



Title	医療における自然言語処理インターフェースに関する研究
Author(s)	竹村, 匡正
Citation	大阪大学, 2003, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/43882
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

氏 名	たけむらただまさ 竹村匡正
博士の専攻分野の名称	博 士 (保健学)
学 位 記 番 号	第 17709 号
学 位 授 与 年 月 日	平成 15 年 3 月 25 日
学 位 授 与 の 要 件	学位規則第 4 条第 1 項該当 医学系研究科保健学専攻
学 位 論 文 名	医療における自然言語処理インターフェースに関する研究
論 文 審 査 委 員	(主査) 教 授 村瀬 研也 (副査) 教 授 大和谷 厚 教 授 稲邑 清也 教 授 芦田 信之

論 文 内 容 の 要 旨

[目 的] 医療情報システムにおいて、入出力の合理化、また蓄積された情報を多角的に入出力することを考慮した場合、自然言語処理インターフェースは極めて合理的である。これらを実装した場合の、言語知識を抽出する方法論の確立に向けて、その利用における可能性を提示することを目的とする。

[方 法・結 果]

1. 新しい Medical Record のための分類法・並びに情報構造の確立

自然言語処理を踏まえ、認知科学的な観点から医療行為者の典型的状況変化に基づいて記録が可能になる新しい分類法（格フレーム分類法）、ならびにそれを記述するための構造（Situation-Oriented Medical Record）を考案し、これに基づく診療記録ナビゲーションシステム、また、これを実装した電子カルテシステムのプロトタイプを構築した。

2. 言語知識取得のための大規模文章情報（Corpus）の利用

次の課題となったのが、自然言語で書かれた患者の情報を如何に自動的に処理するのかということであった。その実現のための知識取得を目的として、実際に現場で記述されている多くの文章を集積し、そこから種々の言語知識を取得することを考案した。そして、医療の文章を分析する上での最小構成単位である用語集（和・英・対訳含めて約 30 万語）を作成した。

3. Corpus の作成

コーパスには、一般的な言語知識を内包しているものも重要であるとともに、ある現場における特定の情報抽出も効果が期待できる。そこで、医療においてシステム化が困難であり、より多く自然言語が用いられている現場でのコーパス作成、またこれらの現場における遠隔コミュニケーションの適応を踏まえ、精神病院とその附属施設における医療従事者のコミュニケーション、また、看護領域における在宅患者のフォローアップ、調査などをネットワーク上で可能にするシステムを開発し、運用した。

また、インターネットを用いたコーパス作成法として、巡回エンジンを用いて医療文章を自動的に取得するシステムを開発、また、大阪大学医学部付属病院で記述された一年分の放射線診断レポートや、医療情報学連合大会などの電子論文集なども用いて各分野のコーパスを作成した。

4. Corpus からの知識取得

コーパスからの知識取得・利用として、文章の価値を判定するという観点から、大量の文章情報においてユーザーが要求する文章を合理的に検索するという場面を想定し、英英訳辞書を用いて、説明用語の頻度から算出した重み付けを各文章に与えることを考案し、これを先の巡回ロボットシステムと組み合わせて Web 上検索エンジンとして応用的に実装した。また、放射線診断レポート（2000 年・日本語記載 48564 件）、に対して、日本語の医療用語集（239903 語）を用いて用語解析を行い、利用されている用語の基本的統計量の算出とともに、言語知識として語句の分野依存性に着目しこれを抽出した。具体的には、モダリティ・撮影部位別（69 種類）に対して、各分類で用いられている用語に対して「TF・IDF 重み付け」を応用して重要度を数量化し、これに基づく重要語を算出した。これには、例えば「X-TV・胸部」の場合なら、「気管支鏡、気胸」などが抽出された。また、これらの重要語を用いてテストデータ（1999 年放射線診断レポート・701 個）の自動分類を試み、実際の分類と照合した。結果、テストデータに対して文字数の多い（400 字以上）のレポートでは、各重要語のもつ重要度の和を各分類において上位 80 個採用した場合による自動分類は、80%の正答率であった。

[総括]

格フレームや Situation-oriented medical record という情報構造に基づいて構築した分類法は、他の分類法以上に自然言語を扱うことが可能である。また近年の計算機の発展により、実際に利用されている言語そのものを大量に扱うことが可能になり、コーパスを利用するという方法論を提示できた。そして、これらの方法論によって得られた言語知識は、医療分野においてこそ独自の利用法が多く考えられる。というのは、医療分野は MML (Medical Markup Language) や DICOM-SR などに見られるように、多くの規格化・標準化が進められているためである。すなわち、これらの規格の存在によって、実際に記述される文章の内容の関係をこれら言語分析に基づく知識によるブリッジングがしやすく、またその効果は極めて大きいものと考えられる。

今後は、言語解析の方法論の確立・精緻化とともに、現場での自然言語入力情報と、各規格との関係を明らかにし、自動自然言語処理に繋げてゆく所存である。

論文審査の結果の要旨

本論文は、医療情報システムが本来持つべき機能である入力情報の網羅性と、入力業務の効率性の両方を満たすための自然言語入力インターフェースの実装への方針について述べている。具体的には、医療の情報であっても結局その情報を扱うのはヒトであることから、ヒトの認知に即した新しい情報モデルを提示し、また、実際に医療現場で入力された文章情報である放射線読影レポートを大規模に集積し、統計的な自然言語処理の手法を導入した上で、システムによる自動的な内容抽出・分類を試みている。このような試みは新規性に富み、またこれからの医療における自然言語処理研究における主流となることが予想される。よって、本論文は学位の授与に値すると考えられる。