

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2009-128968  
(P2009-128968A)

(43) 公開日 平成21年6月11日(2009.6.11)

(51) Int.Cl. F I テーマコード (参考)  
**G06F 17/28 (2006.01)** G O 6 F 17/28 U 5 B O 9 1  
**G06Q 50/00 (2006.01)** G O 6 F 17/60 1 2 6 K

審査請求 未請求 請求項の数 8 O L (全 10 頁)

(21) 出願番号 特願2007-300178 (P2007-300178)  
 (22) 出願日 平成19年11月20日(2007.11.20)

(71) 出願人 000005496  
 富士ゼロックス株式会社  
 東京都港区赤坂九丁目7番3号  
 (74) 代理人 100114937  
 弁理士 松本 裕幸  
 (72) 発明者 鷹台 基行  
 神奈川県足柄上郡中井町境430 グリー  
 ンテクなかい 富士ゼロックス株式会社内  
 (72) 発明者 増市 博  
 神奈川県足柄上郡中井町境430 グリー  
 ンテクなかい 富士ゼロックス株式会社内  
 (72) 発明者 三浦 康秀  
 神奈川県足柄上郡中井町境430 グリー  
 ンテクなかい 富士ゼロックス株式会社内  
 Fターム(参考) 5B091 AA15 AB17 CA02 CA05 CC04  
 CC16

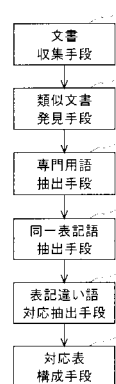
(54) 【発明の名称】 表記ゆれ解析装置

(57) 【要約】

【課題】 表記が異なるが同じ概念を表す同義語と推定される単語の組を効率よく抽出する。

【解決手段】 類似文書発見手段2が、文書収集手段1に記憶されている電子カルテのうち記述内容が類似するものを選出し、専門用語抽出手段3が、前記選出された類似の電子カルテの記述に出現する診療に関する単語を抽出し、同一表記語抽出手段4が、前記抽出された各単語のうち、各電子カルテに共通して出現する共通語を特定する。そして、表記違い語対応抽出手段5が、前記抽出された各単語のうちの共通語以外で且つ異なる電子カルテから抽出された単語の組を対象に、各単語の概念が同一であるか(同義語であるか)を判定し、対応表構成手段6が、同義語と判定された単語の組を対応表形式で出力する。

【選択図】 図1



## 【特許請求の範囲】

## 【請求項 1】

診療結果を記述した電子的な文書を複数記憶する記憶手段と、  
記述内容が類似する複数の文書を選出する選出手段と、  
文書中の記述に出現する診療に関する単語を抽出する抽出手段と、  
前記選出した各文書から抽出した単語のうち各文書に共通して出現する単語を特定する  
特定手段と、

前記選出した各文書から抽出した単語のうち前記特定した単語以外で且つ異なる文書から  
抽出した単語の組を対象に、単語の概念の同一性に係る所定の基準を満たすか否かを判  
定する判定手段と、

所定の基準を満たすと判定された単語の組を出力する出力手段と、  
を備えたことを特徴とする表記ゆれ解析装置。

## 【請求項 2】

前記文書には、その作成主体を示す作成主体情報が付加されており、

前記判定手段は、前記対象の各単語を抽出した文書に付加された作成主体情報が異なる  
場合に所定の基準を満たすと判定することを特徴とする請求項 1 に記載の表記ゆれ解析装  
置。

## 【請求項 3】

前記判定手段は、前記対象の各単語の表記が類似する場合に所定の基準を満たすと判定  
することを特徴とする請求項 1 又は請求項 2 に記載の表記ゆれ解析装置。

## 【請求項 4】

前記判定手段は、前記対象の各単語に係る構文が共通性を有する場合に所定の基準を満  
たすと判定することを特徴とする請求項 1 乃至請求項 3 のいずれか 1 項に記載の表記ゆれ  
解析装置。

## 【請求項 5】

前記判定手段は、前記対象の各単語を抽出した文書における該単語の出現位置が共通性  
を有する場合に所定の基準を満たすと判定することを特徴とする請求項 1 乃至請求項 4 の  
いずれか 1 項に記載の表記ゆれ解析装置。

## 【請求項 6】

前記判定手段は、前記対象の各単語を抽出した文書に診断結果として記述された病名又  
はその識別コードが共通する場合に所定の基準を満たすと判定することを特徴とする請求  
項 1 乃至請求項 5 のいずれか 1 項に記載の表記ゆれ解析装置。

## 【請求項 7】

前記判定手段は、前記対象の各単語を抽出した文書の作成時期が共通性を有する場合に  
所定の基準を満たすと判定することを特徴とする請求項 1 乃至請求項 6 のいずれか 1 項に  
記載の表記ゆれ解析装置。

## 【請求項 8】

コンピュータを、

診療結果を記述した電子的な文書を複数記憶する記憶手段と、

記述内容が類似する複数の文書を選出する選出手段と、

文書中の記述に出現する診療に関する単語を抽出する抽出手段と、

前記選出した各文書から抽出した単語のうち各文書に共通して出現する単語を特定する  
特定手段と、

前記選出した各文書から抽出した単語のうち前記特定した単語以外で且つ異なる文書から  
抽出した単語の組を対象に、単語の概念の同一性に係る所定の基準を満たすか否かを判  
定する判定手段と、

所定の基準を満たすと判定された単語の組を出力する出力手段として機能させるための  
表記ゆれ解析プログラム。

## 【発明の詳細な説明】

## 【技術分野】

10

20

30

40

50

## 【0001】

本発明は、同じ概念であるが表記が異なる表記ゆれを解析する表記ゆれ解析装置およびプログラムに関する。

## 【背景技術】

## 【0002】

一般に、同じ概念に対して複数の表記（単語）が存在しており、コミュニケーションを阻害するひとつの原因となっている。この表記の違い（表記ゆれ）は、地理、時期、教育などの違いに起因していることはほぼ明確である。つまり、同じような言葉遣いをしている文書は地理、時期、著者などが類似している。これら同じような言葉遣いをしているであろう文書集合の背景を、以下では「サイト」と呼ぶ。

10

## 【0003】

ここで、表記ゆれに関し、これまでに種々の発明が提案されている。

例えば、同義語がその表記の直後に丸括弧などで示されることが多いという事実を利用して、同義語対を抽出する方法が提案されている（特許文献1参照）。

例えば、単語間の同義関係の推定に、それらが出現する文書におけるそれらの単語の近傍の単語の統計情報を利用して判定する発明が提案されている（特許文献2参照）。

例えば、同一意味内容の文書（単言語パラレルコーパス）を用いて、表記ゆれを発見する発明が提案されている（特許文献3参照）。

## 【0004】

【特許文献1】特開平11-328205号公報

20

【特許文献2】特開平11-212975号公報

【特許文献3】特開2006-251843号公報

## 【発明の開示】

## 【発明が解決しようとする課題】

## 【0005】

ここで、医学の分野では、病院ごとに内部で用いられている専門用語の表記が異なる、同じ病院内でも科が違えば表記が異なる、昔使われた病名が今では違う呼び方になっている、出身大学が違うなどの理由で同じ科の医師でも表記が異なる、といったことが起きている。そして、医療は複数の機関や医師や看護師などの医療従事者が協力しあって問題解決を図る行為であるものの、これらの表記の違いによるコミュニケーションミスが、医療ミスの原因のひとつになっている。

30

## 【0006】

そこで、病院内で使われる用語の統一を図るために用語集を作成したり、広く用語の統一を図るため標準病名マスターなどのターミノロジを作成することが行われてきた。しかしながら、用語集やターミノロジの作成には手間がかかる上に、これらの手段は医療従事者に対して、決められた語彙を用いるように強制しなければならず、必ずしも上手く利用できていない。

## 【0007】

一方、表記のゆれに対してサイト間に対応表あるいは別名辞書を作成し、異なるサイト間でのコミュニケーションの際に機械的な表記の変換を行うことにより、正確に内容を伝える方法を考えることができる。しかしながら、この場合も対応表の作成に手間がかかるという問題があった。

40

## 【0008】

本発明は、上記従来事情に鑑みなされたものであり、表記が異なるが同じ概念を表す同義語と推定される単語の組を効率よく抽出することを目的としている。

## 【課題を解決するための手段】

## 【0009】

請求項1に記載の本発明は、診療結果を記述した電子的な文書を複数記憶する記憶手段と、記述内容が類似する複数の文書を選出する選出手段と、文書中の記述に出現する診療に関する単語を抽出する抽出手段と、前記選出した各文書から抽出した単語のうち各文書

50

に共通して出現する単語を特定する特定手段と、前記選出した各文書から抽出した単語のうち前記特定した単語以外で且つ異なる文書から抽出した単語の組を対象に、単語の概念の同一性に係る所定の基準を満たすか否かを判定する判定手段と、所定の基準を満たすと判定された単語の組を出力する出力手段と、を備えたことを特徴とする表記ゆれ解析装置である。

【0010】

請求項2に記載の本発明は、請求項1に記載の表記ゆれ解析装置において、前記文書には、その作成主体を示す作成主体情報が付加されており、前記判定手段は、前記対象の各単語を抽出した文書に付加された作成主体情報が異なる場合に所定の基準を満たすと判定することを特徴とする。

10

【0011】

請求項3に記載の本発明は、請求項1又は請求項2に記載の表記ゆれ解析装置において、前記判定手段は、前記対象の各単語の表記が類似する場合に所定の基準を満たすと判定することを特徴とする。

【0012】

請求項4に記載の本発明は、請求項1乃至請求項3のいずれか1項に記載の表記ゆれ解析装置において、前記判定手段は、前記対象の各単語に係る構文が共通性を有する場合に所定の基準を満たすと判定することを特徴とする。

【0013】

請求項5に記載の本発明は、請求項1乃至請求項4のいずれか1項に記載の表記ゆれ解析装置において、前記判定手段は、前記対象の各単語を抽出した文書における該単語の出現位置が共通性を有する場合に所定の基準を満たすと判定することを特徴とする。

20

【0014】

請求項6に記載の本発明は、請求項1乃至請求項5のいずれか1項に記載の表記ゆれ解析装置において、前記判定手段は、前記対象の各単語を抽出した文書に診断結果として記述された病名又はその識別コードが共通する場合に所定の基準を満たすと判定することを特徴とする。

【0015】

請求項7に記載の本発明は、請求項1乃至請求項6のいずれか1項に記載の表記ゆれ解析装置において、前記判定手段は、前記対象の各単語を抽出した文書の作成時期が共通性を有する場合に所定の基準を満たすと判定することを特徴とする。

30

【0016】

請求項8に記載の本発明は、コンピュータを、診療結果を記述した電子的な文書を複数記憶する記憶手段と、記述内容が類似する複数の文書を選出する選出手段と、文書中の記述に出現する診療に関する単語を抽出する抽出手段と、前記選出した各文書から抽出した単語のうち各文書に共通して出現する単語を特定する特定手段と、前記選出した各文書から抽出した単語のうち前記特定した単語以外で且つ異なる文書から抽出した単語の組を対象に、単語の概念の同一性に係る所定の基準を満たすか否かを判定する判定手段と、所定の基準を満たすと判定された単語の組を出力する出力手段として機能させるための表記ゆれ解析プログラムである。

40

【発明の効果】

【0017】

請求項1に記載の表記ゆれ解析装置によると、記述内容が類似する医療文書（例えば電子カルテや医療レポートなどの電子的な文書）を選出し、当該類似文書における異なる単語の組を対象に概念の同一性を判定するため、同義語と推定される単語の組を効率よく抽出することができる。

【0018】

請求項2に記載の表記ゆれ解析装置によると、医療文書の作成主体（医師又はその医師が所属する科や病院など）が同じ場合には表記ゆれが少ない（つまり、或る概念の表記が統一化されている）一方、作成主体が異なる場合には或る概念について異なる表記を用い

50

る場合が多いという性質に着目することで、記述内容が類似する医療文書における異なる単語の組について、医療文書の作成主体が異なる場合に同義語としての評価を肯定的に行うという基準に基づいて、同義語と推定される単語の組を抽出することができる。

【0019】

請求項3に記載の表記ゆれ解析装置によると、各単語の構成文字が一部異なる場合や構成文字が同じでその並びが異なる場合のように、表記が類似する単語は同一の概念を表す場合が多いという性質に着目することで、記述内容が類似する医療文書における異なる単語の組について、各単語の表記が類似する場合に同義語としての評価を肯定的に行うという基準に基づいて、同義語と推定される単語の組を抽出することができる。

【0020】

請求項4に記載の表記ゆれ解析装置によると、その単語周辺を自然言語処理した際の品詞などの形態素情報や構文情報が類似するなど、出現パターンに構文的な共通性が有る単語は同一の概念を表す場合が多いという性質に着目することで、記述内容が類似する医療文書における異なる単語の組について、各単語に係る構文が共通性を有する場合に同義語としての評価を肯定的に行うという基準に基づいて、同義語と推定される単語の組を抽出することができる。

【0021】

請求項5に記載の表記ゆれ解析装置によると、文書において単語が出現する項目名、絶対的な出現位置、或いは周辺に共通して出現する単語を考慮した相対的な出現位置など、文書における出現位置に共通性が有る単語は同一の概念を表す場合が多いという性質に着目することで、記述内容が類似する医療文書における異なる単語の組について、医療文書における各単語の出現位置が共通性を有する場合に同義語としての評価を肯定的に行うという基準に基づいて、同義語と推定される単語の組を抽出することができる。

【0022】

請求項6に記載の表記ゆれ解析装置によると、診断結果が同一の病名について記述される医療文書でほぼ同じ単語が用いられる一方、異なる単語についてはそれらが同一の概念を表す場合が多いという性質に着目することで、記述内容が類似する医療文書における異なる単語の組について、医療文書に診断結果として記述された病名又はその識別コードが共通する場合に同義語としての評価を肯定的に行うという基準に基づいて、同義語と推定される単語の組を抽出することができる。

【0023】

請求項7に記載の表記ゆれ解析装置によると、疾患の発生に時期的な特徴のある病気（例えばインフルエンザ）については、その診断結果に係る医療文書が同時期に作成されて各医療文書でほぼ同じ単語が用いられる一方、異なる単語についてはそれらが同一の概念を表す場合が多いという性質に着目することで、記述内容が類似する医療文書における異なる単語の組について、医療文書の作成時期が共通性を有する場合に同義語としての評価を肯定的に行うという基準に基づいて、同義語と推定される単語の組を抽出することができる。

【0024】

請求項8に記載の表記ゆれ解析プログラムによると、上記作用効果を奏する表記ゆれ解析装置をコンピュータを利用して実現することができる。

【発明を実施するための最良の形態】

【0025】

本発明を、以下に例示する一実施形態に基づいて具体的に説明する。

本例では、A病院とB病院の2つのサイトがそれぞれ電子カルテを保有しており、2病院が統合するなどの理由により両病院間でのコミュニケーションを円滑にするために、各電子カルテで用いられている単語の対応表を作る必要が生じた場合を例に説明する。

【0026】

図1は、本例に係る表記ゆれ解析装置の機能ブロック図を示している。

本例の表記ゆれ解析装置は、文書収集手段1、類似文書発見手段2、専門用語抽出手段

10

20

30

40

50

3、同一表記語抽出手段4、表記違い語対応抽出手段5、対応表構成手段6、を備えている。

【0027】

文書収集手段1は、双方のサイトの電子カルテを収集して記憶する。

本例で扱う電子カルテは、図2にデータ構成を例示するように、各電子カルテを識別する文書ID、所見・主訴及び病名などの診療結果を記述したテキストデータからなる実データ、電子カルテの作成者や作成日時などの属性データ、を含んでいる。なお、文書収集手段1は、電子カルテの収集の際にその収集元のサイトの識別コード（本例では病院の識別コード）を属性データに付加しており、どのサイトから収集した電子カルテかを事後的に把握できるようにしている。

10

【0028】

類似文書発見手段2は、文書収集手段1に記憶されている電子カルテのうち、記述内容（実データ）が類似するものを選出する。本例では、電子カルテの記述に出現する単語及びその出現頻度に基づくキーワードベクトルの類似性に基づいて、記述内容が類似する電子カルテを選出している。

【0029】

なお、このような統計的な評価手法以外の手法により、記述内容が類似する電子文書を選出してもよい。

例えば、診断結果が同一の病名について記述される電子カルテでほぼ同じ単語が用いられる一方、異なる単語についてはそれらが同一の概念を表す場合が多いという性質に基づいて、診断結果として記述された病名又はその識別コードが共通する電子カルテ同士を類似と判断する。例えば、疾患の発生に時期的な特徴のある病気（例えばインフルエンザ）についてはその診断結果に係る電子カルテが同時期に作成され、各電子カルテでほぼ同じ単語が用いられる一方、異なる単語についてはそれらが同一の概念を表す場合が多いという性質に基づいて、作成時期が共通性を有する電子カルテ同士を類似と判断する。

20

また、上記の各手法を複数組合せてもよく、これにより、類似の電子カルテを選出する精度が高められる。

【0030】

専門用語抽出手段3は、電子カルテの記述に出現する診療に関する単語（例えば、医療分野の専門用語）を抽出する。

30

本例では、公知の固有名称抽出技術を利用して単語抽出を行っているが、例えば、ストップワードと一般語を除去する方法など、他の方法を利用して単語抽出することもできる。

なお、抽出した各単語は、電子カルテ毎にリストとして保持してもよく、抽出元の電子カルテの該当部分にタグを付してもよい。

【0031】

同一表記語抽出手段4は、類似文書発見手段2により選出された類似の各電子カルテから専門用語抽出手段3により抽出された各単語のうち、各電子カルテに共通して出現する単語（共通語）を特定する。

本例では、各単語の共通性を、単なる文字列の同一性を見て判定することで処理の簡略化を図っているが、語の活用などの表記変化を加味して判定してもよい。また、類似の電子カルテが3以上ある場合に、本例では、全ての電子カルテに出現するものを共通語としているが、例えば、少なくとも2つの電子カルテに出現するものを共通語としてもよく、所定割合以上の電子カルテに出現するものを共通語としてもよい。

40

なお、特定した各共通語は、リストとして保持してもよく、電子カルテ毎の単語リスト中の各単語に共通語を示すフラグを付してもよく、抽出元の電子カルテの該当部分に共通語を示すタグを付してもよい。

【0032】

表記違い語対応抽出手段5は、類似文書発見手段2により選出された類似の各電子カルテから専門用語抽出手段3により抽出された各単語のうち、同一表記語抽出手段4により共通語と判定されなかった単語で異なる電子カルテから抽出された単語の組を対象に、各

50

単語の概念が同一であることを判定する。

【0033】

ここで、各単語の概念の同一性の判定（同義語であるかの判定）に係る評価基準としては、例えば以下のようなものが考えられる。

（１）抽出元の電子カルテを作成した主体が異なる場合に同義語として肯定的に評価する。

この評価基準は、電子カルテの作成主体（医師又はその医師が所属する科や病院など）が同じ場合には表記ゆれが少ない（つまり、或る概念の表記が統一化されている）一方、作成主体が異なる場合には或る概念について異なる表記を用いる場合が多いという性質に基づいている。なお、この基準を採用するにあたり、作成主体の名称又はその識別コードが電子カルテに付加（又は属性データとして記録）されていることを要する。

10

【0034】

（２）各単語の表記が類似する場合に同義語として肯定的に評価する。

この評価基準は、各単語の構成文字が一部異なる場合や構成文字が同じでその並びが異なる場合のように、表記が類似する単語は同一の概念を表す場合が多いという性質に基づいている。

【0035】

（３）各単語に係る構文が共通性を有する場合に同義語として肯定的に評価する。

この評価基準は、その単語周辺を自然言語処理した際の品詞などの形態素情報や構文情報（例えば「が格」に出現）が類似するなど、出現パターンに構文的な共通性が有る単語は同一の概念を表す場合が多いという性質に基づいている。なお、この基準を採用するにあたり、形態素解析や構文解析などの自然言語解析機能が必要となる。

20

【0036】

（４）抽出元の電子カルテにおける各単語の出現位置が共通性を有する場合に同義語として肯定的に評価する。

この評価基準は、電子カルテにおいて単語が出現する項目名（フィールド名）、絶対的な出現位置、或いは周辺に共通して出現する単語を考慮した相対的な出現位置など、電子カルテにおける出現位置に共通性が有る単語は同一の概念を表す場合が多いという性質に基づいている。

【0037】

（５）抽出元の電子カルテに診断結果として記述された病名又はその識別コード（ICD10コードや医療報酬コードなど）が共通する場合に同義語として肯定的に評価する。

この評価基準は、診断結果が同一の病名について記述される電子カルテでほぼ同じ単語が用いられる一方、異なる単語についてはそれらが同一の概念を表す場合が多いという性質に基づいている。なお、病名や識別コードの特定は、例えば病名欄に記述された単語を病名と特定し、例えば所定の規則に従った文字列を識別コードと特定するといったように、種々の手法により行うことができる。

30

【0038】

（６）抽出元の電子カルテの作成時期が共通性を有する場合に同義語として肯定的に評価する。

この評価基準は、疾患の発生に時期的な特徴のある病気（例えばインフルエンザ）については、その診断結果に係る電子カルテが同時期に作成されて各電子カルテでほぼ同じ単語が用いられる一方、異なる単語についてはそれらが同一の概念を表す場合が多いという性質に基づいている。なお、この基準を採用するにあたり、作成日時が電子カルテに付加（又は属性データとして記録）されていることを要する。

40

【0039】

このような各評価基準のいずれかを用いて同義語か否かを判定してもよいが、これらを複数組み合わせることで判定することが望ましい。

ここで、評価基準を満たすか否かの判断は、評価基準に対する適合度を算出し、これを予め定めた閾値と比較して行ってもよい。なお、複数の評価基準を用いる場合には、各評

50

価基準に対する適合度を評価基準毎の閾値と個別に比較して同義語の判定を行ってもよく、各評価基準に対する適合度を集約して算出した総括的な適合度を閾値と比較して同義語の判定を行ってもよい。また、各評価基準に対する適合度を評価基準毎の閾値と個別に比較して評価する場合には、或る評価基準について評価し、当該評価基準に適合する場合にのみ次の評価基準について評価するというように、評価対象を絞り込みつつ処理を行うことで、処理負担の軽減を図るようにしてもよい。

#### 【0040】

対応表構成手段6は、表記違い語対応抽出手段5により同義語と判定された単語の組を対応表形式で出力する。すなわち、例えば図3に示すように、同義語と判定された「肝癌」及び「肝臓癌」を並列配置した対応表を出力する。なお、同図においては、その抽出元の電子カルテの作成に係る病院毎に単語を対応付けた形式にて出力している。

対応表の出力は、液晶ディスプレイ等の表示手段により表示出力して利用者に提示してもよく、用紙等の媒体に印刷手段により印刷出力して利用者に提示してもよく、情報を記憶保持する記憶手段に出力して事後的に利用できるようにしてもよい。

#### 【0041】

図4は、本例の表記ゆれ解析装置による対応表作成の処理フローを示している。

類似文書発見手段2が、文書収集手段1に記憶されている電子カルテのうち記述内容が類似するものを選出し(ステップS1)、専門用語抽出手段3が、前記選出された類似の電子カルテの記述に出現する診療に関する単語を抽出し(ステップS2)、同一表記語抽出手段4が、前記抽出された各単語のうち、各電子カルテに共通して出現する共通語を特定する(ステップS3)。そして、表記違い語対応抽出手段5が、前記抽出された各単語のうちの共通語以外で且つ異なる電子カルテから抽出された単語の組を対象に、各単語の概念が同一であるか(同義語であるか)を判定し(ステップS4)、対応表構成手段6が、同義語と判定された単語の組を対応表形式で出力する(ステップS5)。

#### 【0042】

なお、類似文書発見手段2において、キーワードベクトル法を用いて類似の電子カルテを選出する場合には、専門用語抽出手段3による各電子カルテからの単語抽出を先に実行し、その抽出結果に基づいて類似文書発見手段2が各電子文書のキーワードベクトルを求めて電子カルテの類似性を判断するようにしてもよい。

#### 【0043】

図5は、本例に係る表記ゆれ解析装置の主要なハードウェア構成を示している。

すなわち、本例の表記ゆれ解析装置は、各種演算処理を行うCPU、CPUの作業領域となるRAM、基本的な制御プログラムを記憶するROM、本発明に係る各機能を実現するためのプログラム等を記憶するHDD、利用者に対する情報を表示出力する液晶ディスプレイや利用者からの情報の入力を受け付けるマウス・キーボード等の機器とのインターフェースである入出力I/F、他の装置との間で通信を行うインターフェースである通信I/F、等のハードウェア資源を有するコンピュータで構成されている。

#### 【0044】

そして、本発明に係るプログラムをHDDから読み出してRAMに展開し、これをCPUにより実行させることで、本発明に係る記憶手段(文書収集手段1)、選出手段(類似文書発見手段2)、抽出手段(専門用語抽出手段3)、特定手段(同一表記語抽出手段4)、判定手段(表記違い語対応抽出手段5)、出力手段(対応表構成手段6)を、表記ゆれ解析装置のコンピュータに実現している。

#### 【0045】

なお、本発明に係るプログラムは、例えば、当該プログラムを記憶したCD-ROM等の外部記憶媒体を配布する形式や、ネットワークを介して配信する形式により、本発明の実施者に提供される。

また、本発明に係る表記ゆれ解析装置の各機能手段は、本例のようなソフトウェア構成により実現する態様に限られず、専用のハードウェアモジュールで構成してもよい。

また、本発明に係る表記ゆれ解析装置の各機能手段は、本例のように1台のコンピュー

10

20

30

40

50

タに設ける態様に限られず、複数台のコンピュータに分散して設けてもよい。

【図面の簡単な説明】

【0046】

【図1】本発明の一実施形態に係る表記ゆれ解析装置の機能ブロック図である。

【図2】本発明の一実施形態に係る電子カルテのデータ構成を例示する図である。

【図3】本発明の一実施形態に係る対応表を例示する図である。

【図4】本発明の一実施形態に係る対応表作成の処理フロー図である。

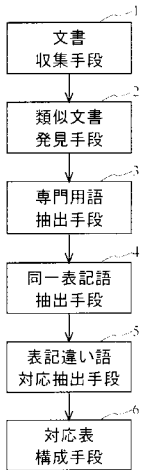
【図5】本発明の一実施形態に係る表記ゆれ解析装置のハードウェア構成図である。

【符号の説明】

【0047】

- 1：文書収集手段、
- 2：類似文書発見手段、
- 3：専門用語抽出手段、
- 4：同一表記語抽出手段、
- 5：表記違い語対応抽出手段、
- 6：対応表構成手段

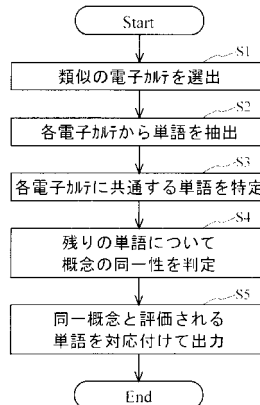
【図1】



【図3】

A病院	B病院
肝癌	肝臓癌
⋮	⋮

【図4】



【図2】

文書ID	
実行日時	
属性情報	

【 図 5 】

