

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6860668号
(P6860668)

(45) 発行日 令和3年4月21日 (2021.4.21)

(24) 登録日 令和3年3月30日 (2021.3.30)

(51) Int. Cl. F I
GO 6 F 16/33 (2019.01) GO 6 F 16/33
GO 6 F 16/36 (2019.01) GO 6 F 16/36

請求項の数 12 (全 24 頁)

(21) 出願番号 特願2019-528581 (P2019-528581)
 (86) (22) 出願日 平成29年11月17日 (2017.11.17)
 (65) 公表番号 特表2020-501255 (P2020-501255A)
 (43) 公表日 令和2年1月16日 (2020.1.16)
 (86) 国際出願番号 PCT/CN2017/111604
 (87) 国際公開番号 W02018/095281
 (87) 国際公開日 平成30年5月31日 (2018.5.31)
 審査請求日 令和1年7月22日 (2019.7.22)
 (31) 優先権主張番号 201611055619.8
 (32) 優先日 平成28年11月25日 (2016.11.25)
 (33) 優先権主張国・地域又は機関
 中国 (CN)

(73) 特許権者 520015461
 アドバンスド ニュー テクノロジーズ
 カンパニー リミテッド
 英国領ケイマン諸島 グランド ケイマン
 ケーワイ1-9008 ジョージ タウ
 ン ホスピタル ロード 27 ケイマン
 コーポレート センター
 (74) 代理人 100188558
 弁理士 飯田 雅人
 (74) 代理人 100205785
 弁理士 ▲高▼橋 史生

早期審査対象出願

最終頁に続く

(54) 【発明の名称】 名前マッチング方法および装置

(57) 【特許請求の範囲】

【請求項 1】

名前をマッチングさせるためのコンピュータにより実施される方法であって、
 マッチング対象の名前を含む入力を受信するステップであって、前記名前が複数の単語
 を含む、ステップと、

前記名前に対応する第1の名前セットを決定するステップであって、前記第1の名前セッ
 トが複数の要素を含む、ステップと、

標準名前セットを生成するために、類似度に基づいて前記複数の単語のそれぞれを前記
 複数の要素のそれぞれとマッチングさせるステップと、

前記名前の1つまたは複数の名前の文字が標準名前の1つまたは複数の標準名前の文字と
 は異なるにもかかわらず、前記名前が前記標準名前セット内の少なくとも1つの標準名前
 と同義であると判定するステップであって、前記マッチング対象の名前が前記標準名前セ
 ット内の前記少なくとも1つの標準名前と同義であるかどうかを判定するために、前記マ
 ッチング対象の名前に対して、省略形検出、呼称検出、多言語検出、または別名検出のう
 ちの少なくとも1つを実行することを含む、ステップと、

前記マッチング対象の名前のマッチング結果を決定するステップであって、

前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前
 と同義であるとの判定に回答して、前記少なくとも1つの標準名前を前記マッチング対象
 の名前の前記マッチング結果として決定するステップと、

前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前

10

20

と同義でないとの判定に応答して、1つまたは複数の類似度アルゴリズムを使用して前記名前を前記標準名前とマッチングさせることによって前記名前の前記マッチング結果を決定するステップであって、前記名前のマッチング結果が、前記複数の単語のそれぞれの複数のマッチング度の重み付き加算を含み、前記複数のマッチング度が、テキストマッチング度、音声マッチング度、および文字列マッチング度を含む、ステップと
を含む、ステップと

を含む、方法。

【請求項2】

マッチングさせる必要がない名前の所定のセットを読み出すステップと、
前記マッチング対象の名前が名前の前記所定のセット内に含まれないと判定するステップと
をさらに含む、請求項1に記載の方法。

10

【請求項3】

前記複数の単語のそれぞれをマッチングするステップが、
前記第1の名前セットの各名前のインデックスを決定するステップであって、前記インデックスが前記名前に関連付けられた予め確立された単語を含む、ステップと、
前記マッチング対象の名前に含まれる各単語を取得するために前記マッチング対象の名前を分割するステップと、
前記第1の名前セットのサブセットを取得するために、文字列マッチングアルゴリズムを使用することによって前記マッチング対象の名前に含まれる各単語と、各インデックスとに対して類似度マッチングを実行するステップであって、前記サブセットが正常にマッチングされた各インデックスによってインデックスされた名前を含み、前記文字列マッチングアルゴリズムが、接頭辞ツリーマッチングアルゴリズム、辞書ツリーマッチングアルゴリズム、文字列類似度マッチングアルゴリズム、および発音類似度マッチングアルゴリズムのうちの1つまたは複数を含む、ステップと、
各サブセットに基づいて前記マッチング対象の名前をマッチングさせるために使用される前記標準名前セットを決定するステップと
を含む、請求項1に記載の方法。

20

【請求項4】

前記標準名前セット内の前記標準名前に含まれる各単語と、前記マッチング対象の名前に含まれる各単語との間の類似度に基づいて、前記標準名前セット内の前記名前に含まれる各単語を前記マッチング対象の名前に含まれる各単語と、位置合わせされた標準名前セットとして、位置合わせするステップをさらに含む、
前記名前が前記少なくとも1つの標準名前と同義であると判定することが、
前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前と同義であるかどうかを判定するために、前記位置合わせされた標準名前セットに基づいて前記マッチング対象の名前を検出するステップを含む、請求項1に記載の方法。

30

【請求項5】

コンピュータシステムによって実行可能な1つまたは複数の命令を記憶したコンピュータ可読記録媒体であって、前記1つまたは複数の命令は、前記コンピュータシステムに、
マッチング対象の名前を含む入力を受信するステップであって、前記名前が複数の単語を含む、ステップと、
前記名前に対応する第1の名前セットを決定するステップであって、前記第1の名前セットが複数の要素を含む、ステップと、
標準名前セットを生成するために、類似度に基づいて前記複数の単語のそれぞれを前記複数の要素のそれぞれとマッチングさせるステップと、
前記名前の1つまたは複数の名前の文字が標準名前の1つまたは複数の標準名前の文字とは異なるにもかかわらず、前記名前が前記標準名前セット内の少なくとも1つの標準名前と同義であると判定するステップであって、前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前と同義であるかどうかを判定するために、前記マ

40

50

ッチング対象の名前に対して、省略形検出、呼称検出、多言語検出、または別名検出のうちの少なくとも1つを実行することを含む、ステップと、

前記マッチング対象の名前のマッチング結果を決定するステップであって、

前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前と同義であるとの判定に応答して、前記少なくとも1つの標準名前を前記マッチング対象の名前の前記マッチング結果として決定するステップと、

前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前と同義でないとの判定に応答して、1つまたは複数の類似度アルゴリズムを使用して前記名前を前記標準名前とマッチングさせることによって前記名前の前記マッチング結果を決定するステップであって、前記名前のマッチング結果が、前記複数の単語のそれぞれの複数のマッチング度の重み付き加算を含み、前記複数のマッチング度が、テキストマッチング度、音声マッチング度、および文字列マッチング度を含む、ステップと

を含む、ステップと

を行わせる、コンピュータ可読記憶媒体。

【請求項6】

マッチングさせる必要がない名前の所定のセットを読み出すステップと、

前記マッチング対象の名前が名前の前記所定のセット内に含まれないと判定するステップと

をさらに行わせる、請求項5に記載のコンピュータ可読記憶媒体。

【請求項7】

前記複数の単語のそれぞれをマッチングするステップが、

前記第1の名前セットの各名前のインデックスを決定するステップであって、前記インデックスが前記名前に関連付けられた予め確立された単語を含む、ステップと、

前記マッチング対象の名前に含まれる各単語を取得するために前記マッチング対象の名前を分割するステップと、

前記第1の名前セットのサブセットを取得するために、文字列マッチングアルゴリズムを使用することによって前記マッチング対象の名前に含まれる各単語と、各インデックスとに対して類似度マッチングを実行するステップであって、前記サブセットが正常にマッチングされた各インデックスによってインデックスされた名前を含み、前記文字列マッチングアルゴリズムが、接頭辞ツリーマッチングアルゴリズム、辞書ツリーマッチングアルゴリズム、文字列類似度マッチングアルゴリズム、および発音類似度マッチングアルゴリズムのうちの1つまたは複数を含む、ステップと、

各サブセットに基づいて前記マッチング対象の名前をマッチングさせるために使用される前記標準名前セットを決定するステップと

を含む、請求項5に記載のコンピュータ可読記憶媒体。

【請求項8】

前記標準名前セット内の前記標準名前に含まれる各単語と、前記マッチング対象の名前に含まれる各単語との間の類似度に基づいて、前記標準名前セット内の前記名前に含まれる各単語を前記マッチング対象の名前に含まれる各単語と、位置合わせされた標準名前セットとして、位置合わせするステップをさらに行わせ、

前記名前が前記少なくとも1つの標準名前と同義であると判定することが、

前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前と同義であるかどうかを判定するために、前記位置合わせされた標準名前セットに基づいて前記マッチング対象の名前を検出するステップを含む、請求項5に記載のコンピュータ可読記憶媒体。

【請求項9】

コンピュータにより実施されるシステムであって、

1つまたは複数のコンピュータと、

前記1つまたは複数のコンピュータに動作可能に接続され、かつ1つまたは複数の命令を記憶したコンピュータ可読記録媒体を備えた1つまたは複数のコンピュータメモリデバイ

10

20

30

40

50

スト

を備え、

前記1つまたは複数の命令は、前記1つまたは複数のコンピュータによって実行されると、前記1つまたは複数のコンピュータに、

マッチング対象の名前を含む入力を受信するステップであって、前記名前が複数の単語を含む、ステップと、

前記名前に対応する第1の名前セットを決定するステップであって、前記第1の名前セットが複数の要素を含む、ステップと、

標準名前セットを生成するために、類似度に基づいて前記複数の単語のそれぞれを前記複数の要素のそれぞれとマッチングさせるステップと、

前記名前の1つまたは複数の名前の文字が標準名前の1つまたは複数の標準名前の文字とは異なるにもかかわらず、前記名前が前記標準名前セット内の少なくとも1つの標準名前と同義であると判定するステップであって、前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前と同義であるかどうかを判定するために、前記マッチング対象の名前に対して、省略形検出、呼称検出、多言語検出、または別名検出のうちの少なくとも1つを実行することを含む、ステップと、

前記マッチング対象の名前のマッチング結果を決定するステップであって、

前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前と同義であるとの判定に回答して、前記少なくとも1つの標準名前を前記マッチング対象の名前の前記マッチング結果として決定するステップと、

前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前と同義でないとの判定に回答して、1つまたは複数の類似度アルゴリズムを使用して前記名前を前記標準名前とマッチングさせることによって前記名前の前記マッチング結果を決定するステップであって、前記名前のマッチング結果が、前記複数の単語のそれぞれの複数のマッチング度の重み付き加算を含み、前記複数のマッチング度が、テキストマッチング度、音声マッチング度、および文字列マッチング度を含む、ステップと

を含む、ステップと

を行わせる、システム。

【請求項10】

マッチングさせる必要がない名前の所定のセットを読み出すステップと、

前記マッチング対象の名前が名前の前記所定のセット内に含まれないと判定するステップと

をさらに行わせる、請求項9に記載のシステム。

【請求項11】

前記複数の単語のそれぞれをマッチングするステップが、

前記第1の名前セットの各名前のインデックスを決定するステップであって、前記インデックスが前記名前に関連付けられた予め確立された単語を含む、ステップと、

前記マッチング対象の名前に含まれる各単語を取得するために前記マッチング対象の名前を分割するステップと、

前記第1の名前セットのサブセットを取得するために、文字列マッチングアルゴリズムを使用することによって前記マッチング対象の名前に含まれる各単語と、各インデックスとに対して類似度マッチングを実行するステップであって、前記サブセットが正常にマッチングされた各インデックスによってインデックスされた名前を含み、前記文字列マッチングアルゴリズムが、接頭辞ツリーマッチングアルゴリズム、辞書ツリーマッチングアルゴリズム、文字列類似度マッチングアルゴリズム、および発音類似度マッチングアルゴリズムのうちの1つまたは複数を含む、ステップと、

各サブセットに基づいて前記マッチング対象の名前をマッチングさせるために使用される前記標準名前セットを決定するステップと

を含む、請求項9に記載のシステム。

【請求項12】

前記標準名前セット内の前記標準名前に含まれる各単語と、前記マッチング対象の名前に含まれる各単語との間の類似度に基づいて、前記標準名前セット内の前記名前に含まれる各単語を前記マッチング対象の名前に含まれる各単語と、位置合わせされた標準名前セットとして、位置合わせするステップをさらに行わせ、

前記名前が前記少なくとも1つの標準名前と同義であると判定することが、

前記マッチング対象の名前が前記標準名前セット内の前記少なくとも1つの標準名前と同義であるかどうかを判定するために、前記位置合わせされた標準名前セットに基づいて前記マッチング対象の名前を検出するステップを含む、請求項9に記載のシステム。

【発明の詳細な説明】

【技術分野】

10

【0001】

本出願は、コンピュータソフトウェア技術の分野に関し、詳細には、名前をマッチングさせるための方法および装置に関する。

【背景技術】

【0002】

人名マッチングは、リスクコントロール分野において非常に重要な技術である。例えば、リスクコントロールシステムは、決定された無許可ユーザの人名をブラックリストに記録する。次いで、リスクコントロール動作を実行するとき、現在サービスを実行している各ユーザについて、各ユーザの人名を、走査してブラックリスト内の人名とマッチングさせる。マッチングが成功した場合、ユーザは、無許可ユーザとみなすことができ、ユーザのサービスは、特定のリスクを防止するために拒否される。

20

【0003】

人名マッチングは、正確な人名マッチングと、人名のファジーマッチングとに分類することができる。対照的に、人名のファジーマッチングは、適切なファジー度を制御するのが困難であるので、技術の観点からより困難である。

【0004】

既存の技術では、人名のファジーマッチングを実行するために、通常、文字列マッチングアルゴリズムが使用され、文字列マッチング度閾値がファジー度を決定する。しかしながら、文字列マッチング度閾値は、すべて経験に従って設定される。脱落を低減するために、文字列マッチング度閾値は、通常、比較的低い値に設定される。その結果、マッチング精度は、比較的低く、リスクコントロールシステムの誤警報率は、比較的高い。

30

【発明の概要】

【発明が解決しようとする課題】

【0005】

本出願の実装形態は、以下の技術的問題を軽減するために、名前をマッチングさせるための方法および装置を提供する。既存の技術では、比較的低いマッチング精度および比較的高いシステム誤警報率は、文字列マッチングアルゴリズムを使用することによって人名のファジーマッチングが実行されるときに起こる。

【課題を解決するための手段】

【0006】

40

前述の技術的問題を軽減するために、本出願の実装形態は、以下のように実施される。

【0007】

本出願の一実装形態は、以下の、マッチング対象の名前を取得するステップと、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定するステップと、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために、マッチング対象の名前を検出するステップと、検出結果に基づいて、マッチング対象の名前のマッチング結果を決定するステップとを含む、名前をマッチングさせるための方法を提供する。

【0008】

本出願の一実装形態は、以下の、マッチング対象の名前を取得するように構成される取

50

得モジュールと、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定するように構成される決定モジュールと、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために、マッチング対象の名前を検出するように構成される検出モジュールと、検出結果に基づいて、マッチング対象の名前のマッチング結果を決定するように構成されるマッチングモジュールとを含む、名前をマッチングさせるための装置を提供する。

【0009】

本出願の実装形態において使用される少なくとも1つの技術的解決策は、以下の有益な効果を達成することができる。名前は、人名を含むことができる。実際には、マッチング対象の人名は、データ入力時の適時性および不確実性、ならびに人名の多様性のために、実際の人名とは異なる場合がある。これも、ファジーマッチングを実行する理由である。本出願の解決策では、この理由のために、マッチング対象の人名は、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために検出され、検出結果に基づいて人名マッチング結果が決定される。ファジー度が経験に従って設定された文字列マッチング度閾値を使用することのみによって制御される既存の技術と比較すると、本出願は、ファジー度を制御する精度を改善するのにより助けになる。そのため、マッチング精度は、改善され得、リスクコントロールシステムの誤警報率が低減され得る。したがって、既存の技術における問題のうちのいくつかまたはすべてが軽減され得る。

【0010】

本出願の実装形態または既存の技術における技術的解決策をより明確に説明するために、以下は、実装形態または既存の技術を説明するために必要とされる添付図面を簡単に説明する。明らかに、以下の説明における添付図面は、本出願のいくつかの実装形態を単に示し、当業者は、創造的な努力なしにこれらの添付図面から他の図面を依然として導出することができる。

【図面の簡単な説明】

【0011】

【図1】本出願の一実装形態による、名前をマッチングさせるための方法を示す概略フローチャートである。

【図2】本出願の一実装形態による、実際の適用シナリオにおいて名前をマッチングさせるための方法における予備選別の実装形態を示す概略フローチャートである。

【図3】本出願の一実装形態による、実際の適用シナリオにおいて名前をマッチングさせるための方法における統合アルゴリズムを使用することによってファジーマッチングを実行することの実装形態を示す概略フローチャートである。

【図4】本出願の一実装形態による、実際の適用シナリオにおいて名前をマッチングさせるための方法の実装形態を示す概略フローチャートである。

【図5】本出願の一実装形態による、名前をマッチングさせるための装置を示す、図1に対応する概略構造図である。

【発明を実施するための形態】

【0012】

本出願の実装形態は、名前をマッチングさせるための方法および装置を提供する。

【0013】

当業者に本出願における技術的解決策をよりよく理解させるために、以下は、本出願の実装形態における添付図面を参照して本出願の実装形態における技術的解決策を明確かつ総合的に説明する。明らかに、記載された実装形態は、本出願の実装形態のほんのいくつかであり、すべてではない。創造的な努力なしに本出願の実装形態に基づいて当業者によって得られるすべての他の実装形態は、本出願の保護範囲内に入るものとする。

【0014】

上記で説明したように、実際には、マッチング対象の人名は、データ入力時の適時性および不確実性、ならびに人名の多様性(主に、人名の「形状(すなわち、文字)」の変化を

10

20

30

40

50

示す)のために、実際の人名とは異なる場合がある。理解の容易さのために、英語の人名が例として使用される。英語の人名の一般的なバリエーションタイプおよび例をTable 1(表1)に示す。

【 0 0 1 5 】

【表 1】

Table 1

英語の人名の一般的なバリエーションタイプ	個人名バリエーションタイプの例	10
スペルミス	Jaxson、Jakson、および Jaxon は、もともと同じ人名を表すことが予想される。SMYTH および SMITH は、もともと同じ人名を表すことが予想される。綴りの誤り。	
表音式綴りの違い	Michel、Michal、および Miguel は、もともと同じ人名を表すことが予想される。類似した発音によって引き起こされるスペルミス。	20
ニックネーム	Mike、Mick、および Mikey は、同じ人のニックネームを表す。	
呼称	Dr.、Mr.などは、人を指す用語であり、それらが人名とともに現れるとき、人名の意味に影響を及ぼさない。	
同義語	Mohamed、Mohammed、および Mohammad は、人名中に現れるとき、同義である。	30
省略形	人名 ROB は、ROBBIN の省略形であり、ROB および ROBBIN は、同義である。人名 BOB は、BOBBY の省略形であり、BOB および BOBBY は、同義である。	
多国籍言語差	ロシア語の BORISOV は、BORISOVNA とは異なる意味を持ち、スペイン語の JUANLTA は、JUAN と同義である。	40

【 0 0 1 6 】

人名をマッチングさせるために文字列マッチング方法のみが使用される場合、特定の文字を「スペルミスすること」の様々なタイプに対してマッチングが試みられ得ることが可能である(実際には、マッチングは偶然に成功し、しっかりとした根拠はない)。しかしながら、マッチング精度は、他のバリエーションタイプに対しては非常に低い。

【 0 0 1 7 】

本出願の解決策では、人名が同義であるが人名の文字が同一ではない(言い換えれば、名前が同義であるが形状において異なる)状況が総合的に考慮され得、さらにマッチング精度が改善され得るように、省略形検出、呼称(address term)検出、多言語検出、または

別名検出のような特定の検出が他のバリエーションタイプのために実行され得る。「異なる形状」は、スペルミスによる「間違い」である異なる形状である可能性があるが、以下の実装形態では、主に他のバリエーションタイプによる「合理的で正しい」異なる形状である可能性があることに留意することに価値がある。

【0018】

本出願の解決策は、人名マッチングに適用可能であるだけでなく、人名以外の名前、例えば、地名または物体名のマッチングにも適用可能である。

【0019】

本出願の解決策について以下に詳細に説明する。

【0020】

図1は、本出願の一実装形態による、名前をマッチングさせるための方法を示す概略フローチャートである。手順を実行するプログラムがインストールされ得るデバイスは、限定はしないが、パーソナルコンピュータ、大型または中型コンピュータ、コンピュータクラスタ、モバイルフォン、タブレットコンピュータ、スマートウェアラブルデバイス、車両機械などを含む。このプロセスは通常、リスクコントロール分野において使用され得、リスクコントロールシステムまたは関連システムによって実行される。

【0021】

図1に示す手順は、以下のステップを含むことができる。

【0022】

S101. マッチング対象の名前を取得する。

【0023】

本出願の本実装形態では、マッチング対象の名前が属する言語は、限定されず、英語、ロシア語、スペイン語、または中国語であり得る。説明の容易さのために、以下の実装形態は、マッチング対象の名前が属する言語が英語である例を使用することによって主に説明される。

【0024】

S102. マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定する。

【0025】

本出願の本実装形態では、標準名前セットは、より大きい名前セットから選別されるサブセットであり得、または、直接より大きい名前セットであり得る。説明の容易さのために、ここでの選別は、「予備選別」と呼ばれ得る。例えば、背景でのシナリオでは、より大きい名前セットは、リスクコントロールシステムによって保持されるブラックリストであり得る。

【0026】

前の状況では、サブセットは、マッチング対象の名前と類似の名前のみを含むセットであり得る。ここでの「類似の」は、類似性をさらに決定する一連の後続の動作が存在するので、それほど厳密でなくてもよい。その後「精選別」が実行されることを意味する。その後の精選別の作業量が低減され得、精選別の適切性が改善され得るように、予備選別を通してマッチング範囲が迅速に低減され得、それは、本出願の解決策の効率を改善することにつながる。

【0027】

S103. マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために、マッチング対象の名前を検出する。

【0028】

本出願の本実装形態では、検出され決定される必要がある「名前が同義であるが、名前の文字が同一ではない」は、主にTable 1(表1)内の1つまたは複数のバリエーションタイプによって引き起こされ、検出は、以下で詳細に説明する、省略形検出、呼称検出、多言語検出、または別名検出のような検出のうちの少なくとも1つを含むことができる。

10

20

30

40

50

【 0 0 2 9 】

本出願の本実装形態では、複数の検出が存在するとき、複数の検出は、順番に連続して実行され得る。マッチング対象の名前のマッチング結果が検出プロセスにおいて決定され得る場合、残りの検出が実行されてなくてもよい。確かに、実行効率を改善するために、複数の検出は、並列して実行され得、次いで、検出結果が要約され得る。

【 0 0 3 0 】

S104. 検出結果に基づいて、マッチング対象の名前のマッチング結果を決定する。

【 0 0 3 1 】

本出願の本実装形態ではステップS103を実行することによって、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないことが判定された場合、マッチング対象の名前のマッチング結果は、直接決定され得る。この状況では、ステップS103における検出プロセスは、実際には、マッチング対象の名前のマッチングプロセス全体である。

【 0 0 3 2 】

マッチング対象の名前が、以下の条件、すなわち、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないことを満たさない場合、マッチング対象の名前のマッチング結果を決定するために、別のマッチング方法を使用することによって、マッチング対象の名前に対してマッチングがさらに実行され得る。

【 0 0 3 3 】

図1に示す方法を通して、名前は、人名を含むことができる。実際には、マッチング対象の人名は、データ入力時の適時性および不確実性、ならびに人名の多様性のために、実際の人名とは異なる場合がある。これも、ファジーマッチングを実行する理由である。本出願の解決策では、この理由のために、マッチング対象の人名は、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために検出され、検出結果に基づいて人名マッチング結果が決定される。ファジー度が経験に従って設定された文字列マッチング度閾値を使用することのみによって制御される既存の技術と比較すると、本出願は、ファジー度を制御する精度を改善するのにより助けになる。そのため、マッチング精度は、改善され得、リスクコントロールシステムの誤警報率が低減され得る。したがって、既存の技術における問題のうちのいくつかまたはすべてが軽減され得る。

【 0 0 3 4 】

図1における方法に基づいて、本出願の本実装形態は、以下に説明する方法および拡張解決策のいくつかの実装形態をさらに提供する。

【 0 0 3 5 】

本出願の本実装形態において、マッチング対象の異なる名前の複雑さは、異なり得、含まれる情報も異なり得る。少ない量の情報または単純な情報特徴を有するマッチング対象のいくつかの名前について、マッチングがマッチング対象の名前に対して実行される場合であっても、得られるマッチング結果の値が期待に達することは困難である。例えば、名前が、英語の人名「Jim」、「Jimmy」、「David」、「John」、または「Mike」のような単純で一般的すぎる場合、マッチングが成功しても、特定の人を決定するのは困難である。

【 0 0 3 6 】

この状況による名前マッチングのために使用される処理リソースの浪費を軽減するために、マッチング対象の名前は、マッチングを続けるかどうかを判定するために、取得された後に最初にフィルタリングされ得る。加えて、マッチング対象の名前がホワイトリストに含まれている場合、同様の問題が存在し、マッチング対象の名前はまた、この方法を使用することによって処理され得る。

【 0 0 3 7 】

ステップS101およびS102について、マッチング対象の名前が取得された後、マッチング

対象の名前をマッチングさせるために使用される標準名前セットが決定される前に、以下のステップ、すなわち、マッチングされる必要がない名前の所定のセットを取得するステップと、マッチング対象の名前がマッチングされる必要がない名前のセット内に含まれるかどうかを判定するステップと、そうである場合、後続のステップを実行し続けるステップとがさらに実行され得る。そうでなければ、マッチング対象の名前に対してマッチングが実行されなくてもよい。

【0038】

本出願の本実装形態では、(「予備選別」の以前の状況に対応する)ステップS102の実施は、説明のための例として使用される。ステップS102について、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定することは、以下、すなわち、

10

【0039】

類似度マッチングを実行する方法の複数の実装形態も存在する。単語分割マッチングがマッチング対象の名前に対して実行され得、またはフルテキストマッチングがマッチング対象の名前に対して実行され得る。

【0040】

20

単語分割マッチングの一例では、マッチング対象の名前に含まれる各単語と、第1の名前セットに含まれる各単語とに対して類似度マッチングを実行することによって、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定することは、以下、すなわち、第1の名前セットに含まれる各名前のインデックスを取得することであって、名前のインデックスが名前に含まれる任意の単語である、取得することと、マッチング対象の名前に含まれる各単語を取得するためにマッチング対象の名前を分割することと、第1の名前セットのサブセットを取得するためにマッチング対象の名前に含まれる各単語と各インデックスとに対して類似度マッチングを実行することとを含むことができる。取得されるサブセットは、正常にマッチングされた各インデックスによってインデックスされた名前を含み、各サブセットに基づいて、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定する。

30

【0041】

前述の例におけるインデックスは、予め確立され、インデックスに基づいて単語分割マッチングを実行することの利点は、マッチングプロセスにおいてセット内に必要な名前を取得する速度が効果的に加速され得ることである。単語分割マッチングがインデックスに基づいて実行されない場合、効率が影響を受ける可能性があることを除いて、単語分割マッチングが依然として実施され得る(例えば、セットを格納するデータテーブルが、単語分割マッチングのためのSelect文を使用することによって、必要とされる名前について直接問い合わせられる)。

【0042】

40

さらに、マッチング対象の名前に含まれる各単語と、各インデックスとに対して類似度マッチングを実行することは、以下、すなわち、文字列マッチングアルゴリズムを使用することによって、マッチング対象の名前に含まれる各単語と、各インデックスとに対して類似度マッチングを実行することを含むことができ、文字列マッチングアルゴリズムは、1つまたは複数のアルゴリズム、例えば、接頭辞ツリーマッチングアルゴリズム、辞書ツリーマッチングアルゴリズム、文字列類似度マッチングアルゴリズム、および発音類似度マッチングアルゴリズムを含むことができる。ここで、文字列マッチングアルゴリズムを使用することが単によりよい方法であるか、または、類似度マッチングを実施するために使用され得るテキストマッチングアルゴリズムのような別のアルゴリズムが使用され得る。

50

【 0 0 4 3 】

各サブセットに基づいてマッチング対象の名前をマッチングさせるために使用される標準名前セットを決定する複数の実装形態が存在し得る。例えば、Nの文字列マッチングアルゴリズムが使用される場合、単語分割を通して取得されるMの単語の各々について、各文字列マッチングアルゴリズムが、Nのサブセットを対応して取得するために単語を各インデックスとマッチングさせるために使用される。次いで、Nのサブセットの和集合が取られる。Mの和集合の合計が取得され、Mの和集合における発生総数が指定された閾値を超える各名前が選択され、標準名前セットとして決定される。別の例について、Nのサブセットが取得され、次いで、Nのサブセットの共通部分が取られ、標準名前セットとして決定される。

10

【 0 0 4 4 】

「予備選別」プロセスについて上記で説明した。理解の容易さのために、本出願の実装形態は、図2に示すように、実際の適用シナリオにおいて名前をマッチングさせるための方法における予備選別の実装形態を示す概略フローチャートをさらに提供する。

【 0 0 4 5 】

実際の適用シナリオにおいて、第1の名前セットが英語の人名のリストであり、リスト内の各名前のインデックスが予め決定され(対応する名前以外の関連情報がインデックスを使用することによってインデックスされ得る)、インデックスがインデックスに対応する名前内の単語であり、各インデックスがインデックスに対応する単語を一次キーとして使用することによって確立されるインデックステーブル内に含まれ、マッチング対象の名前が英語の人名「Kit Wai Jackson Wong」であり、人名を分割するための区切り文字としてスペースが使用されると仮定し、単語分割結果がTable 2(表2)に示されている。

20

【 0 0 4 6 】

【表 2】

Table 2

マッチング対象 の名前	Kit Wai Jackson Wong			
単語分割後	単語 1:Kit	単語 2:Wai	単語 3:Jackson	単語 4:Wong

30

【 0 0 4 7 】

図2において、M=4であり、単語分割結果は、{Kit, Wai, Jackson, Wong}である。4つのマッチングアルゴリズム、すなわち、マッチング結果1、2、3、および4を使用することによって取得される単一の単語のインデックスに対応する名前セットを出力するために、単語分割を通して取得される各単語は、接頭辞ツリーマッチングアルゴリズム、辞書ツリーマッチングアルゴリズム、文字列類似度マッチングアルゴリズム(Simstringアルゴリズムなど)、および発音類似度マッチングアルゴリズム(metaphoneアルゴリズムなど)を使用することによって、各インデックスとマッチングされる。

40

【 0 0 4 8 】

各単語の総合的なマッチング結果を取得するために、各単語のマッチング結果の和集合が取られる。

【 0 0 4 9 】

少なくとも2つの総合的なマッチング結果に含まれる名前がセットを形成するために選択され、セットは、生成された予備選別結果として使用される。

【 0 0 5 0 】

本出願の本実装形態では、標準名前セットが決定された後、マッチング対象の名前は、

50

検出され得る。しかしながら、検出の前に1つまたは複数のタイプの前処理がさらに実行され得る。前処理は、その後の検出結果の信頼性を改善するのに役立つ。前処理は、位置合わせ処理、大文字および小文字の統一処理、簡略化された伝統的な処理などを含むことができる。

【0051】

一例として位置合わせ処理が使用される。ステップS103について、マッチング対象の名前を検出する前に、以下の処理、すなわち、標準名前セット内の名前に含まれる各単語と、マッチング対象の名前に含まれる各単語との間の類似度に基づいて、標準名前セット内の名前に含まれる各単語をマッチング対象の名前に含まれる各単語と位置合わせすることがさらに実行され得る。

10

【0052】

実際には、類似度に基づいて位置合わせ処理を実行することの原理は、類似度ベースの最大化原理とすることができる。具体的には、最大の類似度を有する単語を含む名前の位置が位置合わせされる。

【0053】

例えば、マッチング対象の名前が「Kate Lee Smith」であり、標準名前セット内の「Smith Catherine Lee」のような名前の位置合わせ結果がTable 3(表3)に示されていると仮定する。

【0054】

【表3】

20

Table 3

マッチング対象の名前	Kate	Lee	Smith
標準名前セット内の位置合わせ処理後に得られる名前	Catherine	Lee	Smith

30

【0055】

さらに、位置合わせ処理が実行されると、ステップS103について、マッチング対象の名前を検出することは、以下、すなわち、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために、位置合わせされた標準名前セットに基づいてマッチング対象の名前を検出することを含むことができる。

【0056】

本出願の本実装形態では、ステップS103における検出は、人名のファジーマッチングのマッチング精度を改善するための重要な部分である。ステップS103について、マッチング対象の名前を検出することは、以下、すなわち、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために、マッチング対象の名前に対して、省略形検出、呼称検出、多言語検出、または別名検出のうちの少なくとも1つを実行することを含む。いくつかの検出について、以下に個別に説明する。

40

【0057】

本出願の本実装形態では、頭字語検出が省略形検出において最も一般的である。頭字語検出に加えて、部分詞の省略による省略形検出が存在する。一実装形態では、マッチング対象の名前に対して省略形検出を実行することは、以下、すなわち、所定の省略形対比組合せデータを取得することであって、各省略形対比組合せが、少なくとも1つの単語と、単語の省略形との間の省略形マッピング関係を反映する、取得することと、省略形対比組

50

合せデータに基づいて、マッチング対象の名前に含まれる単語が標準名前セット内の名前に含まれる単語との省略形マッピング関係を有するかどうかを検出することと、検出結果に基づいて、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定することとを含むことができる。

【0058】

例えば、英語の人名「Ben Williams」が「B. Williams」と略される場合があると前もって決められている場合、「Ben Williams」および「B. Williams」は、省略形対比組合せとして使用され得る。マッチング対象の名前および標準名前セット内の名前が省略形対比組合せであることが検出された場合、名前が同義であるが、名前の文字が同一ではない状況が存在すると判定され得る。

10

【0059】

本出願の本実装形態では、マッチング対象の名前に対して呼称検出を実行することは、以下の、すなわち、所定の呼称データを取得することと、マッチング対象の名前が呼称を含むかどうかを検出することとであって、呼称が、呼称を含む名前の意味に影響を及ぼさないとみなされる、検出することと、検出結果に基づいて、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定することとを含むことができる。

【0060】

人名に関して、呼称は、一般に、尊称(Mr.またはMiss.など)または肩書き(Dr.またはProf.など)のような、元の名前の少なくとも一部に添えられる単語である。人名マッチング環境では、呼称は、呼称に対応する元の名前の少なくとも一部の意味に影響を及ぼさない。したがって、マッチング対象の名前が呼称を含み、呼称以外の他の部分が標準名前セット内の名前と一致する可能性があるとして判定された場合、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないと判定され得る。

20

【0061】

本出願の本実装形態では、マッチング対象の名前に対して別名検出を実行することは、以下、すなわち、所定の別名対比組合せデータを取得することと、各別名対比組合せが、少なくとも1つの単語と、単語の別名との間の別名マッピング関係を反映する、取得することと、別名対比組合せデータに基づいて、マッチング対象の名前に含まれる単語が、標準名前セット内の名前に含まれる単語との別名マッピング関係を有するかどうかを検出することと、検出結果に基づいて、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定することとを含むことができる。

30

【0062】

実際には、別名は、別名に対応する名前のニックネーム(例えば、Table 1(表1)中のMickは、Mikeyのニックネーム)、または別の分野における別名に対応する名前の同義語名を含むことができる。後者について、異なる分野は、異なる地域(例えば、異なる国または異なる州)、異なる言語(例えば、異なる国の言語または異なる国籍の言語)、異なる産業などであり得る。

40

【0063】

対応して、以下のうちの少なくとも1つが実行され、すなわち、マッチング対象の名前のニックネームが検出されるか、または異なる分野におけるマッチング対象の名前の同義語名が検出される。

【0064】

別名は、別名に対応する名前の意味に影響を及ぼさない。したがって、マッチング対象の名前が標準名前セット内の名前の別名であると判定された場合、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないと判定され得る。

【0065】

50

上記で説明した対比組合せデータおよび呼称データの記憶形態は、本出願では限定されないことに留意することに価値がある。一般的な方法は、対比組合せデータおよび呼称データを対応するデータテーブル内に記憶し、データが使用される必要があるとき、データベースからデータを読み取ることである。

【0066】

本出願の本実装形態では、マッチング対象の名前に対して多言語検出を実行することは、以下、すなわち、マッチング対象の名前に対応する言語を決定することと、言語または別の言語のうちの少なくとも1つの綴り変形同義語規則または綴り変形同音異義語規則のうちの少なくとも1つを取得することと、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために、綴り変形同義語規則または綴り変形同音異義語規則のうちの少なくとも1つに従ってマッチング対象の名前を検出することを含むことができる。

10

【0067】

多言語検出は、以下の状況、例えば、英語の「Pooh」がドイツ語では「Puh」と綴られ、これらの2つの単語が人名において現れるときには同義である状況を主に目的とする。

【0068】

文字列マッチング度を計算するために使用されるアルゴリズムが以前に実行された検出において使用され得ることに留意することに価値がある。

【0069】

本出願の本実装形態では、検出結果に基づいてマッチング対象の名前のマッチング結果を決定することは、以下、すなわち、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないと判定された場合、その少なくとも1つの名前をマッチング対象の名前のマッチング結果として決定すること、または、そうでない場合(具体的には、前の検出を通してマッチングが失敗したとき)、1つまたは複数の類似度アルゴリズムを使用することによってマッチング対象の名前を標準名前セット内の名前とマッチングさせることによってマッチング対象の名前のマッチング結果を決定することを含むことができる。

20

【0070】

本出願の本実装形態では、複数の類似度マッチングは、異なる次元に基づくことができ、したがって、マッチング結果の信頼性を改善するのに役立つことができる。そのような着想によれば、類似度アルゴリズムは、テキストマッチング度を計算するために使用されるアルゴリズム(nグラムアルゴリズムなど)、音声マッチング度を計算するために使用されるアルゴリズム(Phonexアルゴリズムなど)、文字列マッチング度を計算するために使用されるアルゴリズム(Jaro-Winklerアルゴリズムなど)などであり得る。

30

【0071】

複数の類似度アルゴリズムが使用されるとき、総合的なマッチング結果を得るために、類似度アルゴリズムに対応するマッチング結果が総合的に測定され得る。測定方法は、本出願では限定されず、一般的な方法は、重み付け加算である。

【0072】

例えば、nグラムアルゴリズムが使用されるとき、アルゴリズム入力は、マッチング対象の名前に含まれる各単語、および単語の位置合わせされた位置における各単語であり、アルゴリズム出力は、位置合わせされた単語の各対のテキストマッチング度であり、F1と表記される。

40

【0073】

Jaro-Winklerアルゴリズムが使用されるとき、アルゴリズム入力は、マッチング対象の名前に含まれる単語、および単語の位置合わせされた位置における各単語であり、アルゴリズム出力は、位置合わせされた単語の各対の文字列マッチング度であり、F2と表記される。

【0074】

Phonexアルゴリズムが使用されるとき、アルゴリズム入力は、マッチング対象の名前に

50

含まれる単語、および単語の位置合わせされた位置における各単語であり、アルゴリズム出力は、位置合わせされた単語の各対の音声マッチング度であり、F3と表記される。

【 0 0 7 5 】

以下の式、

$$F=w1*F1+w2*F2+w3*F3、ここで、w1+w2+w3=1$$

において示されるように、テキストマッチング度、文字列マッチング度、および音声マッチング度に対して重み付け加算を実行することによって、位置合わせされた単語の各対の総合的なマッチング度Fが得られる。

【 0 0 7 6 】

単語の各対の総合的なマッチング度Fに基づいて、平均値を計算することによって、マッチング対象の名前と、標準名前セット内の名前との間のマッチングの結果が得られる。例えば、Table 3(表3)中の名前の対について、得られたマッチング結果がTable 4(表4)に示されている。

【 0 0 7 7 】

【表 4】

Table 4

検出された名前	Kate	Lee	Smith
リスト内の名前	Catherine	Lee	Smith
総合的なマッチング度	0.872	1	1
マッチング結果	$(0.872+1+1)/3=0.957$		

【 0 0 7 8 】

予備選別後の検出およびマッチングプロセスについては、上記で説明した。このプロセスでは、複数のマッチング関連アルゴリズムが使用され得る。本出願の解決策の実施中、使用され得るアルゴリズムは、統合され得、このプロセスは、統合されたアルゴリズムを使用することによってファジーマッチングを実行するプロセスである。

【 0 0 7 9 】

本出願の本実装形態では、統合アルゴリズムを使用することによってファジーマッチングが実行された後、特定の規則に基づくなんらかの事後フィルタリングがさらに実行され得、例えば、マッチング結果のマッチング度は、テキスト記述情報にマッピングされるか、またはマッチング度は、特定のシナリオに基づいて適切に補償もしくは低減される。

【 0 0 8 0 】

前述の説明に基づいて、より直感的に、本出願の一実装形態は、図3に示すように、実際の適用シナリオにおいて名前をマッチングさせるための方法において統合アルゴリズムを使用することによってファジーマッチングを実行する実装形態を示す概略フローチャートをさらに提供する。図3において、様々な方法において様々な検出を実行し、マッチング度を計算するシーケンスは、単なる一例であり、本出願を限定するものではない。

【 0 0 8 1 】

図3において、マッチング対象の名前のマッチングが前の検出のうちのいずれか1つを介して成功した場合、マッチング結果が直接出力され得、そうでない場合、マッチング対象の名前のマッチング結果は、マッチング度計算(例えば、テキストマッチング度計算、音声マッチング度計算、および文字列マッチング度計算)の1つまたは複数方法を使用することによって決定され、出力され得る。

【0082】

もちろん、様々な方法における様々な検出およびマッチング度計算も完全に実行され得、次いで、様々な方法における様々な検出結果およびマッチング度計算結果が、マッチング対象の文字列のマッチング結果を決定するために総合的に考慮され得る。

【0083】

さらに、本出願の一実装形態は、図4に示すように、実際の適用シナリオにおいて、名前をマッチングさせるための方法の実装形態を示す概略フローチャートをさらに提供する。図4において、特定の規則に基づいて事前決定することは、以下、すなわち、マッチング対象の名前がマッチングされる必要がない名前のセット内に含まれていないかどうかを判定することをさらに含むことができ、ここで、走査リストインデックスが標準名前セット内の名前のインデックスである。

10

【0084】

図3および図4におけるステップについては、上記で詳細に説明しており、簡略化のために、詳細は、ここでは省略される。

【0085】

本出願の本実装形態において提供される名前をマッチングさせるための方法について、上記で説明している。図5に示すように、同じ発明思想に基づいて、本出願の実施形態は、対応する装置をさらに提供する。

【0086】

図5は、本出願の本実装形態による、名前をマッチングさせるための装置を示す、図1に対応する概略構造図である。破線は、オプションのモジュールを表し、装置は、以下、すなわち、マッチング対象の名前を取得するように構成される取得モジュール501と、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定するように構成される決定モジュール502と、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するためにマッチング対象の名前を検出するように構成される検出モジュール503と、検出結果に基づいてマッチング対象の名前のマッチング結果を決定するように構成されるマッチングモジュール504とを含む。

20

【0087】

オプションで、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定する前に、決定モジュール502は、マッチングさせる必要がない名前の所定のセットを取得し、マッチング対象の名前がマッチングさせる必要がない名前のセット内に含まれないと判定するように構成される。

30

【0088】

オプションでは、決定モジュール502がマッチング対象の名前をマッチングさせるために使用される標準名前セットを決定するように構成されることは、以下を含む。

【0089】

決定モジュール502は、マッチング対象の名前をマッチングさせるために使用され得る第1の名前セットを取得し、マッチング対象の名前に含まれる各単語および第1の名前セット内の名前に含まれる各単語に対して類似度マッチングを実行することによって、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定するように構成される。

40

【0090】

オプションで、決定モジュール502が、マッチング対象の名前に含まれる各単語および第1の名前セット内の名前に含まれる各単語に対して類似度マッチングを実行することによって、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定するように構成されることは、以下を含む。

【0091】

決定モジュール502は、第1の名前セット内に含まれる各名前のインデックスを取得することであって、名前のインデックスが名前に含まれる任意の単語である、取得することと

50

、マッチング対象の名前に含まれる各単語を取得するためにマッチング対象の名前を分割することと、第1の名前セットのサブセットを取得するためにマッチング対象の名前に含まれる各単語と、各インデックスとに対して類似度マッチングを実行することを行うように構成される。取得されるサブセットは、正常にマッチングされた各インデックスによってインデックスされた名前を含み、サブセットに基づいて、マッチング対象の名前をマッチングさせるために使用される標準名前セットを決定する。

【0092】

オプションで、決定モジュール502が、マッチング対象の名前に含まれる各単語と、各インデックスとに対して類似度マッチングを実行するように構成されることは、以下を含む。

10

【0093】

決定モジュール502は、文字列マッチングアルゴリズムを使用することによって、マッチング対象の名前に含まれる各単語と、各インデックスとに対して類似度マッチングを実行するように構成され、文字列マッチングアルゴリズムは、以下、すなわち、接頭辞ツリーマッチングアルゴリズム、辞書ツリーマッチングアルゴリズム、文字列類似度マッチングアルゴリズム、および発音類似度マッチングアルゴリズムのうちの少なくとも1つを含む。

【0094】

オプションで、装置は、以下、すなわち、検出モジュール503がマッチング対象の名前を検出する前に、標準名前セット内の名前に含まれる各単語と、マッチング対象の名前に含まれる各単語との間の類似度に基づいて、標準名前セット内の名前に含まれる各単語をマッチング対象の名前に含まれる各単語と位置合わせするように構成される位置合わせモジュール505をさらに含む。

20

【0095】

検出モジュール503がマッチング対象の名前を検出するように構成されることは、以下を含む。

【0096】

検出モジュール503は、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために、位置合わせされた標準名前セットに基づいてマッチング対象の名前を検出するように構成される。

30

【0097】

オプションで、検出モジュール503がマッチング対象の名前を検出するように構成されることは、以下、すなわち、検出モジュール503が、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために、マッチング対象の名前に対して、省略形検出、呼称検出、多言語検出、または別名検出のうちの少なくとも1つを実行するように構成されることを含む。

【0098】

オプションで、検出モジュール503がマッチング対象の名前に対して省略形検出を実行するように構成されることは、以下を含む。

【0099】

40

検出モジュールは、所定の省略形対比組合せデータを取得することであって、各省略形対比組合せが、少なくとも1つの単語と、単語の省略形との間の省略形マッピング関係を反映する、取得することと、省略形対比組合せデータに基づいて、マッチング対象の名前に含まれる単語が標準名前セット内の名前に含まれる単語との省略形マッピング関係を有するかどうかを検出することと、検出結果に基づいて、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定することを行うように構成される。

【0100】

オプションで、検出モジュール503がマッチング対象の名前に対して呼称検出を実行するように構成されることは、以下を含む。

50

【0101】

検出モジュール503は、所定の呼称データを取得することと、マッチング対象の名前が呼称を含むかどうかを検出することと、呼称が、呼称を含む名前の意味に影響を及ぼさないとみなされる、検出することと、検出結果に基づいて、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定することとを行うように構成される。

【0102】

オプションで、別名は、別名に対応する名前のニックネーム、または異なる分野における別名に対応する名前の同義語名を含む。検出モジュール503がマッチング対象の名前に対して別名検出を実行するように構成されることは、以下を含む。

10

【0103】

検出モジュール503は、以下、すなわち、マッチング対象の名前のニックネームを検出すること、または異なる分野におけるマッチング対象の名前の同義語名を検出することのうちの少なくとも1つを実行するように構成される。

【0104】

オプションで、検出モジュール503がマッチング対象の名前に対して多言語検出を実行するように構成されることは、以下、すなわち、検出モジュール503が、マッチング対象の名前に対応する言語を決定し、言語の綴り変形同義語規則または綴り変形同音異義語規則のうちの少なくとも1つを取得し、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定するために、綴り変形同義語規則または綴り変形同音異義語規則のうちの少なくとも1つに従ってマッチング対象の名前を検出するように構成されることを含む。

20

【0105】

オプションで、検出モジュール503がマッチング対象の名前に対して別名検出を実行するように構成されることは、以下を含む。

【0106】

検出モジュール503は、所定の別名対比組合せデータを取得することと、各別名対比組合せが、少なくとも1つの単語と、単語の別名との間の別名マッピング関係を反映する、取得することと、別名対比組合せデータに基づいて、マッチング対象の名前に含まれる単語が、標準名前セット内の名前に含まれる単語との別名マッピング関係を有するかどうかを検出することと、検出結果に基づいて、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないかどうかを判定することとを行うように構成される。

30

【0107】

オプションで、マッチングモジュール504が検出結果に基づいてマッチング対象の名前のマッチング結果を決定するように構成されることは、以下を含む。

【0108】

マッチングモジュール504は、検出モジュール503が、マッチング対象の名前が標準名前セット内の少なくとも1つの名前と同義であるが、名前の文字が同一ではないと判定した場合、マッチング対象の名前のマッチング結果として少なくとも1つの名前を決定し、そうでない場合、1つまたは複数の類似度アルゴリズムを使用することによってマッチング対象の名前を標準名前セット内の名前とマッチングさせることによってマッチング対象の名前のマッチング結果を決定するように構成される。

40

【0109】

オプションで、類似度アルゴリズムは、以下、すなわち、テキストマッチング度を計算するために使用されるアルゴリズム、音声マッチング度を計算するために使用されるアルゴリズム、または文字列マッチング度を計算するために使用されるアルゴリズムのうちの少なくとも1つを含む。

【0110】

オプションで、名前は、人名である。

50

【0111】

本出願の実装形態において提供される装置は、方法と一対一の対応関係にある。したがって、装置および方法は、同様の有益な技術的效果を有する。方法の有益な技術的效果については、上記で詳細に説明しており、したがって、装置の有益な技術的效果は、簡略化のためにここでは省略される。

【0112】

加えて、前述の装置および方法の適用シナリオは、本出願において限定されない。背景で言及したりスクコントロール分野(例えば、資金洗浄対策およびユーザ認証のような分野)に加えて、本出願の解決策は、名前マッチング技法を使用する必要がある任意の他の分野に適用可能である。

10

【0113】

1990年代において、技術的改善がハードウェアの改善(例えば、ダイオード、トランジスタ、またはスイッチのような回路構造に対する改善)であるか、またはソフトウェアの改善(方法手順に対する改善)であるかは、明確に区別され得る。しかしながら、技術が発展するにつれて、多くの方法手順に対する現在の改善は、ハードウェア回路構造に対する直接の改善とみなされ得る。設計者は通常、対応するハードウェア回路構造を得るために、改善された方法手順をハードウェア回路にプログラムする。したがって、方法手順は、ハードウェアエンティティモジュールを使用することによって改善され得る。例えば、プログラマブル論理デバイス(PLD)(例えば、フィールドプログラマブルゲートアレイ(FPGA))がそのような集積回路であり、PLDの論理機能がデバイスプログラミングを通してユーザによって決定される。設計者は、チップ製造業者に特定用途向け集積回路チップを設計および製造することを要求することなく、デジタルシステムをPLDに「統合する」ためにプログラミングを実行する。加えて、現在、集積回路チップを手動で製造する代わりに、そのようなプログラミングは、たいていは「論理コンパイラ」ソフトウェアを使用することによって実施される。論理コンパイラソフトウェアは、プログラムを開発および記述するために使用されるソフトウェアコンパイラと同様である。元のコードは、コンパイルのための特定のプログラミング言語で記述される必要がある。言語は、ハードウェア記述言語(HDL)と呼ばれる。Advanced Boolean Expression Language(ABEL)、Altera Hardware Description Language(AHDL)、Confluence、Cornell University Programming Language(CUP L)、HDCal、Java(登録商標) Hardware Description Language(JHDL)、Lava、Lola、MyHDL、PALASM、およびRuby Hardware Description Language(RHDL)のような多くのHDLが存在する。超高速集積回路ハードウェア記述言語(VHDL)およびVerilogが最も一般的に使用される。当業者は、方法手順がいくつかの記述されたハードウェア記述言語を使用して論理的にプログラムされ、集積回路にプログラムされると、論理的方法手順を実施するハードウェア回路が容易に得られ得ることも理解すべきである。

20

30

【0114】

任意の適切な方法を使用することによってコントローラが実装され得る。例えば、コントローラは、マイクロプロセッサまたはプロセッサ、および、マイクロプロセッサもしくはプロセッサ、論理ゲート、スイッチ、特定用途向け集積回路(ASIC)、プログラマブル論理コントローラ、もしくは内蔵マイクロコントローラによって実行され得る(ソフトウェアまたはファームウェアのような)コンピュータ可読プログラムコードを記憶するコンピュータ可読媒体であり得る。コントローラの例は、限定はしないが、以下のマイクロプロセッサ、すなわち、ARC625D、Atmel AT91SAM、Microchip PIC18F26K20、およびSilicone Labs C8051F320を含む。メモリコントローラはまた、メモリの制御ロジックの一部として実装され得る。当業者は、コンピュータ可読プログラムコードを使用することによってコントローラを実装することに加えて、コントローラが、論理ゲート、スイッチ、特定用途向け集積回路、プログラマブル論理コントローラ、および内蔵マイクロコントローラの形態において同じ機能を実装することを可能にするために、方法ステップに対して論理プログラミングが実行され得ることを知る。したがって、コントローラは、したがって、コントローラは、ハードウェア構成要素とみなされ得、コントローラにおける様々な機能を実

40

50

装するように構成される装置も、ハードウェア構成要素内の構造とみなされ得る。または、様々な機能を実装するように構成される装置は、方法を実施するソフトウェアモジュールと、ハードウェア構成要素内の構造の両方とさえみなされ得る。

【0115】

前述の実装形態において示したシステム、装置、モジュール、またはユニットは、コンピュータチップもしくはエンティティを使用することによって実装され得、または特定の機能を有する製品を使用することによって実装され得る。典型的な実装デバイスは、コンピュータである。コンピュータは、例えば、パーソナルコンピュータ、ラップトップコンピュータ、セルラー電話、カメラ付き携帯電話、スマートフォン、携帯情報端末、メディアプレーヤ、ナビゲーションデバイス、電子メールデバイス、ゲームコンソール、タブレットコンピュータ、ウェアラブルデバイス、またはこれらのデバイスの任意の組合せであり得る。

10

【0116】

説明の容易さのために、上記の装置について、機能を様々なユニットに分割することによって説明した。確かに、本出願が実施されるとき、ユニットの機能は、1つまたは複数のソフトウェアおよび/またはハードウェアにおいて実装され得る。

【0117】

当業者は、本開示の実装形態が方法、システム、またはコンピュータプログラム製品として提供され得ることを理解すべきである。したがって、本開示は、ハードウェアのみの実装形態、ソフトウェアのみの実装形態、またはソフトウェアとハードウェアの組合せを用いる実装形態の形態を使用することができる。さらに、本開示は、コンピュータ使用可能プログラムコードを含む1つまたは複数のコンピュータ使用可能記憶媒体(限定はしないが、ディスクメモリ、CD-ROM、光メモリなどを含む)上に実装されるコンピュータプログラム製品の形態を使用することができる。

20

【0118】

本開示について、本開示の実装形態に基づく方法、デバイス(システム)、およびコンピュータプログラム製品のフローチャートおよび/またはブロック図を参照して説明した。フローチャートおよび/もしくはブロック図内の各プロセスおよび/もしくは各ブロック、ならびにフローチャートおよび/もしくはブロック図内のプロセスおよび/もしくはブロックの組合せを実施するために、コンピュータプログラム命令が使用され得ることに留意することに価値がある。これらのコンピュータプログラム命令は、コンピュータ、または別のプログラム可能データ処理デバイスのプロセッサによって実行される命令がフローチャート内の1つもしくは複数のプロセスおよび/またはブロック図内の1つもしくは複数のブロックにおける指定された機能を実施するためのデバイスを生成するように、機械を生成するために、汎用コンピュータ、専用コンピュータ、内蔵プロセッサ、または別のプログラム可能データ処理デバイスのプロセッサに提供され得る。

30

【0119】

これらのコンピュータプログラム命令は、コンピュータ可読メモリ内に記憶された命令が指示デバイスを含む人工物を生成するように、特定の方法で動作するようにコンピュータまたは別のプログラム可能データ処理デバイスに命令することができるコンピュータ可読メモリ内に記憶され得る。指示デバイスは、フローチャート内の1つもしくは複数のプロセスおよび/またはブロック図内の1つもしくは複数のブロックにおける指定された機能を実施する。

40

【0120】

これらのコンピュータプログラム命令は、一連の動作ならびに動作およびステップがコンピュータまたは別のプログラム可能データ処理デバイス上で実行され、それによってコンピュータ実施処理を生成するように、コンピュータまたは別のプログラム可能データ処理デバイスにロードされ得る。したがって、コンピュータまたは別のプログラム可能データ処理デバイス上で実行される命令は、フローチャート内の1つもしくは複数のプロセスおよび/またはブロック図内の1つもしくは複数のブロックにおける指定された機能を実施

50

するためのステップを提供する。

【0121】

典型的な構成において、コンピューティングデバイスは、1つまたは複数のプロセッサ(CPU)と、1つまたは複数の入力/出力インターフェースと、1つまたは複数のネットワークインターフェースと、1つまたは複数のメモリとを含む。

【0122】

メモリは、非永続的メモリ、ランダムアクセスメモリ(RAM)、不揮発性メモリ、および/またはコンピュータ可読媒体内にある別の形態、例えば、読み出し専用メモリ(ROM)またはフラッシュメモリ(フラッシュRAM)を含むことができる。メモリは、コンピュータ可読媒体の一例である。

10

【0123】

コンピュータ可読媒体は、任意の方法または技術を使って情報を記憶することができる永続的、非永続的、可動、および不動媒体を含む。情報は、コンピュータ可読命令、データ構造、プログラムモジュール、または他のデータであり得る。コンピュータ記憶媒体の例は、限定はしないが、相変化ランダムアクセスメモリ(PRAM)、スタティックランダムアクセスメモリ(SRAM)、ダイナミックランダムアクセスメモリ(DRAM)、別のタイプのランダムアクセスメモリ(RAM)、読み出し専用メモリ(ROM)、電氣的消去可能プログラマブル読み出し専用メモリ(EEPROM)、フラッシュメモリもしくは別のメモリ技術、コンパクトディスク読み出し専用メモリ(CD-ROM)、デジタル多用途ディスク(DVD)もしくは別の光記憶装置、カセット磁気テープ、磁気テープ/磁気ディスク記憶装置もしくは別の磁気記憶デバイス、または任意の他の非伝送媒体を含む。コンピュータ記憶媒体は、コンピューティングデバイスによってアクセス可能な情報を記憶するために使用され得る。本明細書における説明に基づいて、コンピュータ可読媒体は、変調データ信号および搬送波のような一時的コンピュータ可読媒体(一時的媒体)を含まない。

20

【0124】

「含む」、「備える」の用語、またはそれらの任意の他の変形は、非排他的な包含をカバーすることを意図しているので、要素のリストを含むプロセス、方法、製品、またはデバイスが、それらの要素を含むだけでなく、明示的に挙げられていない他の要素も含むか、または、そのようなプロセス、方法、製品、またはデバイスに固有の要素をさらに含むことにさらに留意することに価値がある。より多くの制約なしで、「...を含む」によって先行される要素は、その要素を含むプロセス、方法、製品、またはデバイス内の追加の同一の要素の存在を排除しない。

30

【0125】

当業者は、本出願の実装形態が方法、システム、またはコンピュータプログラム製品として提供され得ることを理解すべきである。したがって、本出願は、ハードウェアのみの実装形態、ソフトウェアのみの実装形態、またはソフトウェアとハードウェアの組合せを用いる実装形態の形態を使用することができる。加えて、本開示は、コンピュータ使用可能プログラムコードを含む1つまたは複数のコンピュータ使用可能記憶媒体(限定はしないが、ディスクメモリ、CD-ROM、光メモリなどを含む)上に実装されるコンピュータプログラム製品の形態を使用することができる。

40

【0126】

本出願について、コンピュータ、例えば、プログラムモジュールによって実行されるコンピュータ実行可能命令の一般的な文脈で説明することができる。一般に、プログラムモジュールは、特定のタスクを実行するか、または特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、構成要素、データ構造などを含む。本出願は、分散型コンピューティング環境においても実施され得る。分散型コンピューティング環境では、通信ネットワークを介して接続されるリモート処理デバイスによってタスクが実行される。分散型コンピューティング環境では、プログラムモジュールは、記憶デバイスを含むローカルとリモートの両方のコンピュータ記憶媒体内に配置され得る。

【0127】

50

本明細書の実装形態について、漸進的な方法で説明している。実装形態の同じまたは類似の部分について、実装形態に参照が行われ得る。各実装形態は、他の実装形態との違いに焦点を当てる。具体的には、システムの実装形態は、方法の実装形態と基本的に類似しており、したがって、簡単に説明される。関連部分について、方法の実装形態における関連説明に参照がなされ得る。

【 0 1 2 8 】

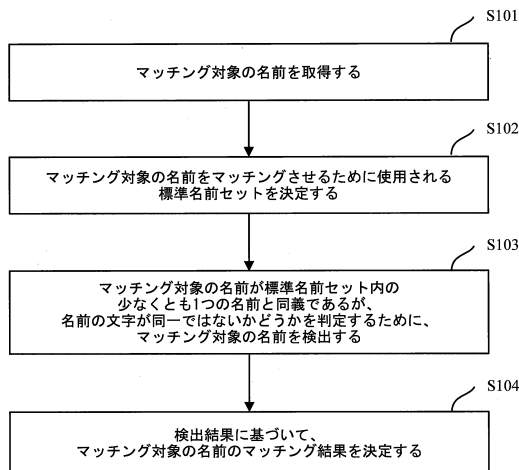
前述の説明は、本出願の実装形態であり、本出願を限定することを意図していない。当業者は、本出願に対して様々な修正および変更をすることができる。本出願の要旨および原理から逸脱することなくなされるいかなる修正、同等の置き換え、または改善も、本出願の特許請求の範囲内に入るものとする。

【 符号の説明 】

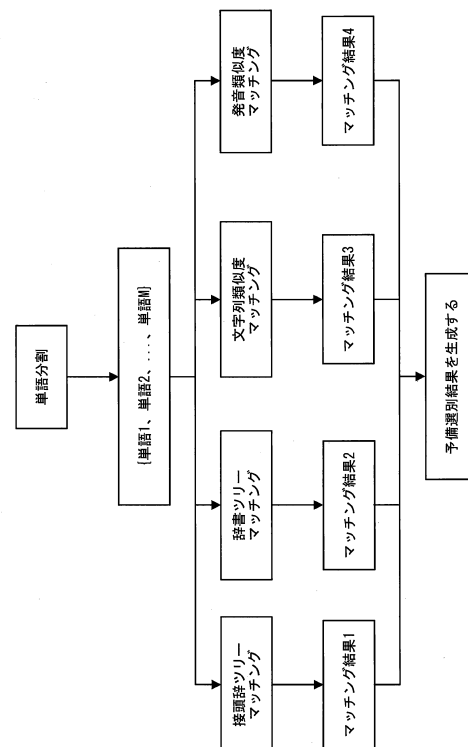
【 0 1 2 9 】

- 501 取得モジュール
- 502 決定モジュール
- 503 検出モジュール
- 504 マッチングモジュール
- 505 位置合わせモジュール

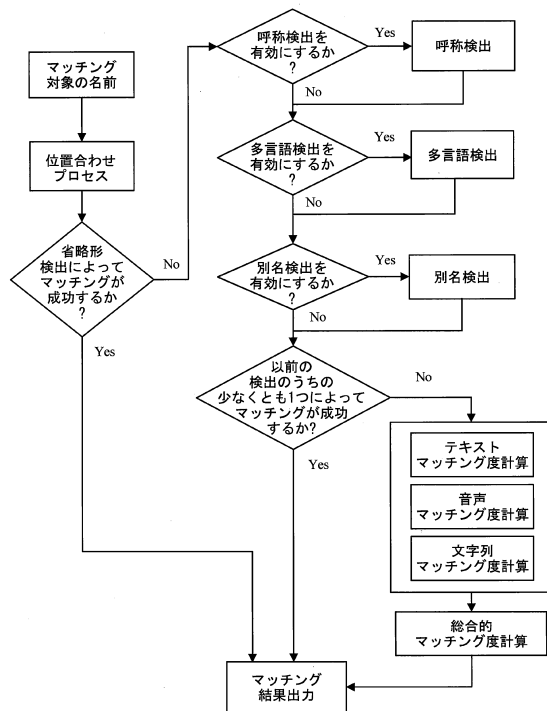
【 図 1 】



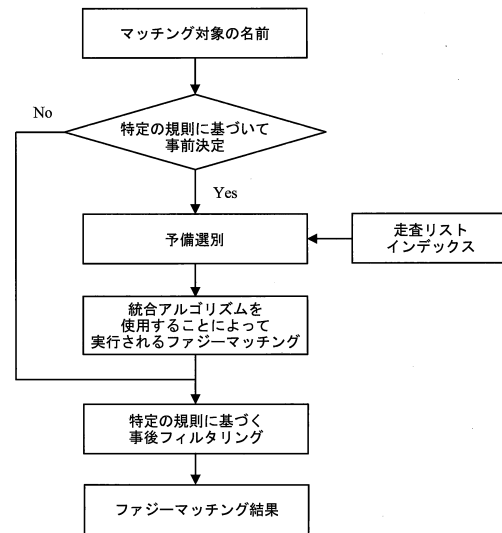
【 図 2 】



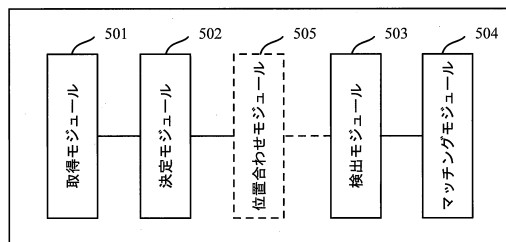
【図 3】



【図 4】



【図 5】



フロントページの続き

(72)発明者 チンチン・ソン

中華人民共和国・311121・ゼジャン・ハンジョウ・ユ・ハン・ディストリクト・ウェスト・
ウェン・イ・ロード・ナンバー・969・ビルディング・3・5 / エフ・アリババ・グループ・リ
ーガル・デパートメント

審査官 甲斐 哲雄

(56)参考文献 特表2010-519655(JP, A)

米国特許出願公開第2014/0222416(US, A1)

米国特許出願公開第2004/0024760(US, A1)

米国特許出願公開第2008/0091674(US, A1)

(58)調査した分野(Int.Cl., DB名)

G06F 16/00 - 16/958