

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7526316号
(P7526316)

(45)発行日 令和6年7月31日(2024.7.31)

(24)登録日 令和6年7月23日(2024.7.23)

(51)国際特許分類	F I
G 0 6 F 40/247 (2020.01)	G 0 6 F 40/247
G 0 6 F 40/30 (2020.01)	G 0 6 F 40/30
G 1 6 H 70/00 (2018.01)	G 1 6 H 70/00

請求項の数 8 (全23頁)

(21)出願番号	特願2023-93632(P2023-93632)	(73)特許権者	521162399 之江実験室 中国 3 1 1 1 2 1 浙江省杭州市余杭区中泰街道科創大道之江実験室
(22)出願日	令和5年6月7日(2023.6.7)	(74)代理人	100128347 弁理士 西内 盛二
(65)公開番号	特開2024-27087(P2024-27087A)	(72)発明者	李 勤 松 中国 3 1 1 1 2 1 浙江省杭州市余杭区中泰街道科 創 大道之江 実 験 室
(43)公開日	令和6年2月29日(2024.2.29)	(72)発明者	俞 紫 怡 中国 3 1 1 1 2 1 浙江省杭州市余杭区中泰街道科 創 大道之江 実 験 室
審査請求日	令和5年6月7日(2023.6.7)	(72)発明者	楊 宗 峰 中国 3 1 1 1 2 1 浙江省杭州市余杭区中泰街道科 創 大道之江 実 験 室
(31)優先権主張番号	202210980851.1		
(32)優先日	令和4年8月16日(2022.8.16)		
(33)優先権主張国・地域又は機関	中国(CN)		
早期審査対象出願			最終頁に続く

(54)【発明の名称】 汎用モデルに基づく標準的な医学用語管理システム及び方法

(57)【特許請求の範囲】

【請求項1】

汎用モデルに基づく標準的な医学用語管理システムであって、用語情報処理モジュール、用語関係ネットワーク構築モジュール、用語ライブラリ自己増幅モジュール及び可視化管理モジュールを含み、

前記用語情報処理モジュールは、オープンソース標準用語データを整理統合し、前記オープンソース標準用語データの定義及び用語の意味に対して用語の分類を行って、臨床所見、身体部位及び形態学的異常を含む用語カテゴリ及び用語カテゴリに対応する、偏側、頻度、主体症状を含む細分の属性の名称を取得し、且つ構築されたシーケンスラベリングモデルによってカルテテキストを分割してカルテテキストの細分の属性の名称に対応する細分の属性内容を取得することに用いられ、各種類の前記用語カテゴリがいずれも1種類の主要な細分の属性の名称及び複数種類の副次的な細分の属性の名称を含み、前記細分の属性内容は細分の属性に対応する具体的な表現語であり、

前記用語関係ネットワーク構築モジュールは、前記細分の属性の内容から語義類似度が最も高い標準的な語義単語を検索して標準化マッピング関係を確立して、前記標準化マッピング関係を補正及び追加し、用語関係ネットワークの設立を完了することに用いられ、

前記用語関係ネットワーク構築モジュールによる具体的な過程は、

前記シーケンスラベリングモデルによってカルテテキストを分割してカルテテキストの細分の属性を取得し、用語関係ネットワーク構築モジュールによって前記標準的な単語から任意の標準的な語義単語を検索し、アルゴリズムによってカルテテキストの細分の属性

と前記任意の標準的な語義単語との語義類似度を取得し、前記カルテテキストの細分の属性の語義類似度が最も高い標準的な単語を選択して、前記カルテテキストの細分の属性と選択された前記語義類似度が最も高い標準的な単語との間に標準化マッピング関係を確立するステップと、

アルゴリズムによって前記カルテテキストを分割して副次的な細分の属性を取得し、用語ライブラリには説明タイプを定義できる親ノードが存在し、且つ分割して取得された主要な細分の属性にはタイプ情報が含まれない場合、副次的な細分の属性の用語ライブラリにおける親ノードから情報を選択して主要な細分の属性を情報追加して、情報に対応する細粒度サブノードに補正し、標準化マッピング関係の補正を完了し、前記説明タイプは細分の属性の単語を分類して定義することに用いられ、前記細粒度サブノードは情報のより細粒度のサブノードであるステップと、

10

単一の用語カテゴリが複数のタイプの副次的な細分の属性を含むことに起因して補正後の主要な細分の属性が複数ある場合、用語ライブラリにおける最も多く存在する主要な細分の属性の親ノードを最後の主要な細分の属性として選択するステップと、

アルゴリズムによって分割して取得された結果は各次元の情報に欠損があって、データソース自身に階層構造が存在することである場合、用語ライブラリにおける前記用語カテゴリの親ノード分割結果によって情報補完を行い、用語関係ネットワークの設立を完了するステップと、を含み、

前記用語ライブラリ自己増幅モジュールは、異なる前記用語カテゴリに対して処理を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得することに用いられ、

20

前記用語ライブラリ自己増幅モジュールは具体的に、語義類似度によって標準化マッピングの有効性を判断し、標準化マッピングが有効である場合、直接にマッピング結果とし、標準化マッピングが無効である場合、前記カルテテキストの細分の属性から前記標準的な単語における他の可能な細分の属性の名称及び標準的な語義単語への標準化マッピングを、アルゴリズムが推奨する、手動で監査する必要があるマッピング結果として改めて計算し、異なる用語カテゴリに対して処理を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得し、

前記可視化管理モジュールは、前記用語ライブラリに対して可視化管理を行うことに用いられる

30

ことを特徴とする汎用モデルに基づく標準的な医学用語管理システム。

【請求項 2】

前記可視化管理モジュールは、標準的な医学用語クエリユニット、用語編集ユニット、用語監査ユニット及び用語マッピングユニットを含み、

前記標準的な医学用語クエリユニットは、ユーザーのプレビューデータを提供し、標準的な用語の可視化クエリを行い、異なる次元の選別条件に応じて対応する用語を選別するとともに、ユーザーログインインターフェース入口及び用語管理インターフェース入口を提供し、標準的な用語のトップレベルカテゴリ、同義語、親ノード用語、サブノード用語及びノ又は属性情報を含む次元の標準的な用語情報を表示し、それと同時に関連用語詳細インターフェースに入る入口を提供することに用いられ、

40

前記用語編集ユニットは、ユーザーが可視化インターフェースを介して標準的な用語の追加、削除及びノ又は補正、標準的な用語の同義語の追加及びノ又は削除、標準的な用語の属性情報の補正を含む用語エンティティの編集を行うことに用いられ、

前記用語監査ユニットは、用語監査者が第三者監査を行い、監査者が用語監査結果の判定を行うことに用いられ、

前記用語マッピングユニットは、ユーザーが異なるソースの医学データに対して標準化マッピングを行い、マッピング不可能なデータに対して推奨キューを構築し、用語管理者による用語のワンクリックマッピング及び推奨キューの形成を確保することに用いられる

ことを特徴とする請求項 1 に記載の汎用モデルに基づく標準的な医学用語管理システム。

【請求項 3】

50

汎用モデルに基づく標準的な医学用語管理方法であって、ステップS 1～ステップS 4を含み、

前記ステップS 1では、用語情報処理モジュールによってオープンソース標準用語データを整理統合し、前記オープンソース標準用語データの定義及び用語の意味に対して用語の分類を行って、臨床所見、身体部位及び形態学的異常を含む用語カテゴリ及び用語カテゴリに対応する、偏側、頻度、主体症状を含む細分の属性の名称を取得し、且つ構築されたシーケンスラベリングモデルによって分割して前記細分の属性の名称に対応する細分の属性の内容を取得し、前記細分の属性の内容が標準的な単語であり、各種類の前記用語カテゴリがいずれも1種類の主要な細分の属性の名称及び複数種類の副次的な細分の属性の名称を含み、前記細分の属性の内容は細分の属性に対応する具体的な表現語であり、

10

前記ステップS 2では、前記シーケンスラベリングモデルによってカルテテキストを分割してカルテテキストの細分の属性を取得し、用語関係ネットワーク構築モジュールによって前記標準的な単語から語義類似度が最も高い標準的な語義単語を検索して標準化マッピング関係を確立して、前記標準化マッピング関係を補正及び追加し、用語関係ネットワークの設立を完了し、

前記ステップS 2はサブステップS 2 1～サブステップS 2 4を含み、

前記サブステップS 2 1では、前記シーケンスラベリングモデルによってカルテテキストを分割してカルテテキストの細分の属性を取得し、用語関係ネットワーク構築モジュールによって前記標準的な単語から任意の標準的な語義単語を検索し、アルゴリズムによってカルテテキストの細分の属性と前記任意の標準的な語義単語との語義類似度を取得し、前記カルテテキストの細分の属性の語義類似度が最も高い標準的な単語を選択して、前記カルテテキストの細分の属性と選択された前記語義類似度が最も高い標準的な単語との間に標準化マッピング関係を確立し、

20

前記サブステップS 2 2では、アルゴリズムによって前記カルテテキストを分割して副次的な細分の属性を取得し、現在のカルテテキストには説明タイプを定義できる親ノードが存在し、且つ分割して取得された主要な細分の属性にはタイプ情報が含まれない場合、副次的な細分の属性の用語ライブラリにおける親ノードから情報を選択して主要な細分の属性を情報追加して、情報に対応する細粒度サブノードに補正し、標準化マッピング関係の補正を完了し、前記説明タイプは細分の属性の単語を分類して定義することに用いられ、前記細粒度サブノードは情報のより細粒度のサブノードであり、

30

前記サブステップS 2 3では、単一の用語カテゴリが複数のタイプの副次的な細分の属性を含むことに起因して補正後の主要な細分の属性が複数ある場合、用語ライブラリにおける最も多く存在する主要な細分の属性の親ノードを最後の主要な細分の属性として選択し、

前記サブステップS 2 4では、アルゴリズムによって分割して取得された結果は各次元の情報に欠損があって、データソース自身に階層構造が存在することである場合、用語ライブラリにおける前記用語カテゴリの親ノード分割結果によって情報補完を行い、用語関係ネットワークの設立を完了し、

前記ステップS 3では、用語ライブラリ自己増幅モジュールによって異なる用語カテゴリに対して処理を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得し、

40

前記ステップS 3は具体的に、語義類似度によって標準化マッピングの有効性を判断し、標準化マッピングが有効である場合、直接にマッピング結果とし、標準化マッピングが無効である場合、前記カルテテキストの細分の属性から前記標準的な単語における他の可能な細分の属性の名称及び標準的な語義単語への標準化マッピングを、アルゴリズムが推奨する、手動で監査する必要があるマッピング結果として改めて計算し、異なる用語カテゴリに対して処理を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得し、

前記ステップS 4では、可視化管理モジュールによって前記用語ライブラリに対して可視化管理を行い、前記可視化管理が標準的な医学用語クエリユニット、用語編集ユニット

50

、用語監査ユニット及び用語マッピングユニットを備える

ことを特徴とする汎用モデルに基づく標準的な医学用語管理方法。

【請求項 4】

前記ステップ S 1 は具体的に、サブステップ S 1 1 ~ サブステップ S 1 2 を含み、

前記サブステップ S 1 1 では、用語情報処理モジュールによってオープンソース標準用語データを整理統合し、前記オープンソース標準用語データの定義及び用語の意味に対して用語の分類を行って、元の用語の階層関係、関連関係及び / 又は符号化情報を保持してソースタグを付けて、用語カテゴリ及び用語カテゴリに対応する細分の属性の名称を取得し、

前記サブステップ S 1 2 では、事前訓練言語モデルによってシーケンスラベリングモデルを構築して、前記シーケンスラベリングモデルによってオープンソース標準用語データを分割して細分の属性の内容を取得する

ことを特徴とする請求項 3 に記載の汎用モデルに基づく標準的な医学用語管理方法。

【請求項 5】

前記サブステップ S 1 2 は具体的に、サブステップ S 1 2 1 ~ サブステップ S 1 2 4 を含み、

前記サブステップ S 1 2 1 では、事前訓練言語モデルによってオープンソース標準用語データにおける各文字 / 単語のベクトル表現を計算し、

前記サブステップ S 1 2 2 では、前記ベクトル表現に対して自己注意メカニズムによって各文字 / 単語の出力状態を計算し、

前記サブステップ S 1 2 3 では、前記出力状態に対して完全接続ニューラルネットワークにより計算を行って各文字 / 単語が各種類の細分の属性の名称に属する確率を取得し、シーケンスラベリングモデルの構築を完了し、

前記サブステップ S 1 2 4 では、前記オープンソース標準用語データを前記シーケンスラベリングモデルによって分割して細分の属性の内容を取得し、前記細分の属性の内容が標準的な単語であり、前記細分の属性の内容が主要な細分の属性の内容及び副次的な細分の属性の内容を含む

ことを特徴とする請求項 4 に記載の汎用モデルに基づく標準的な医学用語管理方法。

【請求項 6】

前記サブステップ S 2 1 は具体的に、サブステップ S 2 1 1 ~ サブステップ S 2 1 2 を含み、

前記サブステップ S 2 1 1 では、前記カルテテキストの細分の属性及び前記標準的な単語から前記カルテテキストの細分の属性と前記標準的な単語を含む文セット、前記カルテテキストの細分の属性及び前記標準的な単語の左側に位置する文セット、並びに前記カルテテキストの細分の属性及び前記標準的な単語の右側に位置する文セットを取得し、

前記サブステップ S 2 1 2 では、いずれか 1 対の文セットにおけるいずれか 1 対の文のベクトル表現の類似度の平均値を計算して、最終的な平均値の組み合わせによって語義類似度を取得し、語義類似度が最も高い標準的な単語を選択して標準化マッピング関係を確立する

ことを特徴とする請求項 3 に記載の汎用モデルに基づく標準的な医学用語管理方法。

【請求項 7】

前記ステップ S 3 は具体的に、サブステップ S 3 1 ~ サブステップ S 3 5 を含み、

前記サブステップ S 3 1 では、語義類似度によって標準化マッピングの有効性を判断し、標準化マッピングが無効である場合、カルテテキストを分割して取得したカルテテキストの細分の属性と標準的な語義単語との距離を計算し、

前記サブステップ S 3 2 では、標準的な語義単語を距離に応じて降順で配列し、上位 K 個の標準的な単語を取って細分の属性との距離が最も近い標準的な単語セットを構成し、

前記サブステップ S 3 3 では、細分の属性の名称における標準的な語義単語の推奨スコアを計算し、すべての細分の属性の名称を推奨スコアに応じて降順で配列し、上位 k 1 個の細分の属性の名称を取ってアルゴリズムが推奨する細分の属性から標準的な用語システ

10

20

30

40

50

ムにおける細分の属性の名称へマッピングするアルゴリズム推奨キューを構成し、

前記サブステップS 3 4では、すべての標準的な単語を語義類似度に応じて降順で配列し、上位k 2個の標準的な単語を取って細分の属性から細分の属性の名称における標準的な単語へマッピングするアルゴリズム推奨キューを構成し、

前記サブステップS 3 5では、すべてのアルゴリズム推奨キューを組み合わせ、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得する

ことを特徴とする請求項3に記載の汎用モデルに基づく標準的な医学用語管理方法。

【請求項8】

前記ステップS 4は具体的に、ユーザーが異なる選別条件に応じて、標準的な医学用語クエリユニットによって対応する用語を選別することと、ユーザーが用語編集ユニットによって標準的な用語の追加、削除及び/又は補正、標準的な用語の同義語の追加及び/又は削除、標準的な用語の属性情報の補正を含む用語エンティティの編集を行うことと、用語監査者が用語監査ユニットによって用語監査結果の判定を行うことと、ユーザーが用語マッピングユニットによって異なるソースの医学データに対して標準化マッピングを行い、マッピング不可能なデータに対して推奨キューを構築し、用語管理者による用語のワンクリックマッピング及び推奨キューの形成を確保することと、を含む

ことを特徴とする請求項3に記載の汎用モデルに基づく標準的な医学用語管理方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は医学分野の技術分野に関し、特に汎用モデルに基づく標準的な医学用語管理システム及び方法に関する。

【背景技術】

【0002】

ビッグデータ時代の到来に伴い、医療ビッグデータの医学分野での応用価値も徐々に掘り起こして利用されるようになる。ところが、この過程において、マルチソースの医学データは統一された標準を有せず、過去に規範的な制約がないため、各業界において既存の医療データを最大限に合理的に利用できず、これも医学情報化構築及び医学研究発展の妨害要素となる。

【0003】

現在、各国内外の権威機構が発表した標準的な医学用語集は既に異なる分野に広く応用及び認可されているが、複数のニーズシーンに対応できて異なるサービスプラットフォーム間の壁を突破できる中国語の標準的な医学用語システムがまだない。このため、今の段階において、中国語の標準的な用語の現実の世界におけるカバー率が40%よりも低いだけであり、医療人工知能の応用ニーズを満たすことができない。

【0004】

既存の中国語医学用語管理方法において、第1としては、snowflakeアルゴリズムによって同義の医学用語をマーキングし、概念を中心として統合した後、単一分類におけるツリーデータ構造医学辞書を構築し、第2としては、英語の医学用語リソース例えばSNOMED CTに対して機械翻訳及び大量の手動補正を行う。

【0005】

第1形態においては、異なるソースから取得された医学用語を分類して、異なる意味の医学用語を取得し、概念を中心として同義の集まり及び語義の分類を実現し、更に異なる意味の医学概念をsnowflakeアルゴリズムによってマーキングし、異なる概念識別子に基づいて統合して医学概念リストを取得するとともに、正確な分類を満たすようにJaccard類似度アルゴリズムによって同じソースの医学用語に対して類似度の計算を行う。このような方法によってインクリメントIDを実現し、IDの一意性及びインデックス性能を確保する。その構築された用語システムは構造が単一で、拡張性が低く、多様なデータソース及び絶えず更新している応用シーンニーズを満たすための汎化能力及び柔軟性がない。用語内部及び用語間の関連情報を効果的に利用しないため、用語集の後続の

10

20

30

40

50

メンテナンス及び拡張のために依然として多くの人件費を要する必要がある。1つの標準的な用語概念を追加する場合を例とし、従来の技術案は元の用語リストに同義の概念が存在するかどうかを判断することにより概念の分類又は追加を行うことができるだけである。追加した概念と元の用語リストにおける用語との間に階層関係がある場合、新たな概念を有効なノードの位置に自動的に追加できず、追加概念と既存概念との間に情報の関連がある場合、それも識別できず、従って関連情報を記憶するのであり、従来技術は用語自身の既存情報を効果的に利用して同類用語の集まりを行うことができず、手動で管理及びメンテナンスする必要があり、標準化された用語管理システム及び反復閉ループを効果的に形成することができない。

【0006】

第2形態においては、大量の手動補正を行わずに機械翻訳のみを行う条件において、翻訳結果が専門概念から大きく外れることを回避できず、中国国内外言語の使用習慣の相違も無視しやすく、用語をマッピングする際に取得したリコール率が極めて低いことが多い。リコール率を向上させる必要がある場合、管理及びメンテナンスのために倍になる人件費及び資金コスト、並びに長い時間サイクルを要する。

【0007】

現在、構築された用語システムは構造が単一で、拡張性が低く、多様なデータソース及び業務ニーズを満たすための汎化能力がなく、用語ライブラリを持続的且つ効率的に管理及び最適化するための柔軟性がない。主な表現は以下のとおりである。第(1)として、ソース情報を保留せず、標準的な用語と非標準的な用語とのマッピング関係を確立・記憶しない。第(2)として、用語内部及び用語間の関連関係を合理的に利用せず、標準的な用語概念の追加の場合には、元の用語リストに同義の概念が存在するかどうかを判断することにより概念の分類又は追加を行うことができるだけであり、追加する必要がある概念と元の用語リストにおける用語との間に階層関係がある場合、新たな概念を有効なノードの位置に自動的に追加できず、追加する必要がある概念は特定の応用ニーズを満たす必要があって、より高い階層に位置する場合、その中の関連情報を効果的に利用してそのサブクラスの同類用語の集まりを行い、第(3)として、用語ライブラリの後続の反復は多くの人件費を要し、効率的で標準化された用語管理システム及び反復閉ループを形成できない。

【0008】

実際のニーズシーンを例とし、用語ライブラリがDRG又はDIPシステムの業務ニーズを満たす必要がある場合、用語を関連する標準データベース例えばICD-10医療保険バージョン、ICD-9-CM-3医療保険バージョンにマッピング又は遡及できるように確保する必要があり、1つの標準概念を例とし、臨床所見概念「下肢痛」は現在確立された標準的な用語ライブラリに存在せず且つ追加される必要がある場合、それを「四肢痛」及び「下肢臨床所見」のノードに自動的に追加して上記2つの用語概念のサブノードとする必要があるとともに、元に「四肢痛」ノードにある「下腿痛」、「足部痛」、「左下肢痛」を「下肢痛」のサブノードに一括で補正する必要があり、そして、身体部位情報「下肢」を発病部位に自動的に関連付けする必要がある。明らかに、従来技術はまだ上記用語の自動化管理プロセスを完了できない。

【0009】

このために、上記技術的問題を解決するように汎用モデルに基づく標準的な医学用語管理システム及び方法を提供する。

【発明の概要】**【発明が解決しようとする課題】****【0010】**

上記技術的問題を解決するために、本発明は汎用モデルに基づく標準的な医学用語管理システム及び方法を提供する。

【課題を解決するための手段】**【0011】**

10

20

30

40

50

本発明が用いる技術案は以下のとおりである。

【0012】

汎用モデルに基づく標準的な医学用語管理システムであって、オープンソース標準用語データを整理統合し、前記オープンソース標準用語データの定義及び用語の意味に対して用語の分類を行って、用語カテゴリ及び用語カテゴリに対応する細分の属性名称を取得し、且つ構築されたシーケンスラベリングモデルによってカルテテキストを分割してカルテテキストの細分の属性名称に対応する細分の属性内容を取得することに用いられ、各種類の前記用語カテゴリがいずれも1種類の主要な細分の属性名称及び複数種類の副次的な細分の属性名称を含む用語情報処理モジュールと、前記細分の属性内容から語義類似度が最も高い標準的な語義単語を検索して標準化マッピング関係を確立して、前記標準化マッピング関係を補正及び追加し、用語関係ネットワークの設立を完了するための用語関係ネットワーク構築モジュールと、異なる前記用語カテゴリに対してインクリメント推奨を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得するための用語ライブラリ自己増幅モジュールと、前記用語ライブラリに対して可視化管理を行うための可視化管理モジュールと、を備える。

10

【0013】

更に、前記可視化管理モジュールは、ユーザーのプレビューデータを提供し、標準的な用語の可視化クエリを行い、異なる次元の選別条件に応じて対応する用語を選別するとともに、ユーザーログインインターフェース入口及び用語管理インターフェース入口を提供し、標準的な用語のトップレベルカテゴリ、同義語、親ノード用語、サブノード用語及び/又は属性情報を含む次元の標準的な用語情報を表示し、それと同時に関連用語詳細インターフェースに入る入口を提供するための標準的な医学用語クエリユニットと、ユーザーが可視化インターフェースを介して標準的な用語の追加、削除及び/又は補正、標準的な用語の同義語の追加及び/又は削除、標準的な用語の属性情報の補正を含む用語エンティティの編集を行うための用語編集ユニットと、用語監査者が第三者監査を行い、監査者が用語監査結果の判定を行うための用語監査ユニットと、ユーザーが異なるソースの医学データに対して標準化マッピングを行い、マッピング不可能なデータに対して推奨キューを構築し、用語管理者による用語のワンクリックマッピング及び推奨キューの形成を確保するための用語マッピングユニットと、を備える。

20

【0014】

本発明は汎用モデルに基づく標準的な医学用語管理方法を更に提供し、用語情報処理モジュールによってオープンソース標準用語データを整理統合し、前記オープンソース標準用語データの定義及び用語の意味に対して用語の分類を行って、用語カテゴリ及び用語カテゴリに対応する細分の属性名称を取得し、且つ構築されたシーケンスラベリングモデルによって分割して前記細分の属性名称に対応する細分の属性内容を取得し、前記細分の属性内容が標準的な単語であり、各種類の前記用語カテゴリがいずれも1種類の主要な細分の属性名称及び複数種類の副次的な細分の属性名称を含むステップS1と、前記シーケンスラベリングモデルによってカルテテキストを分割してカルテテキストの細分の属性を取得し、用語関係ネットワーク構築モジュールによって前記標準的な単語から語義類似度が最も高い標準的な語義単語を検索して標準化マッピング関係を確立して、前記標準化マッピング関係を補正及び追加し、用語関係ネットワークの設立を完了するステップS2と、用語ライブラリ自己増幅モジュールによって異なる用語カテゴリに対してインクリメント推奨を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得するステップS3と、可視化管理モジュールによって前記用語ライブラリに対して可視化管理を行い、前記可視化管理が標準的な医学用語クエリユニット、用語編集ユニット、用語監査ユニット及び用語マッピングユニットを備えるステップS4と、を含む。

30

40

【0015】

更に、前記ステップS1は具体的に、用語情報処理モジュールによってオープンソース標準用語データを整理統合し、前記オープンソース標準用語データの定義及び用語の意味に対して用語の分類を行って、元の用語の階層関係、関連関係及び/又は符号化情報を保

50

持してソースタグを付けて、用語カテゴリ及び用語カテゴリに対応する細分の属性名称を取得するサブステップS 1 1と、事前訓練言語モデルによってシーケンスラベリングモデルを構築して、前記シーケンスラベリングモデルによってオープンソース標準用語データを分割して細分の属性内容を取得するサブステップS 1 2と、を含む。

【0016】

更に、前記サブステップS 1 2は具体的に、事前訓練言語モデルによってオープンソース標準用語データにおける各文字/単語のベクトル表現を計算するサブステップS 1 2 1と、前記ベクトル表現に対して自己注意メカニズムによって各文字/単語の出力状態を計算するサブステップS 1 2 2と、前記出力状態に対して完全接続ニューラルネットワークにより計算を行って各文字/単語が各種の細分の属性カテゴリに属する確率を取得し、シーケンスラベリングモデルの構築を完了するサブステップS 1 2 3と、前記オープンソース標準用語データを前記シーケンスラベリングモデルによって分割して細分の属性内容を取得し、前記細分の属性内容が標準的な単語であり、前記細分の属性内容が主要な細分の属性内容及び副次的な細分の属性内容を含むサブステップS 1 2 4と、を含む。

10

【0017】

更に、前記ステップS 2は具体的に、前記シーケンスラベリングモデルによってカルテテキストを分割してカルテテキストの細分の属性を取得し、用語関係ネットワーク構築モジュールによって前記標準的な単語から任意の標準的な語義単語を検索し、アルゴリズムによってカルテテキストの細分の属性と前記任意の標準的な語義単語との語義類似度を取得し、語義類似度が最も高い標準的な単語を選択して標準化マッピング関係を確立するサブ

20

ステップS 2 1と、アルゴリズムによって分割して副次的な細分の属性を取得し、現在のカルテテキストには説明タイプを定義できる親ノードが存在し、且つ分割して取得された主要な細分の属性にはタイプ情報が含まれない場合、主要な細分の属性を情報追加して情報に対応する細粒度サブノードに補正し、標準化マッピング関係の補正を完了するサブステップS 2 2と、単一の用語カテゴリが複数のタイプの副次的な細分の属性を含むことに起因して補正後の主要な細分の属性が複数ある場合、最も多く存在する主要な細分の属性のカルテテキストにおける親ノードを最後の主要な細分の属性として選択するサブステップS 2 3と、アルゴリズムによって分割して取得された結果は各次元の情報に欠損があって、データソース自身に階層構造が存在することである場合、前記用語カテゴリの親ノード分割結果によって情報補完を行い、用語関係ネットワークの設立を完了するサブステップS 2 4と、を含む。

30

【0018】

更に、前記サブステップS 2 1は具体的に、前記カルテテキストの細分の属性及び前記標準的な単語から前記カルテテキストの細分の属性と前記標準的な単語を含む文セット、前記カルテテキストの細分の属性及び前記標準的な単語の左側に位置する文セット、並びに前記カルテテキストの細分の属性及び前記標準的な単語の右側に位置する文セットを取得するサブステップS 2 1 1と、いずれか1対の文セットの平均値を計算して、最終的な平均値の組み合わせによって語義類似度を取得し、語義類似度が最も高い標準的な単語を選択して標準化マッピング関係を確立するサブステップS 2 1 2と、を含む。

40

【0019】

更に、前記ステップS 3は具体的に、語義類似度によって標準化マッピングの有効性を判断し、標準化マッピングが有効である場合、直接にマッピング結果とし、標準化マッピングが無効である場合、前記カルテテキストの細分の属性から前記標準的な単語における他の可能な細分の属性カテゴリ及び標準的な語義単語への標準化マッピングを、アルゴリズムが推奨する、手動で監査する必要があるマッピング結果として改めて計算し、異なる用語カテゴリに対してインクリメント推奨を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得する。

【0020】

更に、前記ステップS 3は具体的に、語義類似度によって標準化マッピングの有効性を

50

判断し、標準化マッピングが無効である場合、カルテテキストを分割して取得したカルテテキストの細分の属性と標準的な語義単語との距離を計算するサブステップS 3 1と、標準的な語義単語を距離に応じて降順で配列し、上位K個の標準的な単語を取って細分の属性との距離が最も近い標準的な単語セットを構成するサブステップS 3 2と、細分の属性カテゴリにおける標準的な語義単語の推奨スコアを計算し、すべての細分の属性カテゴリを推奨スコアに応じて降順で配列し、上位k 1個の細分の属性カテゴリを取ってアルゴリズムが推奨する細分の属性から標準的な用語システムにおける細分の属性カテゴリへマッピングするアルゴリズム推奨キューを構成するサブステップS 3 3と、すべての標準的な単語を語義類似度に応じて降順で配列し、上位k 2個の標準的な単語を取って細分の属性から細分の属性カテゴリにおける標準的な単語へマッピングするアルゴリズム推奨キューを構成するサブステップS 3 4と、すべてのアルゴリズム推奨キューを組み合わせ、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得するサブステップS 3 5と、を含む。

10

【0021】

更に、前記ステップS 4は具体的に、ユーザーが異なる選別条件に応じて、標準的な医学用語クエリユニットによって対応する用語を選別することと、ユーザーが用語編集ユニットによって標準的な用語の追加、削除及び/又は補正、標準的な用語の同義語の追加及び/又は削除、標準的な用語の属性情報の補正を含む用語エンティティの編集を行うことと、用語監査者が用語監査ユニットによって用語監査結果の判定を行うことと、ユーザーが用語マッピングユニットによって異なるソースの医学データに対して標準化マッピングを行い、マッピング不可能なデータに対して推奨キューを構築し、用語管理者による用語のワンクリックマッピング及び推奨キューの形成を確保することと、を含む。

20

【発明の効果】

【0022】

本発明の有益な効果は以下のとおりである。第1として、本発明が構築した標準的な医学用語ライブラリは異なる科学研究及び業務使用シーンを満たすことができ、異なるプラットフォーム及び権威機構の標準的な医学用語と互換的であり、又はそれに遡及することができ、設計された構造体が用語の各種類の組み合わせ拡張をサポートすることができ、第2として、本発明は複数の標準的な用語の間且つ単一標準的な用語内部の関連情報を保持して利用し、アルゴリズム及び論理ルールによって標準的な用語概念の自動追加及び集まりの推奨を行い、第3として、本発明は標準的な用語ライブラリに対する一部の人件費による効率的で半自動化の管理及びメンテナンスをサポートすることができる。各権威機構が発表した標準的な医学用語及び多施設臨床データと良好なフィードバックメカニズムを形成する。

30

【図面の簡単な説明】

【0023】

【図1】図1は本発明に係る汎用モデルに基づく標準的な医学用語管理システムの模式図である。

【図2】図2は本発明の実施例に係る用語関係ネットワークの構造図である。

【発明を実施するための形態】

40

【0024】

以下の少なくとも1つの例示的な実施例についての説明は実際に説明的なものであって、本発明及びその応用又は使用を制限するものではない。本発明の実施例に基づいて、当業者が創造的な労力を要することなく取得したすべての他の実施例は、いずれも本発明の保護範囲に属する。

【0025】

図1を参照し、汎用モデルに基づく標準的な医学用語管理システムであって、オープンソース標準用語データを整理統合し、前記オープンソース標準用語データの定義及び用語の意味に対して用語の分類を行って、用語カテゴリ及び用語カテゴリに対応する細分の属性名称を取得し、且つ構築されたシーケンスラベリングモデルによってカルテテキストを

50

分割してカルテテキストの細分の属性名称に対応する細分の属性内容を取得することに用いられ、各種類の前記用語カテゴリがいずれも1種類の主要な細分の属性名称及び複数種類の副次的な細分の属性名称を含む用語情報処理モジュールと、前記細分の属性内容から語義類似度が最も高い標準的な語義単語を検索して標準化マッピング関係を確立して、前記標準化マッピング関係を補正及び追加し、用語関係ネットワークの設立を完了するための用語関係ネットワーク構築モジュールと、異なる前記用語カテゴリに対してインクリメント推奨を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得するための用語ライブラリ自己増幅モジュールと、前記用語ライブラリに対して可視化管理を行うための可視化管理モジュールと、を備え、前記可視化管理モジュールは、ユーザーのプレビューデータを提供し、標準的な用語の可視化クエリを行い、異なる次元の選別条件に応じて対応する用語を選別するとともに、ユーザーログインインターフェース入口及び用語管理インターフェース入口を提供し、標準的な用語のトップレベルカテゴリ、同義語、親ノード用語、サブノード用語及び/又は属性情報を含む次元の標準的な用語情報を表示し、それと同時に関連用語詳細インターフェースに入る入口を提供するための標準的な医学用語クエリユニットと、ユーザーが可視化インターフェースを介して標準的な用語の追加、削除及び/又は補正、標準的な用語の同義語の追加及び/又は削除、標準的な用語の属性情報の補正を含む用語エンティティの編集を行うための用語編集ユニットと、用語監査者が第三者監査を行い、監査者が用語監査結果の判定を行うための用語監査ユニットと、ユーザーが異なるソースの医学データに対して標準化マッピングを行い、マッピング不可能なデータに対して推奨キューを構築し、用語管理者による用語のワンクリックマッピング及び推奨キューの形成を確保するための用語マッピングユニットと、を備える。

10

20

【0026】

実施例 汎用モデルに基づく標準的な医学用語管理方法であって、下記ステップを含む。

【0027】

ステップS1では、用語情報処理モジュールによってオープンソース標準用語データを整理統合し、前記オープンソース標準用語データの定義及び用語の意味に対して用語の分類を行って、用語カテゴリ及び用語カテゴリに対応する細分の属性名称を取得し、且つ構築されたシーケンスラベリングモデルによって分割して前記細分の属性名称に対応する細分の属性内容を取得し、前記細分の属性内容が標準的な単語であり、各種類の前記用語カテゴリがいずれも1種類の主要な細分の属性名称及び複数種類の副次的な細分の属性名称を含み、ステップS11では、用語情報処理モジュールによってオープンソース標準用語データを整理統合し、前記オープンソース標準用語データの定義及び用語の意味に対して用語の分類を行って、元の用語の階層関係、関連関係及び/又は符号化情報を保持してソースタグを付けて、用語カテゴリ及び用語カテゴリに対応する細分の属性名称を取得し、オープンソース標準用語データの間に対応関係がある場合、該情報を一括で保持し、機構又は個人が異なるシーンニーズに直面する場合に依然としてデータを効果的に利用及び変換することができ、情報交換をスムーズに行うように確保する。

30

【0028】

ステップS12では、事前訓練言語モデルによってシーケンスラベリングモデルを構築して、前記シーケンスラベリングモデルによってオープンソース標準用語データを分割して細分の属性内容を取得する。

40

【0029】

ステップS121では、事前訓練言語モデルによってオープンソース標準用語データにおける各文字/単語のベクトル表現を計算し、オープンソース標準用語データXについては、それに含まれるすべての文字及びすべての単語を一体に接合する場合、

$$X = (x_1, \dots, x_m, x_{m+1}, \dots, x_{m+n})$$

と示されてもよく、ここで、

50

X_1, \dots, X_m

が X に含まれる合計 m 個の文字であり、
 X_{m+1}, \dots, X_{m+n}

が X に含まれる合計 n 個の単語である。文字 / 単語 x_i の医学用語 X における開始位置が $head[x_i]$ と記され、終了位置が $tail[x_i]$ と記される。まず事前に訓練された Bert モデルによって各文字 / 単語のベクトル表現を $emb(x_i)$ として計算し、ステップ S 1 2 2 では、前記ベクトル表現に対して自己注意メカニズムによって各文字 / 単語の出力状態を計算し、各文字 / 単語 x_i の x_j に対する注意重み $w_{i,j}$ を計算し、

【数 1】

$$w_{i,j} = \frac{\exp(\text{emb}(x_i) \cdot M_1 \cdot \text{emb}(x_j) + \text{pos}(x_i, x_j) \cdot M_2)}{\sum_{k=1}^{m+n} \exp(\text{emb}(x_i) \cdot M_1 \cdot \text{emb}(x_k) + \text{pos}(x_i, x_k) \cdot M_2)}$$

ここで、

$$\text{pos}(x_i, x_j) = (\text{head}[x_i] - \text{head}[x_j], \text{head}[x_i] - \text{tail}[x_j], \text{tail}[x_i] - \text{head}[x_j], \text{tail}[x_i] - \text{tail}[x_j])$$

が x_i と x_j の相対位置ベクトルを示し、 M_1 と M_2 がいずれも事前に訓練された Bert モデルによって訓練して取得した重み行列パラメータである。自己注意メカニズムによって各文字 / 単語 x_i の位置での出力状態 $output(x_i)$ を計算し、

【数 2】

$$\text{output}(x_i) = \sum_{j=1}^{m+n} w_{i,j} \cdot \text{emb}(x_j)$$

ステップ S 1 2 3 では、前記出力状態に対して完全接続ニューラルネットワークにより計算を行って各文字 / 単語が各種類の細分の属性カテゴリに属する確率を取得し、シーケンスラベリングモデルの構築を完了し、最後に出力状態に対して 1 層の完全接続ニューラルネットワークによって文字 / 単語 x_i が各種類の細分の属性に属する確率分布 $prob(x_i)$ を計算し、

【数 3】

$$\text{prob}(x_i) = \text{softmax}(W \cdot \text{output}(x_i) + b)$$

ここで、 W が行列パラメータであり、 b がベクトルパラメータであり、それらがシーケンスラベリングモデル全体によって訓練して取得したもの

であり、 W の行数と b の次元がいずれも医療用語の細分の属性のカテゴリ数に等しく、 softmax がベクトルの softmax 関数を示す。 $prob(x_i)$ はオープンソース標準用語データに対するシーケンスラベリングモデルの最終出力であって、確率からなるベクトルであり、次元がオープンソース標準用語データの細分の属性カテゴリの数に等しい。 $prob(x_i)$ の k 番目の要素は文字 / 単語 x_i が第 k 種類の細分の属性カテゴ

10

20

30

40

50

りに属する確率を示し、最も大きな要素はアルゴリズムが許可する x_i が属する細分の属性カテゴリ及び対応する確率を示す。

【0030】

ステップS124では、前記オープンソース標準用語データを前記シーケンスラベリングモデルによって分割して細分の属性内容を取得し、前記細分の属性内容が標準的な単語であり、前記細分の属性内容が主要な細分の属性内容及び副次的な細分の属性内容を含む。

【0031】

すべてのカテゴリの用語に対して方向性のある内部細分の属性情報構造の設計を行い、各種類の用語カテゴリにはいずれも1種類の主要な細分の属性名称及び複数種類の副次的な細分の属性名称が含まれるとともに、異なるカテゴリ用語間の関係の定義を完了する。

10

【0032】

ステップS2では、前記シーケンスラベリングモデルによってカルテテキストを分割してカルテテキストの細分の属性を取得し、用語関係ネットワーク構築モジュールによって前記標準的な単語から語義類似度が最も高い標準的な語義単語を検索して標準化マッピング関係を確立して、前記標準化マッピング関係を補正及び追加し、用語関係ネットワークの設立を完了し、図2を参照し、図2における「Acute pancytopenia with myelofibrosis」とは「骨髓線維症を伴う急性汎骨髄症」のsnomed ctの用語集における対応する英語概念を指す。

【0033】

カルテテキストから分割して取得されたカルテテキストの細分の属性 a については、前記標準的な単語からカルテテキストの細分の属性 a との語義類似度が最も高い標準的な語義単語

20

\hat{a}

を検索し、カルテテキストの細分の属性 a と標準的な語義単語

\hat{a}

との間に標準化マッピング関係を確立する。例えば、漢方製剤の名称情報におけるケイカを物質であるドケイガイ、スイダンカにマッピングすることができる。操作の名称用語における粒子放射を物理エネルギーである電離放射線及び物質である放射性同位体にマッピングすることができる。それと同時に、属性チェーン「活性成分」、「物理エネルギー」、「目標物質」によって情報の関連付けを行う。

30

【0034】

ステップS21では、前記シーケンスラベリングモデルによってカルテテキストを分割してカルテテキストの細分の属性を取得し、用語関係ネットワーク構築モジュールによって前記標準的な単語から任意の標準的な語義単語を検索し、アルゴリズムによってカルテテキストの細分の属性と前記任意の標準的な語義単語との語義類似度を取得し、語義類似度が最も高い標準的な単語を選択して標準化マッピング関係を確立し、ステップS211では、前記カルテテキストの細分の属性及び前記標準的な単語から前記カルテテキストの細分の属性と前記標準的な単語を含む文セット、前記カルテテキストの細分の属性及び前記標準的な単語の左側に位置する文セット、並びに前記カルテテキストの細分の属性及び前記標準的な単語の右側に位置する文セットを取得し、

40

a'

で前記標準的な単語におけるカルテテキストの細分の属性 a と同じ細分の属性カテゴリを有する任意の標準的な単語を示す場合、カルテテキストの細分の属性 a と任意の標準的な単語

a'

との語義類似度は2つの部分に分けられ、第1部分はそれらのベクトル表現の類似度即ち

50

$\cos(\text{emb}(a), \text{emb}(a'))$

であり、ここで、 \cos が計算ベクトルのコサイン類似度を示す。第 2 部分はカルテテキストの細分の属性 a と任意の標準的な単語

a'

の文脈との類似度であり、この文脈とは病院のカルテテキストにおけるカルテテキストの細分の属性 a と任意の標準的な単語

a'

10

を含む文、並びにカルテテキストの細分の属性 a 及び任意の標準的な単語

a'

の左側又は右側に位置する文を指す。

$\text{sen}_0(a)$

、
 $\text{sen}_{-1}(a)$

20

及び

$\text{sen}_{+1}(a)$

でそれぞれ病院から取得された大量のカルテテキストにおけるカルテテキストの細分の属性 a を含む文セット、カルテテキストの細分の属性 a の左側に位置する文セット及びカルテテキストの細分の属性 a の右側に位置する文セットを示し、同様に

$\text{sen}_0(a')$

30

、
 $\text{sen}_{-1}(a')$

及び

$\text{sen}_{+1}(a')$

40

でそれぞれ任意の標準的な単語

a'

を含む文セット、任意の標準的な単語

a'

の左側に位置する文セット及び任意の標準的な単語

a'

の右側に位置する文セットを示す。

50

【 0 0 3 5 】

ステップ S 2 1 2 では、いずれか 1 対の文セットの平均値を計算して、最終的な平均値の組み合わせによって語義類似度を取得し、語義類似度が最も高い標準的な単語を選択して標準化マッピング関係を確立する。

【 0 0 3 6 】

文セット

$sen_0(a)$

と

$sen_0(a')$

10

との語義類似度はこの 2 つのセットにおけるいずれか 1 対の文のベクトル表現類似度の平均値として定義され、計算方法は、以下の通りであり、

【数 4】

$$\text{sim}(sen_0(a), sen_0(a')) = \frac{\sum_{u \in sen_0(a)} \sum_{v \in sen_0(a')} \cos(\text{emb}(u), \text{emb}(v))}{|sen_0(a)| \cdot |sen_0(a')|}$$

20

ここで、

$|sen_0(a)|$

と

$|sen_0(a')|$

がそれぞれセット

$sen_0(a)$

30

及び

$sen_0(a')$

における文の個数を示す。 $emb(u)$ と $emb(v)$ がそれぞれ文 u 及び v のベクトル表現を示し、事前に訓練された言語モデルにより計算して取得したものである。 cos が計算ベクトルのコサイン類似度を示す。同じ方法で計算してカルテテキストの細分の属性 a と任意の標準的な単語

40

a'

の左側の文セットとの語義類似度

$\text{sim}(sen_{-1}(a), sen_{-1}(a'))$

及び右側の文セットとの語義類似度

$\text{sim}(sen_{+1}(a), sen_{+1}(a'))$

50

を取得する。最終的に上記 2 つの部分の結果を組み合わせることでカルテテキストの細分の属性 a と任意の標準的な単語

a'

との語義類似度

$\text{sim}(a, a')$

を取得し、

【数 5】

$$\text{sim}(a, a') = w_1 \cos(\text{emb}(a), \text{emb}(a')) + w_2 \text{sim}(\text{sen}_0(a), \text{sen}_0(a')) \\ + w_3 \text{sim}(\text{sen}_{-1}(a), \text{sen}_{-1}(a')) + w_4 \text{sim}(\text{sen}_{+1}(a), \text{sen}_{+1}(a'))$$

10

ここで、

w_1, \dots, w_4

が手動で設定した重みパラメータであり、本実施例において $w_1 = w_2 = 0.4$ 、 $w_3 = w_4 = 0.1$ を取る。オープンソース標準用語データから該細分の属性カテゴリにおける任意の標準的な単語

20

a'

をトランスバースし、語義類似度が最も高い標準的な単語

$\hat{a} = \arg \max_{a'} (\text{sim}(a, a'))$

即ちカルテテキストの細分の属性 a の標準化マッピング結果を選択し、標準化マッピング関係の構築を完了する。

30

【0037】

ステップ S 2 2 では、アルゴリズムによって分割して副次的な細分の属性を取得し、現在のカルテテキストには説明タイプを定義できる親ノードが存在し、且つ分割して取得された主要な細分の属性にはタイプ情報が含まれない場合、主要な細分の属性を情報追加して情報に対応する細粒度サブノードに補正し、標準化マッピング関係の補正を完了し、例えば、データ「喀黄膿痰」については、アルゴリズムによって分割して取得した結果は、評価対象が痰であり、一般説明 # 1 が黄であり、一般説明 # 2 が膿であるということであり、細分の属性「黄」及び「膿」がそれぞれ細分の属性「色」及び「炎症形態」のノードにあるため、評価対象「痰」を「喀痰の性状」及び「喀痰の色」に補正し、又は上記用語の同一親ノード「喀痰の外観」を評価対象として用いる必要がある。

40

【0038】

ステップ S 2 3 では、単一の用語カテゴリが複数のタイプの副次的な細分の属性を含むことに起因して補正後の主要な細分の属性が複数ある場合、最も多く存在する主要な細分の属性のカルテテキストにおける親ノードを最後の主要な細分の属性として選択し、ステップ S 2 4 では、アルゴリズムによって分割して取得された結果は各次元の情報に欠損があって、データソース自身に階層構造が存在することである場合、前記用語カテゴリの親ノード分割結果によって情報補完を行い、用語関係ネットワークの設立を完了する。

【0039】

例えば、原データ「生体外血管を介する治療」については、用語を分割してから主語を取

50

得できない場合、その親ノード「薬物の静脈バイパス移植 [誘導] 圧迫療法」を分割して取得した情報で補完し、操作方法が「移植」である。データソースが階層構造を有しない場合、主語を該カテゴリにおける最も高いノード即ち「操作方法」で代替する。オープンソース標準用語データには一般的に親子階層構造の情報があるため、属性関係ネットワークを改善する必要があるだけである。取得された細分の属性が非常に完全な医学用語については、該用語と細分の属性単語とを定義された関係タイプによって関連付けする必要があるだけであり、細分の属性の完全性がない場合、上記論理に基づいて補完してから関連付けすれば、用語関係ネットワークを形成することができる。

【0040】

ステップS3では、用語ライブラリ自己増幅モジュールによって異なる用語カテゴリに対してインクリメント推奨を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得し、前記ステップS3は具体的に、語義類似度によって標準化マッピングの有効性を判断し、標準化マッピングが有効である場合、直接にマッピング結果とし、標準化マッピングが無効である場合、前記カルテテキストの細分の属性から前記標準的な単語における他の可能な細分の属性カテゴリ及び標準的な語義単語への標準化マッピングを、アルゴリズムが推奨する、手動で監査する必要があるマッピング結果として改めて計算し、異なる用語カテゴリに対してインクリメント推奨を行い、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得する。

10

【0041】

処理する必要があるカルテテキストはタイルフォーマットであり、即ち自身に階層関係が含まれない場合、取得された細分の属性情報によって階層構造の判断を行い、先に親子関係を関連付けし、更に属性関係を関連付けする。

20

【0042】

アルゴリズムはHIS、EMRシステムにおけるデータ記憶ルールに従って、異なる分類における用語に対応する選別ソースを特定する。例えば、症状、疾病類用語を電子カルテにおける診断、主訴、既往歴、補助検査ソースから取得し、操作類用語を試験検査、画像学的検査、手術などのソースから取得することができ、薬物の名称を医者 の指示、外来処方箋などのソースから取得することができる。病院から大量のカルテテキストデータを取得し、マッピングを完了できない情報の場合には、同様に分割して取得された細分の属性単語の名称、用語カテゴリなどに基づいて、アルゴリズムが推奨する記憶及び補正対象のキューを形成する。カルテテキストから分割して取得されたカルテテキストの細分の属性 a については、上記方法でそれを標準的な用語システムにおける標準的な単語

30

\hat{a}

にマッピングし、それらの語義類似度が条件

$$\text{sim}(a, \hat{a}) > c$$

40

を満たす場合、今回の標準化マッピングが有効であると見なし、直接に今回のマッピング結果を標準的な用語システムに記憶することができ、ここで、c が手動で設定した閾値であり、本実施例において $c = 0.9$ を取る。上記条件を満たさない場合、カルテテキストの細分の属性 a から標準的な単語

\hat{a}

への標準化マッピングが無効であると見なし、カルテテキストの細分の属性 a からオープンソース標準用語データにおける他の可能な細分の属性カテゴリ及び標準的な単語への標準化マッピングを、アルゴリズムが推奨する、手動で監査する必要があるマッピング結果として改めて計算する必要がある、ステップS31では、語義類似度によって標準化マ

50

ッピングの有効性を判断し、標準化マッピングが無効である場合、カルテテキストを分割して取得したカルテテキストの細分の属性と標準的な語義単語との距離を計算し、オープンソース標準用語データにおけるすべての標準的な単語のセットを A とし、カルテテキストの細分の属性カテゴリ C におけるすべての標準的な単語セットを A_C とし、カルテテキストから分割して取得したカルテテキストの細分の属性 a 及び標準的な用語システムにおける標準的な単語

$$a' \in A$$

については、それらのベクトル表現に基づいてそれらの距離を

$$\text{dist}(a, a') = \|\text{emb}(a) - \text{emb}(a')\|$$

10

として計算し、ここで、

$$\|\cdot\|$$

が計算ベクトルのモジュロを示す。

【 0 0 4 3 】

ステップ S 3 2 では、標準的な語義単語を距離に応じて降順で配列し、上位 K 個の標準的な単語を取って細分の属性との距離が最も近い標準的な単語セットを構成し、すべての標準的な単語

20

$$a' \in A$$

を

$$\text{dist}(a, a')$$

に応じて降順で配列し、上位 K 個の標準的な単語を取って細分の属性 a との距離が最も近い標準的な単語セット $A(a, K)$ を構成し、本実施例において $K = 1000$ を取る。

30

【 0 0 4 4 】

ステップ S 3 3 では、細分の属性カテゴリにおける標準的な語義単語の推奨スコアを計算し、すべての細分の属性カテゴリを推奨スコアに応じて降順で配列し、上位 k_1 個の細分の属性カテゴリを取ってアルゴリズムが推奨する細分の属性から標準的な用語システムにおける細分の属性カテゴリへマッピングするアルゴリズム推奨キューを構成し、標準的な用語システムにおける各細分の属性カテゴリ C の細分の属性 a に対する推奨スコア $\text{score}(a, C)$ は細分の属性カテゴリ C における標準的な単語がセット $A(a, K)$ に含まれる数として定義され、計算公式は、

【数 6】

$$\text{score}(a, C) = \sum_{a' \in A_C} \sigma(a, a'), \quad \sigma(a, a') = \begin{cases} 1, & a' \in A(a, K) \\ 0, & a' \notin A(a, K) \end{cases}$$

40

である。すべての細分の属性カテゴリ C を $\text{score}(a, C)$ に応じて降順で配列し、上位 k_1 個の細分の属性カテゴリを取ってアルゴリズムが推奨する、細分の属性 a から標準的な用語システムにおける細分の属性カテゴリへマッピングするアルゴリズム推奨キュー

$$\text{cand}(a) = \{C_1, C_2, \dots, C_{k_1}\}$$

を構成し、本実施例において $k_1 = 5$ を取る。

50

【 0 0 4 5 】

ステップ S 3 4 では、すべての標準的な単語を語義類似度に応じて降順で配列し、上位 k_2 個の標準的な単語を取って細分の属性から細分の属性カテゴリにおける標準的な単語へマッピングするアルゴリズム推奨キューを構成し、細分の属性カテゴリ

$$C_i \in \text{cand}(a)$$

については、上記方法で標準的な用語システムにおける細分の属性カテゴリ C_i における任意の標準的な単語

$$a'$$

10

と細分の属性 a との語義類似度

$$\text{sim}(a, a')$$

を計算し、すべての標準的な単語

$$a' \in C_i$$

を

20

$$\text{sim}(a, a')$$

に応じて降順で配列し、上位 k_2 個の標準的な単語を取って細分の属性 a から細分の属性カテゴリ C_i における標準的な単語へマッピングするアルゴリズム推奨キュー

$$\text{cand}(a, C_i) = \{a'_{C_i,1}, a'_{C_i,2}, \dots, a'_{C_i,k_2}\}$$

を構成し、本実施例において $k_2 = 3$ を取る。

【 0 0 4 6 】

30

ステップ S 3 5 では、すべてのアルゴリズム推奨キューを組み合わせて、用語関係ネットワークの拡張を完了し、完全な用語ライブラリを取得する。

【 0 0 4 7 】

すべての $\text{cand}(a, C_i)$ を組み合わせて出力とすれば、細分の属性 a から標準的な用語システムへ標準化マッピングを行うアルゴリズムが推奨する、手動で監査及び記憶すべき推奨キュー

$$\{\text{cand}(a, C_1), \text{cand}(a, C_2), \dots, \text{cand}(a, C_{k_1})\}$$

を形成する。

40

【 0 0 4 8 】

例えば、アルゴリズムによって臨床データ「超音波監視による子宮鏡手術をする」を分割して操作方法「超音波監視」を取得し、「子宮鏡」をそれぞれ「超音波誘導」、「顕微鏡検査」に標準化することができ、器具「子宮鏡」を直接マッピングして標準概念「子宮鏡」を取得することができる。該臨床用語が標準的な用語ライブラリ「超音波誘導による子宮鏡検査」の関係ネットワークマップに一致する場合、直接マッピングすることができ、用語の追加を行う必要がない。例えば、アルゴリズムによって臨床データ「両下肢赤腫」を分割して発病部位「両下肢」、相関形態学「赤腫」を取得するが、「赤腫」は標準的な用語ライブラリにおいて直接マッピングできる標準概念がまだない。アルゴリズムは語義に基づいて推奨キューを形成する。

50

【 0 0 4 9 】

表 1 分割・組み合わせについての推奨

シーケンス番号	分割・組み合わせ
1	発病部位：両下肢＋相関形態学：赤腫
2	発病部位：両下肢＋症状主体：赤腫
3	発病部位：両下肢＋相関形態学：腫＋一般説明：赤
4	発病部位：両下肢＋症状主体：腫＋一般説明：赤

10

【 0 0 5 0 】

表 2 標準概念へのマッピングについての推奨

シーケンス番号	概念の分類	標準概念	関係の確立
1	形態学的異常	腫大	サブレベル 同義語
2	形態学的異常	腫脹	サブレベル 同義語
3	形態学的異常	腫物	サブレベル 同義語
4	形態学的異常	血腫	サブレベル 同義語
5	形態学的異常	炎症	サブレベル 同義語
6	限定単語	赤色	サブレベル 同義語
7	臨床所見	赤色	サブレベル 同義語

20

【 0 0 5 1 】

用語管理者はアルゴリズム推奨結果に基づいて分割・組み合わせ及び標準マッピングの選択を行い、用語「両下肢赤腫」を「発病部位」＋「相関形態学」に分割し、用語「赤腫」と標準概念を正しく関連付けする。

【 0 0 5 2 】

同一用語カテゴリにおける用語については、細分の属性が重なる用語は既定の集まりの論理、標準概念の命名ルールに従って概念の追加及び対応する情報関連付けを行うことができ、一部の論理ルールの例は下記表 3 に示される。

30

【 0 0 5 3 】

表 3 論理ルールの例

40

50

用語の分類	重なり属性	概念の命名ルール及び例
臨床所見	発病部位	{追加概念：発病部位＋「関連臨床所見」} 例： 下肢麻痺、下肢赤腫、下肢腫痛、下肢痛であり、重なる細分の属性「発病部位：下肢」がある場合、追加標準概念である下肢関連臨床所見を推奨し、親ノードである臨床所見を推奨することができる
臨床所見	相関形態学	{追加概念：形態学的異常＋「症状」} 例： 歯肉出血、月経期出血、大便出血、鼻血であり、重なる細分の属性「相関形態学：出血」がある場合、追加標準概念である出血症状を推奨し、親ノードである臨床所見を推奨することができる
操作	目的	条件：関わる用語にはいずれも祖孫概念である手術が存在する {追加概念：目的＋「手術」} 例： 治療的角膜光切除術、内視鏡的結腸病変のアルゴンプラズマ凝固法による治療術、頭部血管治療のための超音波であり、重なる細分の属性「目的：治療的」がある場合、追加標準概念である治療的手術を推奨し、親ノードである手術を推奨することができる

10

20

30

【0054】

ステップS4では、可視化管理モジュールによって前記用語ライブラリに対して可視化管理を行い、前記可視化管理が標準的な医学用語クエリユニット、用語編集ユニット、用語監査ユニット及び用語マッピングユニットを備える。

【0055】

前記ステップS4は具体的に、ユーザーが異なる選別条件に応じて、標準的な医学用語クエリユニットによって対応する用語を選別することと、ユーザーが用語編集ユニットによって標準的な用語の追加、削除及び/又は補正、標準的な用語の同義語の追加及び/又は削除、標準的な用語の属性情報の補正を含む用語エンティティの編集を行うことと、用語監査者が用語監査ユニットによって用語監査結果の判定を行うことと、ユーザーが用語マッピングユニットによって異なるソースの医学データに対して標準化マッピングを行い、マッピング不可能なデータに対して推奨キューを構築し、用語管理者による用語のワンクリックマッピング及び推奨キューの形成を確保することと、を含む。

40

【0056】

該モジュールは医学用語標準化管理プラットフォームソフトウェアを構築する。モジュールは標準的な医学用語クエリ、用語編集、用語監査、用語マッピングの4つの部分に分けられる。

【0057】

用語クエリ部分は、ユーザーのプレビューデータを提供し、標準的な用語の可視化クエリを行い、ユーザーが異なる次元の選別条件に応じて対応する用語を選別することができ

50

、それと同時にユーザーログインインターフェース入口及び用語管理インターフェース入口を提供する。該標準的な用語のトップレベルカテゴリ、同義語、親ノード用語、サブノード用語、属性情報を含むが、それらに限らない標準的な用語の多次元情報を表示するとともに、関連用語詳細インターフェースに入る入口を提供する。

【0058】

用語編集部分は、ユーザーが可視化インターフェースを介して標準的な用語の追加、削除、補正、並びに標準的な用語の同義語の追加及び削除、標準的な用語の属性情報の補正を含む用語エンティティの編集を行う。

【0059】

用語監査部分は、用語の更新品質を確保するように用語監査者が第三者監査を行うことに供する。それと同時に、該ページから現在の監査対象の用語をプレビューすることができる。監査詳細ページに用語の具体的な更新内容を表示し、該ページには同時に監査者が用語監査結果の判定を行う。

10

【0060】

用語マッピング部分は、ユーザーが異なるソースの医学データに対して標準化マッピングを行い、4.3における一部の過程を可視化し、マッピング不可能なデータに対して推奨キューを構築し、用語管理者による該プラットフォームにおける用語のワンクリックマッピング及び推奨キューの形成を確保し、管理者が用語ライブラリにおける参照データを反復するものとし、用語ライブラリを更新・拡張する効率及び性能を向上させる。

【0061】

以上の説明は本発明の好適な実施例であって、本発明を制限するためのものではなく、当業者であれば、本発明に対して種々の変更及び変化を行うことができる。本発明の主旨及び原則内に行われるいかなる補正、等価置換、改良などは、いずれも本発明の保護範囲内に含まれるべきである。

20

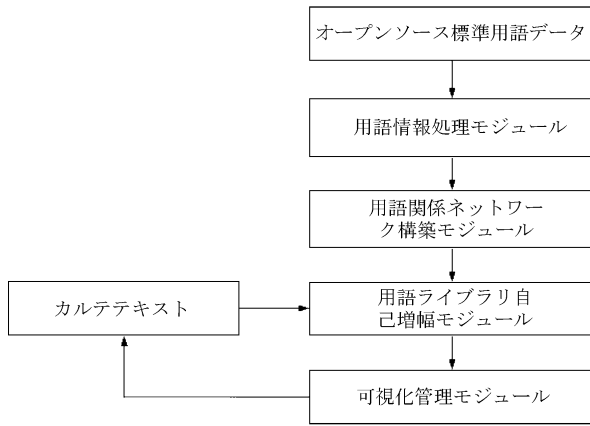
30

40

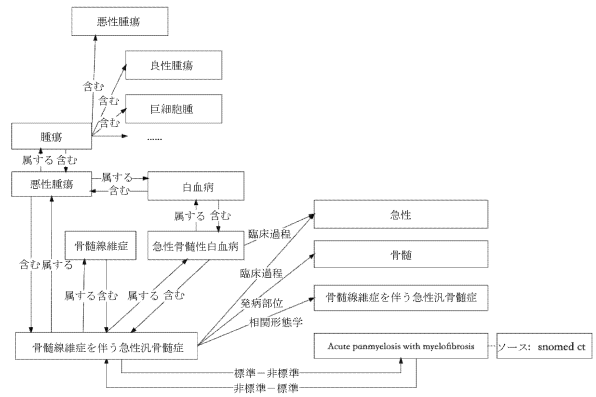
50

【 図面 】

【 図 1 】



【 図 2 】



10

20

30

40

50

フロントページの続き

- (72)発明者 田 雨
中国 3 1 1 1 2 1 浙江省杭州市余杭区中泰街道科 創 大道之江 実 験 室
- (72)発明者 周 天舒
中国 3 1 1 1 2 1 浙江省杭州市余杭区中泰街道科 創 大道之江 実 験 室
- 審査官 中元 淳二
- (56)参考文献 特開 2 0 0 9 - 1 2 8 9 6 8 (J P , A)
米国特許第 0 7 6 1 0 1 9 2 (U S , B 1)
中国特許出願公開第 1 1 0 9 2 9 5 0 3 (C N , A)
- (58)調査した分野 (Int.Cl. , D B 名)
G 0 6 F 4 0 / 0 0 - 4 0 / 5 8
G 0 6 F 1 6 / 0 0 - 1 6 / 9 5 8
G 0 6 Q 1 0 / 0 0 - 9 9 / 0 0
G 1 6 H 1 0 / 0 0 - 8 0 / 0 0