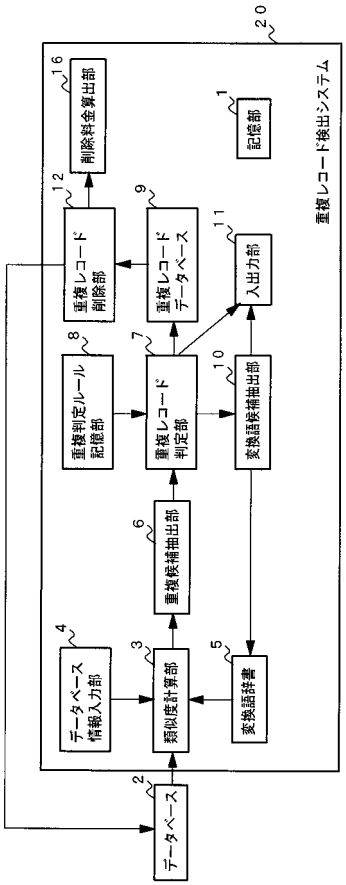
【図１３】



【図１４】


重複レコード検出システム
20
データベース
2
データベース情報入力部
4
類似度計算部
3
重複候補抽出部
6
重複判定ルール記憶部
8
重複レコード判定部
7
重複レコードデータベース
9
変換語候補抽出部
10
変換語辞書
5
検索部
14
データベース登録部
13
検索料金算出部
15
記憶部
1

【図１５】


重複レコード検出システム
20
データベース
2
データベース情報入力部
4
類似度計算部
3
重複候補抽出部
6
重複判定ルール記憶部
8
重複レコード判定部
7
重複レコードデータベース
9
変換語候補抽出部
10
変換語辞書
5
重複レコード削除部
12
削除料金算出部
16
入出力部
11
記憶部
1

フロントページの続き

(72)発明者　齋藤　悠
東京都港区芝五丁目７番１号　日本電気株式会社内

審査官　吉田　誠

(56)参考文献　特開平１１－１８４８８４（ＪＰ，Ａ）
特開平１０－２７５１５９（ＪＰ，Ａ）
特開平０７－１９２０５３（ＪＰ，Ａ）
特開平０６－２６６７６９（ＪＰ，Ａ）
特開２００３－１７３３４５（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１７／３０