

診療記録記載用語の標準化と構造化に関する研究

主任研究者	里村	洋一（千葉大学医学部附属病院）
分担研究者	木村	通男（浜松医科大学）
	廣瀬	康行（琉球大学医学部附属病院）

厚生科学研究費補助金（情報技術開発研究事業）
総括 研究報告書

診療記録記載用語の標準化と構造化に関する研究

主任研究者 里村洋一 千葉大学医学部附属病院 教授

研究要旨 : この研究は、電子化した診療記録を医療機関関係者間で共有するためのデータの標準化を目標としている。診療記録に用いられる用語集に構造を与え、コンピュータによって記録を作成するに当たって、効率よくかつ統一性のあるデータが作成できるよう、インテリジェンスのある辞書に編成しようとするものである。対象を、疾病・症状・診察所見・検査所見に絞って、用語の収集と構造化を試みた。

内外の用語集（MEIDとSNOMED）を利用しこれらから用語の抽出をおこない、階層構造を与え、さらに他の用語との関連を記述する手法を研究した。階層構造や関連の記述には、SNOMEDの区分法と基本用語集を参照して用語の構成を分析した結果を用いた。また、用語辞書編集のツールとして、類似語検索システムを作成した。

分担研究者

木村 通男

浜松医科大学附属病院
医療情報部 教授

廣瀬 康行

琉球大学医学部附属病院
医療情報部 教授

まで閉鎖的と言われてきた医療の体質を強く反映しているからである。

申請者らは、長年、医療情報システムの開発と普及に当たっており、早くからこの問題に挑戦してきた。既に、病名（診断名）については、「ICD10対応標準病名集」の形で、その研究成果を応用した。また、他の医療用語、たとえば医療行為（手術や処置）や病理診断、検査所見などについても、その構造や用語法についての研究を展開してきた。

研究課題は、その一つである症状・診察所見・検査所見について、用語の標準化に必要な要素の抽出と用語の選択、これら用語の相互関係の構造化に挑戦することである。この成果によって、電子化診療録に記載されるデータの一貫性と統一性を得ることが期待できる。

A. 研究目的

情報処理システムが医療の効率を上げ、質の向上をもたらすと期待されているが、それは、個々の医療行為の情報化そのものによってではなく、記録された情報の交流とその共通利用によって実現されるものである。

情報の共有の前提として、データの標準化や通信手法の規格化、さらにそれに伴う危険を回避するための安全が要求される。中でも、データの標準化は最も基本的な要素であって、かつ最も困難な課題である。なぜならば、通信や安全性の問題に比べて、はるかに要因が複雑であり、また、医療の内容に密接しており、これ

B. 研究方法

本研究では、症状と診察や検査の所見を記載するための用語集開発を行う。そのために、次の3段階に分けて研究している。1) 内外の用語集や臨床研究誌から、この領域の用語を収集し、選択して用語の抽出を行

い、次いで、2) これら用語の階層構造を定義して構造化する。さらに、一部の用語について、3) 臨床の現場での利用に供しその実用性を検証する。

本年度はこのうち1)と2)の一部を行うこととし、そのためのソフトウェア開発に主点をおいて研究した。

具体的には、1) 基本語(用語を構成する基本的な意味単位)の抽出と整理 2) 類似語抽出のシステム の二つである。

C. 研究結果

基本語の抽出：既に抽出された日本語の基本語については、研究者らがこの数年来開発してきた11000余りの病名基本語がある、この内、今回の症状や所見に属する用語が266ありこれを優先的に採用することとした。

一方、これを補うために、SNOMEDのFUNCTION軸から、基本語(英語)を抽出する作業をおこなった。現在約400の基本語が抽出されているが日本語訳の付与が進行中である。

C. 1 構造化：

用語を構造化する方法には、大きく分けて2通りある。一つは、ある一定の視点から対象の事象を分類し、その分類概念に基づいて用語を階層的に配置するものである。たとえば、ICD-10準拠の病名集はこの例である。もう一つは、用語をその構成する最小の意味単位に分解し、用語をその構成要素の組合せとして理解し、どの構成要素のからも検索可能なような意味ネットワークを構成するものである。前者は、分類概念が使用者の間で共通のものとして理解されており、かつ、分類の目的に対応した利用が行われる場合に便利である。後者は、構造が複雑で統一的な分類概念が当てはまらないことから、外見では理解されにくい、個々の利用の目的に応じた情報検索を可能とする。特にコンピュータを利用した検索や利用には適している。

本研究では、電子カルテ等によって日常診

療に利用される医学用語の体系化を目指していることから、後者の形の構造化を図る事とした。

C. 2 意味単位：

用語を意味単位に分解する場合には、意味単位とは何かを定義する必要がある。日本語のように主として表意文字を用いる場合には、最小の意味単位として個々の文字を用いることもできる。しかし、表意文字といえども、用字法や対象領域によってその意味が異なることがしばしばであり、一律に一文字を意味単位として扱うことに困難がある。たとえば「疼痛」の「疼」と「痛」を分離するのは言語学的には正当であっても、医療では意味がない。そこで、この研究では、最小意味単位(ATOM)を医学や医療における常識的な認識の範囲に置いて設定する事とした。たとえば、「疼痛」「痛」「痛み」「いたみ」は同義であると見なして、これらを全てATOMとする。これに対して、複数のATOMが組み合わせられた場合に、これを用語(TERM)と呼ぶ。英語ではTERMとATOMの間に単語(WORD)の概念が存在するが、日本語では必要とされない。WORDは形態学的にスペースで区切られる単位であるが、日本語ではこれに相当する区分子が存在しないからである。ただし、英語と日本語の連携を考えると、ATOMとTERMの中間にWORDの概念を挿入する必要がある。この場合、日本語のATOMをそのまま英語のWORDに対応させることができよう。なお、全てのATOMはTERMでもありうる。

C. 3 属性の記述

全ての要素(ATOM、TERM)はそれぞれの中で一定の分類が行われる。すなわち、医学用語全般を扱う場合にATOMの分類は、部位、臓器、病変、病因、症候、所見、機器、手技、薬剤、時間要素、量的要素、社会要素、接続表現、固有名詞などであるが、本研究で対象とする症状・症候の場合には、この分類は単純化され、部位、

臓器、症状、時間要素、量的要素、固有名詞などに限定される。この他の属性として、品詞（形容詞の場合は対応する名詞）、仮名表現、SNOMEDコード、ISO概念表現（ISO/TC215において検討されている医学概念の国際表現）等を持つ。

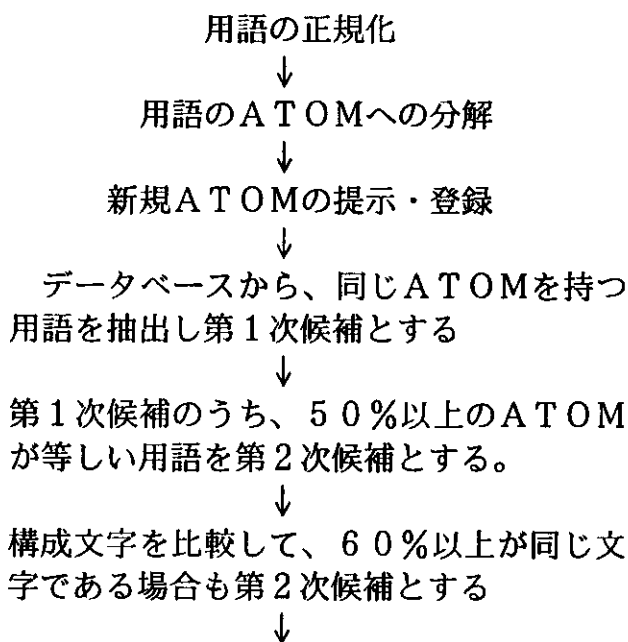
C. 4 辞書編集ツール

上記の様な構造を持った用語集を編集するためには、専用のデータベース検索システムを準備することが必要である。このシステムには以下の能力が要求される。

- 1) 用語を既知のATOMに分解する機能（パーシング）
- 2) 用語の表現を正規化する機能
漢字の統一
1バイト文字の排除
スペース、括弧等記号の整理
- 3) 新規のATOMを登録する機能
- 4) 類似の用語を検索する機能

1)と2)については、研究者等がこの数年間で開発した正規化プログラムが利用できる。本件研究では、用語の類似性を検索するとともに、新規のATOMの発見を支援し登録する機能を持ったシステムの開発を行った。

現在作成されているシステムの処理過程は以下の様である。



類似語の一覧を表示

通常、数万に上る用語を検索しこれらの中から類似性のある用語を見つけ出すのは、骨の折れる作業である。人手で行う場合には、数十人が分担して何ヶ月もの期間をようしてやっと可能な作業である。

本システムを利用することによって、一用語に対して類似のものを抽出するのに数秒しか必要としない。

D. 現段階での考察

辞書編集は最終的に人手、しかもそれぞれの領域の専門家を要する作業である。従来は、この作業の第1段階（リスト作成）から人手に頼り、しかもそれぞれの担当者が必ずしも統一的な基準で行うとは限らない環境で作業が行われてきた。

これでは、統一性のある辞書編成は大変な難事業となってしまう。そこで、コンピュータを利用した能率的かつ一貫性のある辞書編集を支援するためのツールを開発する事が、用語の構造化を達成するための前提となると考えた。

現時点での類似性判定は、形態学的比較（同一文字の数やATOM構成）を中心としたものであるが、SNOMEDの構造や、品詞の情報を参照することができれば、さらに精度の良い作業を支援することができる。辞書編集ツールの精度は、編集作業そのものの結果を反映することから、回帰的な意味を持っている。

E. 結論

本年は、2年計画の初年度にあたり、結論に至る成果は得られていないが、ATOMの抽出とこれを利用した類似語検索システムが半ば完成した。次年度には、この成果を生かして、本格的な辞書作成実験にはいる。

F. 研究発表

医療情報学に投稿予定 2000年度

厚生科学研究費補助金（情報技術開発研究事業）
分担研究報告書

診療記録記載用語の標準化と構造化に関する研究

分担研究者 木村通男 浜松医科大学医療情報部 教授

研究要旨： SNOMED International はもっともよく整備され、多軸記述での記載をするため、個々の軸では単一の木構造をもっている。一方、診療記載の項目について、知的検索のためには、平板なコード集ではなく、木構造をもったものとして扱える必要がある。本研究では、医用画像伝送規格である DICOM が、その画像の部位やレポートのために、SNOMED の部分集合として制定した SNOMED-DICOM-Microglossary が、構造化された記述形式であるため、これを素材として、様々な記述レベルのニーズに対応できるコード集とはどのようなものかについて考察した。

A. 研究目的

診療録の記載内容について、後に意味的な検索を行なうことは、診療内容の向上に益するばかりでなく、情報交換による病診連携、病院管理的価値など、様々な点で有用である。しかしこの実現のためには、記載の用語がコード化されていることが一番である。しかし、上記の必要性のうち、診療情報交換の例を取っても、交換すべき情報のレベルは、初診診療所から専門病院への場合、専門医間での場合と、様々である。用語記述の深さに対するこのような様々なニーズに対応するには、用語のコードが構造化されている必要がある。本研究では、こういったコードが意味的に持つべき構造についての考察をおこなった。

B. 研究方法

主任研究者の昨年度の研究でも用いたように、SNOMED International は、もっとも広範囲に、かつ網羅的に作成された医学用語集であるが、これはそのままでは、平板なコード集である。一方、画像伝送規格である DICOM 規格は、その画像の部位やレポート記載のために、SNOMED を簡略化し、構造化した SNOMED-DICOM-

Microglossary を作成し、公開している。今回はこれの内科的記述について、様々な用途で用いることが出来るか、その構造を検討した。

C. 研究結果

SNOMED-DICOM-Microglossary は、近年 SNOMED 制定母体が提唱している SNOMED-RT に準拠している。SNOMED-RT は完全な網羅的木構造を形成しているので、すべての詳細記述には、必ず親（より一般的な表現）が定義されていた。従って、様々な用途に応用可能であることが判明した。

D. 現時点での考察

専門家が求める記述は、極めて詳細なものであり、SNOMED 的記述で言えば「冠動脈左回旋枝の75%狭窄による心筋前壁の虚血性壊死」といった記述レベル（粒度）を求められるが、これを逆紹介で診療所に返す場合は「前壁心筋梗塞」で十分かもしれないし、医事的には「心筋梗塞」で十分かもしれない。しかし、通常の平板なコードではこれらは同一のレベルに置かれるため、親子

関係（意味的包含関係）は利用できない。

これは、用途ごとに別の表現の病名を記述する必要を意味し、これでは実用的な病名登録システムとは言えない。ICDは統計的分類を旨としたものであるため、ある程度の構造は、コード番号で見取れるが、これは頭数桁だけのことであり、臨床的に必要なのは後ろ数桁のレベルであるのに、これらの部分は構造化されていない。これが、ICD準拠の病名集をそのまま臨床病名登録に利用した際の問題点であると考えられる。従って、よりよい病名登録を実現するには、病名集の網羅性だけでなく、構造に依る一意性（親が必ず一つ、唯一存在すること）を病名登録システムが内部情報として持っていることが望ましい。

E. 結論

構造化された情報を持つ SNOMED-DICOM-Microglossary は、病名記述の様々な粒度のニーズに対応出来るものであることが判明した。今後は病名データベースは、項目を平板なコード集として持つのではなく、木構造を持ったものとして扱える必要があると考える。

F. 研究発表

1. 論文発表

Kimura M., Ohe K., et.al., A Patient Information Exchange Guideline using MML, HL7 and DICOM, International Journal of Medical Informatics, 51:59-68, 1998.

2. 学会発表

Kimura M., et.al., Patient Information Exchange Guideline MERIT-9 -Using Medical Markup Language MML, The 9th Annual World Congress on Medical Informatics (MEDINFO-98), pp.433-437, 1998, Seoul, Korea.

厚生科学研究費補助金（情報技術開発研究事業）
分担研究報告書

診療記録記載用語の標準化と構造化に関する研究

分担研究者 廣瀬 康行 琉球大学医学部附属病院 教授

研究要旨：

診療記録記載用語を標準化・構造化するための前提として、本研究でいうところの ATOM（意味最小単位）や TERM の収集と整理が不可欠である。よって用語辞書編修ツールが開発されたわけだが、このツールは幾つかの機能を併せ持つこととなった。すなわち、SNOMED International と MEID との間の用語対照機能、SNOMED における木構造に即した日本語の ATOM や TERM の入力登録機能、ATOM や TERM の属性表示機能、そして類語検索列挙機能である。本研究では、SNOMED と MEID を参照しつつ用語辞書編修ツールにより用語収集を実施しその実用性を検証するとともに、用語間の意味ネットワークや概念関係の表現方法について考察した。

A. 研究目的

研究計画における分担研究者の役割から、本年度は用語辞書編修ツールを用いて、1) 診療記録記載用語のうち症状や所見を表現する ATOM や TERM を実際に登録してみることに、2) そしてこのツールの機能を評価すること、の二点を主眼とした。またこれらを実施しつつ、このようにして収集登録した用語群が、電子化診療録システム的环境において十二分に活用されるかについて、若干の考察を試みた。

B. 研究方法

用語辞書編修ツールは MEID を母辞書として用語を検索表示する機能を有している。そして、新しい用語（ATOM や TERM）を類語もしくは類似語として、MEID 外辞書に登録することができる。その際に SNOMED と対照させることができるので、SNOMED における木構造に即しつつ、新たな用語を入力登録できる。用語収集は、このような環境において実施した。またこの過程において、どのような用語が・このツールを用い

た結果としてどのような用語と結びつけられていくのかを観察することなどによって、用語間の意味ネットワークや概念関係の表現方法、換言すれば、その様な属性の具備方法について検討した。

C. 研究結果

使用した用語辞書編修ツールでは、新規用語の入録登録の際に MEID を母辞書とするとともに、これを登録の際の「柱」としている。このため、SNOMED の用語を「柱」とした入力登録はできない。よって SNOMED の木構造を直接的に再利用することができないという不利な点を抱えている。いずれにせよ新しい用語は、それが ATOM であれ TERM であれ、登録の際に、既存の辞書を、いわば「柱」として必要としている環境となっている。このことは、新しい用語と他の用語との「関係」は、登録の際の「柱」とされた辞書に内包されている既存の用語関係に、結果的として強く影響されることになる。ツールにおけるこのような作用もしくは副作用は、一概には

良し悪しを断じることはできない。というもののその価値判断は、意図する構造、あるいは、構造のあり方に関する展望等に依存するからである。またこのような作用もしくは副作用こそが、このツールの編集作業とその結果の反映という回帰性を保障している、とも言えよう。しかし標準的な構造の構築を目指す場合、何らかの障害因子を生じせしめてしまう危険性を否定することができなかった。同一の ATOM もしくは TERM は、常に同等の様態で使用されるとは限らないという理由によって、一つの用語の類語もしくは類似語は、広がりをもつこととなっている。つまり、ある程度の・小さな「文脈」が、用語間の類性や類似性に影響することとなる。しかしこのツールでは、用語登録もしくは類似語の際に、この点に関する配慮機能は備えていない。用語の属性に関しては、現時点では、発生頻度のみが表示できる。しかし電子化診療録において有効活用できるためには、特に症状や所見用語においては、類語や類似語という観点もさることながら、まさに共起性に依る群形成を反映する機能も必要となる。もっともこのような群形成には単に共起性のみならず、学術的に合理的な意味ネットワークの存在が立証されることも必要であろう。いずれにせよ、現段階ではそこまで至っていない。

D. 考察

ある用語がなんらかの文書において「使用された後」に検索分類対象とされる際における用語探索とその目的のための分類や構造化と、なんらかの文書のなかにある用語を「使う」ためにその用語を探索する際に使用する機構を下支えするような構造化もしくは用語間構造は同一であると、一般に安易に信じられている。しかし筆者は、両者間の差は動性の程度のみでないであろうと考える。むしろ相異なる知性の運動である相異なる探索機構こそが、それらが求めている・必要な用語間構造に差を生じせしめているように思われ

る。もちろん両者は完全に分離独立しているわけではないが、これまで体系化されてきた辞書や分類は、そのほとんどが前者を目的としており、その意味において静的であり、そして美しい。しかし電子化診療録において有効活用できる用語体系とは、まさに後者を求めている、と考える。この後者を目的として、用語間の意味ネットワークや概念関係の表現法を整え構造化体系を構築していくことまた、その体系は、体系に関する一般的なイメージからすれば、いわば捻れた環のように見えるのではないかと思われる。その具体的な例は、まさに前項に記した結果のとおりである。しかしこのことは、後者を目的とする構造化体系を構築することが不可能であることを意味しているのではない。ただしツールの設計には、注意が必要となる可能性が高い。

E. 結論

以下に列挙した点が明らかになった：

- 1) 新用語の取り込みには解釈の軸となるべき体系が必要であること（この知識世界の座標軸を、分類やコード体系における軸と区別するために「柱」とよんだ）
- 2) 「柱」は単一とは限らないこと
- 3) 新用語を取り込む際に用いた「柱」に依って、その用語と他の用語との連関が、時として半強制的に決定されうること。
- 4) その用語と他の用語との関連を、その「柱」のなかで決定するには、（おそらくはその柱の中に出現する他の用語との）共起性をも同時に知る必要のあること
- 5) その用語を「検索対象とする」のではなく「思い出して書き留める」ためには、共起性における群形成とともに、意味ネットワークによる裏打ちも必要なこと。

E. 発表

2000年に医療情報学会誌に発表予定