

[別添 3]

平成 29 年度厚生労働科学研究費補助金 政策科学総合研究事業

(臨床研究等 ICT 基盤構築・人工知能実装研究事業 総括研究報告書)

カルテ情報の自動構造化システムと疾患数理モデルの逐次的構築，及び，自動構造化機能を有した
入力機構の開発

研究代表者 荒牧英治 奈良先端科学技術大学院大学 研究推進機構

研究要旨

電子カルテは患者情報が全て記録されているものの，非文法的かつ断片化した表現が多く自然言語処理を応用した利活用は困難であった．これを二次利用するため申請者等（申請者荒牧及び分担者河添が所属する研究室主宰者の大江ら）は，2008年から電子カルテから医療用語の自動抽出及び自動コーディングを行う研究に従事してきた．その成果は，日本内科学会の症例報告検索システムなどとして実用化され，現在も用いられている．本研究は，電子カルテの二次利用のさらなる実用化に向けて問題となる次の2つの課題を解決する．

(課題1) 実用化可能な解析精度の達成

(課題2) 電子カルテに組み込み可能な実装の開発

若宮翔子（奈良先端科学技術大学院大学 研究推進機構・博士研究員）

河添悦昌（東京大学医学部附属病院 企画情報運営部・講師）

支援ツールの構築を目的とし，入力支援ツールを開発した．

B. 研究方法

B-1. 病名のサジェスト表示による入力支援

本稿で述べるテキスト入力支援ツール(図1)は，IMEによる仮名漢字変換後の入力された単語，文字の接頭辞から推測される病名，症状名をサジェストし，提示された選択肢から項目を選ぶことで病名の入力を補完する機能を提供する．サジェストの表示は，IMEでの仮名漢字変換後に行うため，基本的にどのIMEとも共存して利用可能である．提案するツールはテキストエディタとしての利

A. 研究目的

医療の現場などで作成される文書の日本語入力に関しては，非文法的な表現，低頻度の複雑な複合名詞，省略型の多用などの特徴があるため，既存の入力方法だけでは，ユーザーに過度に負担を強いることも少なくない．このため，本研究では，病名など複雑な固有名詞を多く含む医療文書作成を支援する知的なテキスト入力

用を想定し、編集後にテキスト文書として保存または、クリップボード経由で他のアプリケーションへデータ転送する利用形態を想定している。

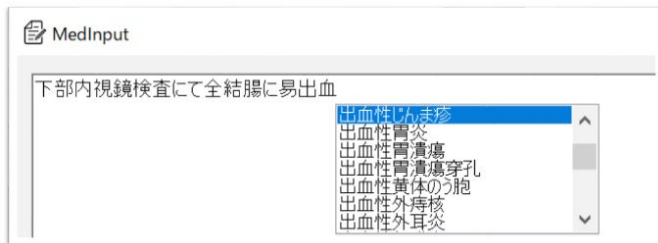


図1 予測病名リストの表示

B-2.病名辞書の構築

辞書作成のソースデータは病名の一覧が記載されたテキストファイルである。病名の一覧データはMEDIS（医療情報システム開発センター）から提供されているものを用いた。このテキストデータからトライ構造による文字ベースの辞書を構築した。トライは、辞書に登録する各見出し語の共通接頭辞を併合することにより構築される木構造である。病名検索では、トライの根節点から葉節点に向かって、検索文字列の各文字を先頭から1回たどるだけで、入力文字列の先頭から始まる全ての接頭辞を探索することができる。従ってトライ全体に格納されている見出し語つまり病名の総数に関係なく、検索文字列の長さに比例した計算時間で検索が終了する。

B-3. 入力サジェストの提示

病名のサジェストは、図1に示すリストボックスウィンドウがポップアップで表示される。ユーザーはリストから、入力しようとしている病名を選択することで、その該当部分に選択された病名が挿入される。該当病名がない場合はEscキーにより、ポップアップを閉じることができる。なお、ポップアップはモードレスで動作しているため、ポップアップ表示中も文字入力を継続して行うことが可能である。

B-4. 入力サジェストの表示制御

入力サジェストの表示はテキストが更新される毎に判断を行う。判断には、キャレットの直前にある文字列を取得して、その文字列をキーワードとして、トライ辞書の検索を行う。

しかし、文字（単語）が入力される毎に、辞書検索を行いヒットがあった場合にサジェストを行うことには問題がある。例えば、辞書には以下のように「た」で始まる病名が含まれているが、入力文が「～した」など動詞の末尾で終了した場合などに、最後の助動詞“た”を病名の開始と判断し、以下の病名がヒットしてしまうという課題がある。

■たこつぼ型心筋症

■たこ壺型心筋障害

上記のように、入力された文字(単語)が病名の可能性が低い場合、入力サジェストの提示を抑止する必要がある。

B-5.入力された単語(文字)が病名の一部であるかの判定

入力サジェスト提示の判断を行うため、提案手法は、既存の文字ベースの病名抽出ツール「MedEX/J」を利用する。つまり、編集集中の文を病名抽出ツールに適応し、入力された単語(文字)が病名の一部かを判定する。予備調査の結果、文が未完結の場合でも病名抽出を高い精度で行えることを確認した。

なお、本手法は、病名(標準病名のみ)を扱い人名やIDは扱わないため、個人情報を不用意に削除・保管することはない。また、機械学習モデル内部においても、単語レベルで情報を保持しており、特定の個人に結びつくこともない。

C. 研究結果

C-1. 病名の一括抽出と病名の標準化

病名抽出機能を入力された文章全体に対して適応することで、文章中に含まれる全ての病名

の確認と、病名の標準化を行うことができる。

図2に病名の一括抽出と事実性の判定を行った例を示す。抽出された陽性の病名は赤で、陰性の病名は青で表示される。処理方法としては、初めに、文章を文に分割し、1文ごとに前述の病名抽出ツールにより病名を抽出している。抽出された病名は同時に自動的に陽性か陰性の極性判断がなされる。

C-2. 抽出された病名の標準名への変更

マウス操作により、抽出された赤字または青字の病名を右クリックすると、その病名に対する標準名およびICD-10コードが表示される(図3)。ユーザーはここでサジェストされた標準病名に置き換えることが可能である。

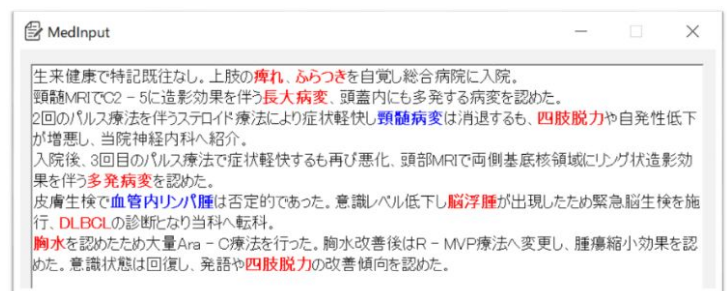


図2 病名の一括抽出と、陽性(赤)/陰性(青)の判断

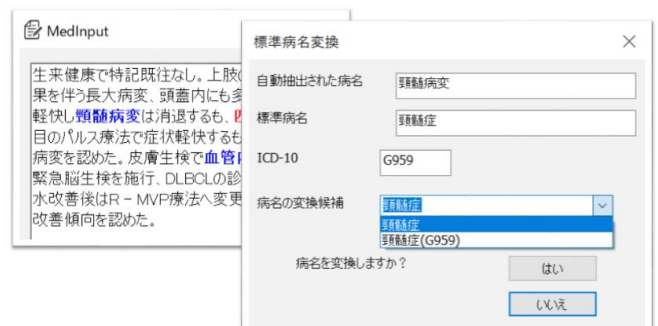


図 3 病名の ICD-10 コードの確認と標準病名への変換

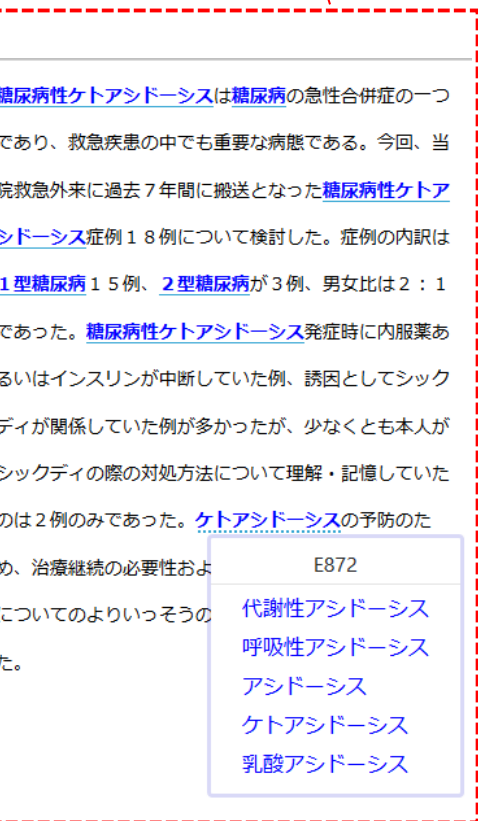


図 4 カルテ入力パレットのインターフェイス

C-3. Web版入力支援ツールの開発

本研究では、さらに入力支援ツールの操作性を高めるために、Web版の入力支援システム(以下、カルテ入力パレット)を構築した(図4)。

図4に示すとおり、Web上の入力枠に医療テキストを入力し、解析ボタンを押下することで極めて簡易に使用することができる。

図4の入力例では、糖尿病に関する症例報告を入力している。傷病名と判断した箇所が信頼度に対応した色で強調表示され、当該箇所にマウスポインタをあわせることで、ICD10コードに対応した標準病名が提示される。提示された標準病名をマウス選択することで、容易に傷病名を変更することが可能である。なお、入力枠の右側に簡易的な人体解剖図が示されており、入力テキスト解析により特定された傷病名に対応した部位が点滅するという直感的な可視化機能を備えている。

D. 考察

本ソフトウェアのユーザービリティの評価として、数名の被験者にこのツールを使用した時の医療テキストの入力時間を計測した。その

結果，このツールを用いた場合とそうでない場合では，入力時間の違いに有意差は見られなかった．しかし，テスト時には事前に被験者によるツールの使い方の練習は全く行わなかったもので，ツールの使い方に慣れてくれば，ある程度入力時間の短縮は期待できると予想している．

E．結論

本研究では病名支援入力ツールを開発した．入力時の支援だけでなく，本ツールはバッチ処理により，文章から一括して病名を抽出する機能も有し，抽出された病名のICD-10コードの確認と病名変更の機能も提供している．今後は，医療従事者により，本システムを評価を実施する予定である．

F. 健康危険情報

特になし．

G．研究発表

1. 論文発表

- 該当なし

2. 学会発表

- Kaoru Ito, Hiroyuki Nagai, Taro Okahisa, Shoko Wakamiya, Tomohide Iwao, Eiji Aramaki: J-MeDic: A Japanese Disease Name Dictionary based on Real Clinical Usage, LREC 2018. (Miyazaki, Japan)
- 矢野憲，岩尾友秀，荒牧英治：MedInput: 病名の自動予測補完による医療テキスト入力支援ツールの構築，言語処理学会 第24回年次大会，2018．

H．知的財産権の出願・登録情報

特になし