

されている。ドイツ語版、スペイン語版、イタリア語版、オランダ語版、ポルトガル語版、日本語版は、UMLSにも収容されている。

(2) 歴史

1990年、医薬品の品質、安全性、有効性に関する承認審査資料などに関する国際的なハーモナイゼーションを目的とした ICH (International Conference on Harmonisation of Technical Requirement for Registration of Pharmaceuticals for Human Use) が開かれ、1994年には ICH の場で医薬品規制に関わる国際医学用語集 (MEDical Dictionary for Drug Regulatory Affairs—MEDDRA) 開発が決定し、1997年には日英語の MedDRA v.2 (Medical dictionary for Drug Regulatory Affairs) が公開された。元々は WHO の副作用用語集 WHO-ART、米国 FDA の COSTART、日本の J-ART、と言った副作用用語を基礎として MedDRA に一本化したものだが、さらに適応症や手術名なども加わった医学用語集となっている。

(3) 入手並びに利用

MedDRA 自体は国際製薬団体連合会 (International Federation of Pharmaceutical Manufacturers and Associations—IFPMA) が所有権を有しており、MedDRA の利用契約は①コア契約 (法人の売り上げ規模による区分)、②ベーシック契約、とから構成されており、コア契約のユーザーには MedDRA への用語追加要請の権限が与えられている。

MedDRA は 5 階層の構造となっており、米国 FDA の副作用報告システム、EU の医薬品監視システム、日本の医薬品安全性報告などに採用され利用されている。

(4) 改訂

当初は年 4 回の改訂であったが、現在は年 2 回、定期的に改訂され、3 月と 9 月にリリースされている。改訂作業のためには 1997 年から国際維持管理機関 (Maintenance and Support Services Organization—MSSO) がおかれており、一貫して管理されている。

MedDRA に対して日本語訳を付与したものが MedDRA/J でこれを管理する組織が JMO (Japanese Maintenance Organization)、その

JMO は (財) 日本公定書協会 JMO 事業部におかれており、MedDRA/J のメンテナンスや利用者への情報提供、セミナー開催なども行っている。国内での利用は JMO と契約して行うこととなる。

2. 1. 5 ICD (International Classification of Diseases)

(1) 概要

ICD は 1900 年に国際死因分類表として制定され、WHO (World Health Organization) によってオーソライズされるもので、当初は International Classification of Diseases これを略して ICD と称していた。ほぼ 10 年毎の改訂とされており、最新版は ICD-10、正式名称も International Statistical Classification of Diseases and Related Health Problem となっている。ICD にはファミリーと呼ばれる関連規格が多数有り、また、ICD を加工した規格も作られ使われている。ICPC (International Classification of Primary Care) などは代表的なものと言え、これは、ドイツ語、フランス語、スペイン語、オランダ語、イタリア語、スウェーデン語、フィンランド語、デンマーク語、ポルトガル語の他にもバスク語やヘブライ語版まで揃っている。がんについては、ICD-O (International Classification of Disease—subclassification for Oncology) が使われている。

ICD-10 には、アメリカ版 (ICD-10AE) やオーストラリア版 (ICD-10AM) 等もある。

(2) 歴史

医学・医療の分野で長い歴史のある分類表であり、米国で死因統計に利用していた分類表を国際的に活用する形で始まり、途中からは WHO 主導で維持・改訂されるようになった。本質的に疾病や死因に関する疫学研究で役立つもので、あくまで統計用の分類体系であった。この為臨床目的には必ずしも十分と言えず、かつ用語集やシソーラスとは異なるものであったが、ICD-9 と ICD-10 には本来の分類体系から用語集・コードブックに近づけた ICD-9CM (Clinical Modification)、ICD-10CM なども開発され使われている。

(3) 入手並びに利用

オリジナルのファイルは WHO のサイトからダ

ダウンロードできる様になっており、各国語版はそれぞれの国で管理されている。我が国では厚生労働省統計情報部によって翻訳され、日本語版「疾病及び関連保健問題の国際統計分類」として印刷物で市販されている。

(4) 改訂

最新の ICD-10 は、1990 年の第 43 回 WHO 総会で承認されたものでアルファベット 1 桁数字 2・3 桁の数字でカテゴリー分類する形となっている。

2. 1. 6 CPT (Current Procedural Terminology)

(1) 概要

CPT は、1966 年から米国医師会で作成している 8,000 以上の医療行為や医療サービスに関するコード集であり、5 桁のコードで表示する。医師の行う行為を全てコード表現出来ることを目的としており、対象となる医療行為には、麻酔、手

術、病理学や実験から投薬まで、含んでいる。医療保険の Medicare で使われている医療コードとして HCPCS (Healthcare Common Procedure Coding System) でも採用されている。

(2) 入手並びに利用

年 1 回改訂され、毎年書籍として販売される。米国医師会からは、CPT を使ったコーディング方法、毎年の改訂箇所に関する解説、その他、コードブックを活用するための補足的な書籍なども販売されている。米国医師会では少しでも CPT が有効活用されるべく低額のライセンス制度を用意している。そして毎年、CPT 利用者のためのシンポジウムも開催しているし、参加者は生涯教育のポイントも受けることが出来る。

(3) 改訂

毎年の改訂で、最新版の 2007 年版では 8,611 のコード数となっている。

2. 2 UMLS と医学用語シソーラスに関する歴史と私見 Some Observation and History on the UMLS and Medical Thesauri¹

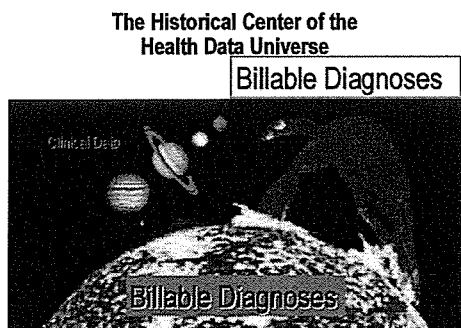
Christopher G. Chute, MD DrPH
Professor and Chair, Biomedical Informatics

Mayo Clinic College of Medicine
Rochester, Minnesota, USA

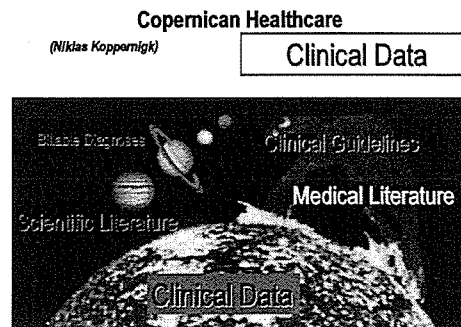
2. 2. 1 はじめに

臨床医学領域において、用語集の果たす中心的な役割を、継続的に改善していくこと作業はまさに、遺産と言える。何故なら、患者を診療し、その内容は臨床データとして、保存される。それが資料として、医療知識として、データベース化されていき、後に、ガイドラインとなり、それが、また診療に役立っていく。この循環に欠かせないものが、それぞれのデータを結ぶ用語である。用語があつてこそ、臨床診療が循環するのであると言えるからである。臨床の場面では、①診療記述、②結果研究、③持続的な質の改善、④保健サービス、⑤臨床疫学、⑥資源管理、⑦ガイドラインの作成、と流れていく。その経過と結果の関係は、確立された伝統と言える。かつては、診療報酬を中心に臨床データが、太陽のまわりをまわる惑星のように、まわっていた。現在は、コペルニクスの地動説のごとく、この立場が逆転し、臨床データを中心に、生物医学文献、臨床ガイドライン、診療報酬請求などが、まわるようになった。

(図 1)



(図 2)



2. 2. 2 医学用語と標準化

医学的概念、事象、観察、介入はどのように、表現すれば良いでしょうか？言語もそうですね。①ニュアンス、詳細、それらを自由に結合し、②時に適った、正確な、決して、時代遅れではないものにする、③自然形であり、親しみやすく、確立されたもので、あれば良いのですが、それらは、曖昧で、不正確、なおかつ予測できないものです。では、コード化に関しては、どうでしょうか？①簡潔明瞭で、正確、②構造化され、一貫性があり、よくまとまっているもの、③分析可能で、操作できるものが良い、硬く、柔軟性のない、退屈なもので、メンテナンスも行き届くものが良いでしょう。

臨床で利用される際、問題となるもの例を示します。HL7(Health Level Seven)の例ですが、HL7では、血液型を表す ABO は、コンピュータ上では、明確に区別ができません。このような例もあります。

¹本論文は、この研究班が招聘した Chute 教授が 2004 年 11 月 24 日に東京大学鉄門記念講堂で行った講演の記録をもとにしたものである。

(図 3)

Does HL7 Solve the Problem for Clinical Use?

- [From HL7 & HL7]
- ① OBX|1|CE|ABO^ABO GROUP|
|O^Type O| Equivalence not
 obvious to
 - ② OBX|1|CE|BLDTYP^ABO GROUP|
|TYPEO^Type O| computer
 - ③ OBX|1|CE|ABOTYPE^ABO GROUP|
|OPOS^Type O|

OBX|1|CE|883-9^ABO GROUP|F-D1250^Group O|
LOINC Codes Solve this Problem

用語の標準化がなくては、①保健データの比較、②データのやりとり、③調査、効果などを調べるような、二次的利用、④意思決定支援となる資源と結びつけることができなくなります。これらを満たすことが、必須条件となります。

事象対概念: 医学情報メタデータの基本原理

ヒポクラテスは、四体液説をとなえ、迷信や呪術を切り離し、科学的に医学を発展させました。アリストテレスは、皆さんもよくご存知の、三段論法で、近代論理学や分類学の基礎を築きました。ガレノスは、人間以外の動物を使用して、解剖なども行ない、それを基礎として、ヒトの医学理論と体系を築きました。

人類初の保健データのモデルは、1542年、ロンドンで、Parish Clerks が集めた黒死病(ペスト)の軌跡をまとめたものでありました。約 60 の疾患がカテゴリーにまとめられ、スプレッドシートを使った、「死亡表」でありました。この表は不定期に発行されました。

後に、世界最初の疫学者となる、ロンドンの商人であった、ジョン・グラント(1662)は、商業算術を駆使し、この死亡表を研究し、出生、婚姻、死亡など、人口動態に現れる規則性を発見し、人口統計をまとめました。近代疫学と人口統計データの成果物と言えます。ですが、その後、200年間、この偉業はとりあげられることはありませんでした。

次に、分類に関して、少々、述べますが、カール・フォン・リンネ(1763-1778)の名があがります。彼は、自然体系において、生物分類を体系化しました。生物の学名を、属名と種小名の 2 語のラテン語で表す二名法を体系づけました。分類の基本単位である種の他に、綱、目、属という、上位の分類を設け、それらを階層的に位置付けました。分類学の父と呼ばれています。

(図 4)

Major Classification Efforts

- Carolus Linnaeus
Carl von Linné
• *Genera Morborum* (1763)
- Underscored Content Difficulty
• Pathophysiology vs Manifestation
e.g. Rabies as psychiatric disease



「学術用語というものは、問診と同じくらい重要なものである。物理科学における、質量と度量と同じである。また、決して、遅れることなく、定められねばならない」 (*William Farr, about Cullenian system: First Annual Report of the Register-General of Births, Deaths, and Marriages in England, London. 1839 p99*)

1853年、ブリュッセルにおいて、世界初の統計委員会が発足された。William Farr (London), Marc d'Espine (Geneva), Achille Gullarad (Paris) …sponsor、とうメンバーでありました。139の注釈からなる、世界で最初の死因分類が、世にでました。

Il y a lieu de former une nomenclature uniforme des causes de décès applicable a tous les pays.
(Sixteenth Annual Report of Registrar-General of England, 1853, Appendix, p. 7)

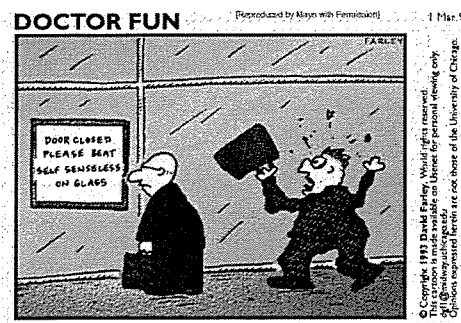
(1) ICD (International Classification of Diseases)

ICD-1: Jacques Bertillon

1893年、シカゴにおいて、1886年版の国際統計研究(161の注釈からなる)が、紹介されました。1898年、米国公衆衛生学会に適合され、10年毎に調査を行うことが発表されました。Bertillonはフランス政府の国際委員会のメンバーで、この研究委員会の事務局長でありました。その後、1900年、1909年、1920年と改訂されていきました。このうち、1909年に改訂されたものが、ICD-2です。ICDは疾病のカタログではありません。国際統計に基づいた、死因分類であり、見せかけだけのものではありません。適切な疾患名や、科学的な分類まで、含まれているものです。1950から

1954年にかけて、米国公衆衛生局の協力のもと、患者のインデックスに適合した。1954から1959年まで、ICD-7は5回の改訂を行ない、1959、1962年に発行されました。このとき、病院記録のインデックスにも適合されました。

(図 5)



1977年、270万ドルの費用をかけて、ICD-9-CM (Clinical Modification) が作成されました。

(図 6)

ICD-9-CM Resolution?

185 Malignant neoplasm of prostate

Contrast:

- Carcinoma in Situ vs.
- Widely Metastatic Cancer

(図 7)

Combination Platter Coding in Diabetes Mellitus

250.0	DM without mention complications
250.1	Diabetes with ketoacidosis
250.2	Diabetes with hyperosmolarity
250.3	Diabetes with other coma
250.4	Diabetes with renal manifestations
250.5	Diabetes with ophthalmic manifestations
250.5	Diabetes with neurological manifestations
250.7	Diabetes with peripheral circulatory disorders
250.8	Diabetes with other specified manifestation
250.9	Diabetes with unspecified complications

Severity?

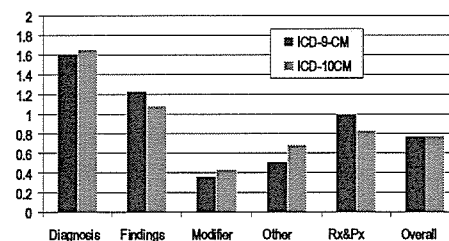
重篤度の変わりとしての、機能がなく、代わりとて、得られるもののみ、しばしば、得られたもののみ、広く、保健サービス研究に使われている。疾患が示唆する重篤度を除いたものを理想的に置き換えた。

臨床的な分類作業というのは、どのようにした、いけばよいのでしょうか？ CPRI (Computerized Patient Record Institute)は、内容を追いかけてながら、コード化することによる影響を調査した。

コード化と語彙は、国際的なワーキンググループの中で、ボランティアが努力している。

(図 8) ICD-9-CM と ICD-10-CM の比較

ICD-9-CM vs ICD-10CM (final draft) Composite Scores [hot off the press!]



(図 9)

Temporal Drift Re-use of Codes ICD-9-CM Procedure Codes

External Fixation example

1990	93.59	Other immobilization...
1991	78.4	Other repair or plastic op. bone
1992	78.1	Application of external fixation device
Note:	78.1	retitled from "periosteal suture"
=>	78.2	"limb shortening procedures"

which used to be in 78.3, now just "lengthening"

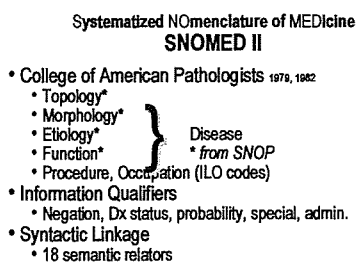
(図 10)

Changes made to DM Coding in 1993 Alone

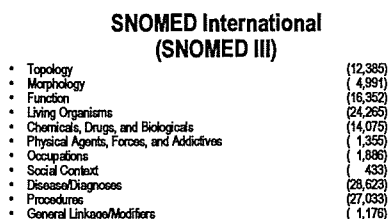
250.02 -> 250.90	250.52 -> 250.50
250.03 -> 250.91	250.53 -> 250.51
250.12 -> 250.10	250.62 -> 250.60
250.13 -> 250.11	250.63 -> 250.61
250.22 -> 250.20	250.72 -> 250.70
250.23 -> 250.21	250.73 -> 250.71
250.32 -> 250.30	250.82 -> 250.80
250.33 -> 250.31	250.83 -> 250.81
250.42 -> 250.40	250.92 -> 250.90
250.43 -> 250.41	250.93 -> 250.91

CPRI が、1996年に行なった、コード化の調査によると、①殆どのコーディングシステムは情報を欠いている、②いくつかのものは、他のものより、無くなっているものがある、③ICD、CPT (Current Procedural Terminology)でさえも、半数以上は臨床上、詳細とされるものが、失われている、④誤った分類へのバイアスは管理されたコーディングシステムを使用する研究においては問題があると結論づけている。新たな、Medical Concept の動きとしては、16世紀のスプレッドシートを拡張し、統制語と複合語を求めるものがあります。Axes(分類軸)のような形にしました。専門用語は、文字列よりはむしろ、概念を表現す

(図 15)

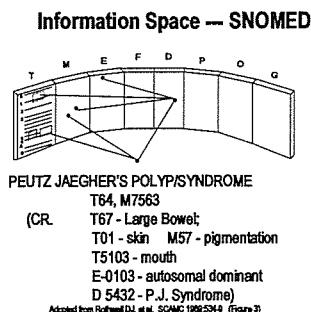


(図 16)



SNOMED III に階層構造を加え、語義を結合させ、関連性を一定の形にし、Kaiser Permanente CMT (Cretified Medical Transcription) のコンテンツと関連性を加えたものが、SNOMED - RT (Reference Terminology)です。

(図 17)



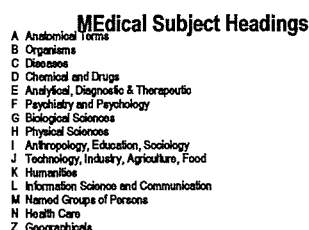
1999 年 4 月、英国の NHS(National Health Service) が発行している、Read Code をマージすることに合意、2001 年、SNOMED CT(Clinical Terms) になり、現在に至ります。

(3) MeSH (Medical Subject Headings)

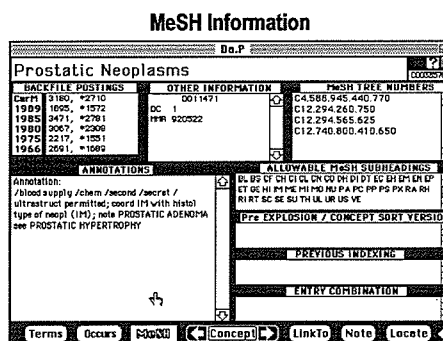
Index Medicus のルーツを振り返ると、1964 年、医学を基礎とする、Medlars のために、電算

化されました。Biomedicine に焦点をあてた、広くカバーするものとして、作成された、医学用語集(シソーラス)が MeSH です。文献検索から発展したもので、2003 年現在、主題語(見出し語、Descriptor) が 21,973 語、同義語(Synonym)が 125,858 語、薬物、化学物質の語が、132,123 語(これは、Supplementary Concept Record という別テーブルになっています)、収録されています。図書館員に焦点をあてて、作成され、基礎科学をベースにしていたため、臨床学的には、貧弱なものでありました。MeSH には A-Z まで、15 のカテゴリーから成り立っていました。

(図 18)



(図 19)



(図 26)

A little closer look:

[from K. Spackman]

- 285.9 Anemia, unspecified
 - Anemia:
 - NOS; essential
 - normocytic, not due to blood loss
 - profound; progressive; secondary
 - Oligocythemia
 - Excludes: anemia (due to):
 - blood loss:
 - acute (285.1)
 - chronic or unspecified (280.0)
 - iron deficiency (280.0-280.8)

用語のマッピングは本当に、容易なのでしょう
か？貧血を例にとると、定義が多少なりとも異な
っています。

(図 27)

The easy part of mapping???

[from K. Spackman]

- SNOMED
 - DC-10010 Anemia
- ICD-9-CM
 - 285.9 Anemia, NOS
- But DC-10010 subsumes chronic blood loss
anemia, and 285.9 explicitly excludes it!

(図 28)

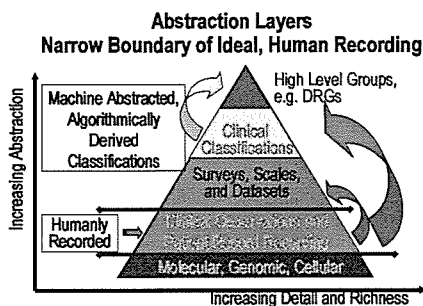
A little closer look:

[from K. Spackman]

- 285.9 Anemia, unspecified
 - Anemia:
 - NOS; essential
 - normocytic, not due to blood loss
 - profound; progressive; secondary
 - Oligocythemia
 - Excludes: anemia (due to):
 - blood loss:
 - acute (285.1)
 - chronic or unspecified (280.0)
 - iron deficiency (280.0-280.8)

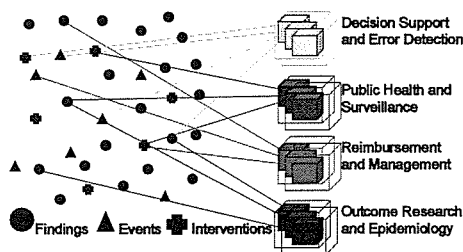
マッピング機能の実用性となると、分類には、
正式なルールが要求されます。ですが、まだない
ものですよね。記録とした、臨床学的要素の存在
を知らねばなりません。臨床学的な事象を再度集
め、用語間のマッピングは成功するでしょうか？

(図 29)



(図 30)

Aggregation Logics by domain
rule-based aggregations



2. 2. 3 挑戦

完璧な用語とは、臨床学的な完成度には、十分
とは言えない、患者の所見、徴候、診断を取られ
ても、それも、適切だとは言えない。

2. 2. 4 結論

今日のヘルスケアにおいて、用語というものは、
患者データを二次的に利用をすべての場面で問
題が潜んでいる。共通システムに向かって、進む
ことは危険であり、起こりやすいものになってい
る。臨床システムを統合するよう、要求されてい
る。

臨床ベースの文献検索をどのようにしたら、漏
れなく、検索できるであろうか？患者に出会っ
てから、臨床ガイドラインができるまで、この間を
いかにスムーズにつなげるかに全てが、かかって
いる。米国では、SNOMED CT の供給率はかな
り、高く、部位+症状、Axes などから、構成でき
る用語集が求められているのであろうか？それ
でも、PubMed の利用率は高く、容易に一次情報
まで、たどりつけるが、診療体系、診察から、診
療費用の請求までとなると、まったく役に立たな
くなってしまふ。日本でも、ICD はレセプト計算
などに利用できるが、MeSH ではどうであらう
か？もちろん、使用できない。その国の医療体制
まで、包含する用語集の確立が急がれると思われ
る。

2. 3 生物学における統制用語集——意義と課題及びその限界

Controlled Terminologies in Biomedicine: Rationale, Challenges, and Limitations¹

William Hersh, MD
Professor and Chair
Department of Medical Informatics and Clinical
Epidemiology
Oregon Health and Science University

2. 3. 1 はじめに

医学領域において、統制された用語集が、なぜ重要なのか？ 個々の専門用語集とその限界にどこまで挑み、どこに進むのか？ をこれから、いくつかの例を紹介しながら、述べたいと思います。

医療情報学というものは、あくまでも情報であります。技術のことを言っているではありません。情報と、人と、そして、テクノロジーを結びつけるものが、医療情報学です。一面から見れば、利用者の広い範疇の中にあると言えます。例えば、患者ケアの中で生まれ、使われているものは、患者に限定したものとなり、医療に従事するものの科学文献と言え、ナレッジベースのものということになります。このように、利用するものによって、それぞれ特有のものが生まれます。

2. 3. 2 統制用語とその実例

NCVHS (National Committee on Vitals and Health Statistics) によれば、① いかなるデータソースであっても、他のデータソースとの交換が可能で、相互運用ができなければならない、② データに一貫性があり、データを比較することができなければならない、③ 継続して、収集され、信頼性のある、質の保てるデータでなければならない、などの条件を満たす必要があります。Voitovich は、「医学は、少ない言葉で、多くの意味を持たねばならない、それは、航空管制や軍隊のように」と述べています。(1999年、Reduction of medical verbiage: fewer words, more meaning. *Annals of Internal Medicine*, 131: 146-147)

また、Chute は、①所見、症状、結果を文書化

し、情報を捉える、②情報を転送し、コミュニケーションを図る、③疾患、治療法などの分類は、知識を組織化する、④ナレッジベースの情報にアクセスし、情報検索をする、⑤意思決定支援を補う働きをする、これらはデータの標準化と用語の標準化に必要な要素であります。

コンピュータには、知性がありません。コンピュータは、自ら何かできるというものではありません。プログラマーや人間が、指示したことしか、実行できません。確かに、コンピュータは数字と記号の処理は早く、正確であり、膨大な記憶装置をもっています。それらは、コンピュータの中で描写された、情報のもつ意味をハッキリさせてくれます。

Common Cold(風邪、感冒)という言葉には、どれくらい違った言い方ができるのでしょうか？ 同義語の中には、

- ① Cold 風邪、感冒
- ② Upper respiratory infection 上気道感染
- ③ URI (同上)
- ④ Pharyngitis, bronchitis, rhinitis, etc 咽頭炎、咳、鼻炎など
- ⑤ Viral syndrome ウイルス性症候群
- ⑥ ...

と言うように、いくつも同義となる語が存在します。

次に、lead では、どうでしょうか？

- ① Hypertension leads to heart disease 心臓疾患を起こす高血圧
- ② An EKG lead 心電図
- ③ Lead poisoning 鉛中毒

¹本論文は、この研究班が招聘した Hersh 教授が 2006 年 3 月 2 日に東京大学鉄門記念講堂で行った講演の記録をもとにしたものである。

④ …

というように、まったく異なる意味になる場合もあります。

(1) Terminology Code

つまるところ、コードを用いて、コンセプト(概念)を表現することが望まれる。①順序だてた数字、あるいは、ランダムな数字、②略語、③上下関係が示せる階層構造、④並列にならべて、コード化する、⑤順序だてて、並べコードかする、等、様々なアプローチがある。ただし、意味(語義)をコード化する、“Semantic Code”は避けるべきである。

では、標準化された医学用語というものは、どのようなものであろうか?①多くの専門家グループや委員会によって、それぞれの目的にあわせて、コード体系が開発されているもの。②特別な語は特有の問題を解決するために、設計されている。他の用語へのマッピングや、他の目的のために利用することは、しばしば問題がおきるものである。③階層構造をもっているものとなる。

(2) 標準的な用語集

- ①診断の分野では、ICD-9、ICD-9-CM、ICD-10、DRGs (Drug-related groups)
 - ②臨床徴候と所見では、SNOMED-CT
 - ③医薬品(薬物)では、National Drug Code (NDC)、National Drug File Reference Terminology (NDF-RT)
 - ④研究の分野では、LOINC (Logical Observation Identifiers Names and Drugs)
 - ⑤(手術)手技と診断学的研究の分野では、CPT-4 (Current Procedural Terminology)、HCPCS (Healthcare Common Procedure Coding System)
 - ⑥医療機器の分野では、Universal Medical Device Nomenclature (UMDN)
- と言ったものがあります。

(3) ICD (International Classification of Disease)

ICDは1893年、死因別につくられた国際統計リストに起源を發します。当初の目的は死亡統計を集めることでありました。後に、WHO (World Health Organization)にその作業は引き継がれ、現在は、国際疾病分類と呼ばれており、死因というより、疾病のコードとう意味をもつようになりました。

1975年、ICDも改訂を重ね、ICD-9となりました。階層構造を数字でもつようになり、数字の桁がそれぞれの階層レベルを表します。因みにICD-9は4桁の数字を持っています。後に、米国用に修正され、5桁のコードをもつ、ICD-9-CM (Clinical modifications)ができました。返済をする多くの団体は、ヘルスケアに携わるプロバイダーはICD-9-CMのコードを割り当てることを求めました。また、予防とスクリーニングに関わってしまうため、“V code”が追加されました。

(図1) ICD-9-CMの例

Example of ICD-9-CM

```
003 Other Salmonella Infections
  003.0 Salmonella Gastroenteritis
  003.1 Salmonella Septicemia
  003.2 Localized Salmonella Infections
    003.20 Localized Salmonella Infection, Unspecified
    003.21 Salmonella Meningitis
    003.22 Salmonella Pneumonia
    003.23 Salmonella Arthritis
    003.24 Salmonella Osteomyelitis
    ...
    003.29 Other Localized Salmonella Infection
  003.8 Other specified salmonella infections
  003.9 Salmonella infection, unspecified
```

15

OHSU

1990年、WHOはICD-10を米国以外の多くの国によって、作成しました。データフォーマット(数字の桁数)を除いた構造を大きく変化させました。米国のヘルスケアシステムに要する費用はかなりの額(5~10億ドル)になると思われませんが、質の高いデータを蓄えることで、取り戻せるであろうと、Libickiは述べています。

Chuteによれば、ICDの限界は、①用語が変わった時、特定のコードの意味を変えることができない、例)新型の肝炎など、②数字の桁数で表すコードを使用することには、問題がある、例)10を超える項目に合わせようとする時、③詳細になると、不適切なものになることがしばしばある、例)ひとつのコードで殆どの癌を表現している、④拡張性がない、部位や重篤度などの表現を加えることができず、関連性を示すこともできない、述べています。

(4) Diagnosis-related groups (DRG)

DRGの意図は、医療サービスの研究のため、ICD-9のコードを集め、グループ化することにあります。数百のコードセットは病気を表し、1980年代に、HCFA (Health Care Financing

Administration) によって、入院治療の支払い請求に適応されました。

(図 2) DRG の例:呼吸器系疾患

DRG examples for respiratory diseases

Respiratory disease w/ major chest operating room procedure, no major complication or comorbidity 75
 Respiratory disease w/ major chest operating room procedure, minor complication or comorbidity 76
 Respiratory disease w/ other respiratory system operating procedure, no complication or comorbidity 77
 Respiratory infection w/ minor complication, age greater than 17 79
 Respiratory infection w/ no minor complication, age greater than 17 80
 Simple Pneumonia w/ minor complication, age greater than 17 89
 Simple Pneumonia w/ no minor complication, age greater than 17 90
 Respiratory disease w/ ventilator support 475
 Respiratory disease w/ major chest operating room procedure and major complication or comorbidity 538
 Respiratory disease, other respiratory system operating procedure and major complication 539

19

OHSU

(5) SNOMED Clinical Terms (SNOMED CT)

<http://www.snomed.org/snomedct/>

2005 年、米国病理学会 (CAP: College of American Pathologist) によって、開発されました。SNOP (Systematized Nomenclature of Pathology) の分類を基にしましたが、1980 年代に医学全般に適応するよう、拡張されました。また、SNOMED CT には、英国に起源をもつ、Read Code がマージされています。SNOMED CT は臨床診断や所見を述べる際に、最も豊富な語彙をもっており、多軸型、あるいは、構成面からのアプローチが特徴です。①用語と用語を結合させることが許されている、例) Lung(肺) + inflammation(炎症)、②修飾語を加えることが許されている、例) 重篤度、悪化、というような組み合わせと症状、徴候を表現することができる。2005 年 7 月現在、① Concept : 366,170 語、② Description : 993,420 語、③約 146 万の関連語が存在している。

(図 3)

SNOMED CT concept hierarchy

Concept Hierarchy	Example
Finding	Swelling of arm
Disease	Pneumonia
Procedure	Biopsy of lung
Observable entity	Tumor stage
Body structure	Structure of thyroid
Organism	DNA virus
Substance	Gastric acid
Pharmaceutical/biologic product	Tamoxifen
Specimen	Urine specimen
Physical object	Suture needle
Physical force	Friction
Events	Flash flood
Environments/geographical	Intensive care unit
Social context	Organ donor

22

OHSU

(図 4)

Some attributes (relationships) for a Disorder and Finding

- Finding Site
- Associated With
 - After
 - Causative Agent
 - Due To
- Associated Morphology
- Severity
- Onset
- Course
- Episodicity
- Pathological Process
- Associated Etiologic
- Finding
- Temporally Follows
- Has Definitional Manifestation
- Occurrence
- Stage
- Subject of information

23

OHSU

2003 年 7 月、CAP は米国国立医学図書館 (NLM: National Library of Medicine) は全米で利用可能な、5 年間分の SNOMED CT ライセンスを得ることに合意。米国内にある、すべての公共、民間の医療機関、研究機関、教育機関、あるいは統計的利用においても、無料で利用可能になった。患者レベルのデータセット、あるいはナレッジベースでも、円コードが可能になり、利用者は多くを取り出さなくとも、再配分が可能になり、UMLS メタシソーラス経由で再配布されています。

(6) National Drug Codes (NDC)

<http://www.fda.gov/cder/ndc>

11 桁のコードからなり、各々、すべての医薬品製剤に付与されている。①最初の 5 桁が製薬メーカーのコードであり、FDA (Foods and Drugs Administration) により、割り当てられている。②次の 4 桁のコードは製品名、用量などを示している。すべての薬剤にコードが付与されているが、製薬メーカーが割り当てているので、同じ薬でありながら、違うコードを持ってしまうという問題

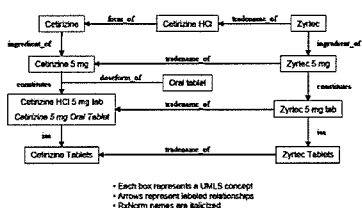
がある。③最後の2桁のコードはパッケージを表しており、一瓶に何錠入っているかなどが、記されている。

(7) National Drug File Reference Terminology (NDF-RT)

新薬の用語をベテランの検査官、NLM、FDAによって、作成されている。Nelsonによると、NLMのRxNormプロジェクトは処方と構成物質のsemantic構造を提供している。

(図 5)

Relationships in RxNorm (Bodenreider, 2004)



27

OHSU

(8) Logical observations, identifiers, and numerical codes (LOINC)

<http://www.regenstrief.org/loinc/>

各試験において、①特性、例)質量濃度、割合、②時、③標本、例)血液、脳脊髄液、④方法、例)定性、定量、などを明記する。

(図 6)

LOINC Examples

- Blood glucose GLUCOSE:MCNC:PT:BLD:QN:
- Serum glucose GLUCOSE:MCNC:PT:SER:QN:
- Urine glucose concentration GLUCOSE:MCNC:PT:UR:QN:
- Urine glucose by dip stick GLUCOSE:MCNC:PT:UR:SQ:TEST STRIP
- Ionized whole blood calcium CALCIUM:FREE:SCNC:PT:BLD:QN:
- 24 hour calcium excretion CALCIUM:TOTAL:MRAT:24H:UR:QN:
- Automated hematocrit HEMATOCRIT:NFR:PT:BLD:QN:AUTOMATED COUNT
- Manual spun hematocrit HEMATOCRIT:NFR:PT:BLD:QN:SPUN
- Erythrocyte MCV ERYTHROCYTE MEAN CORPUSCULAR VOLUME:ENTVOL:PT:RBC:QN:AUTOMATED COUNT
- ESR by Westergren method ERYTHROCYTE SEDIMENTATION RATE:VEL:PT:BLD:QN:WESTERGREN

28

OHSU

(9) 他のヘルスケアの用語集

- ①MeSH (Medical Subject Headings)
- ②CPT-4 (Current Procedural Terminology)
- ③Various nursing vocaburaries

- ④ACR code (American College of Radiology)
- ⑤ Diagnostic and Statistical Manual of Mental Disorders (DSM)
- ⑥ Universal Medical Device Nomenclature (UMD)

2. 3. 3 統制された用語集の限界

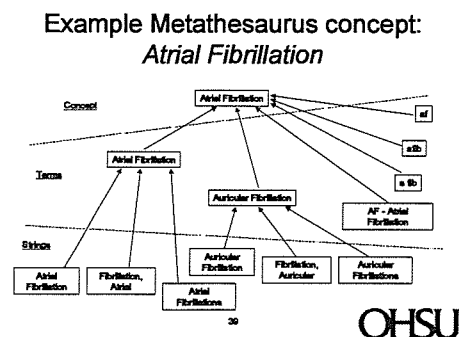
臨床医師に利用される用語を表現しません。臨床所見をあらゆる医学用語の殆どは、変更されず。これが垣根のような存在になります。また、単語を組み合わせて、複合語をつくることを認めて欲しいと思っています。医療請求が優先されるため、どうしても、歪が生じます。例)E/M Code (Evaluation and Management Code), DRGs

(1) UMLS (Unified Medical Language System)

<http://www.nlm.gov/research/umls/>

1980年代、NLMは多くのMedical Systemを調査した。どれも、異なる用語集を持っていた。MeSH、ICD-9、それぞれのアプリケーション特有の用語集、等々。そこで、利用者が適切な資源を、それぞれの用語間でつながる手段が必要であることを見出した。1986年、UMLSプロジェクトをNLMが引き受けることとなった。UMLSは、①メタシソーラス、②Semantic Network、③Specialist hexicon、の3つから、構成されている。メタシソーラスは、医学分野で用いられる、異なる統制語と異なる分類をひとつ、あるいはそれ以上、出現する、コンセプトを情報データベースとしたものである。メタシソーラスは、主な用語間を横断するものです。表記が異なる同義語であっても、同一のコンセプトであれば、同じコードが付与されます。ハッキリと区別の可能な用語であっても、表記の異なるものがあります。すなわち、文字列の違いです。UMLSは、①1対1の関連のみ、案内されている。②元となる用語が存在するもののみ、登録されている。③一元管理できない、階層が存在する。それも元となる用語が存在するもののみである。あと、④拡張性がない。これらはUMLSの限界と言える。だが、現時点では、まだ少ないが、用語の倉庫以上のものである。メタシソーラスのゴールは用語のリンクを作ることでありませぬ。新しい用語をつくるものでは、ありません。望まれているものもあれば、代弁するものもあります。メタシソーラスは、医学用語の標準となるべきであり、それも、実現可能なものであります。

(図 7)



(2) メタシソーラスのコンセプトと関連した他のデータ

・ 1)概念属性、①カテゴリー、②Symantic type、③定義…ドランド等から、2)用語属性、①状態…優先語と同義語、②Lexical Tag…イニシャル、略語、③言語、3)文字列属性、文字列によるもの、例)語順、複数形の変化など、4)Atom 属性…元となる用語、5)関連性、①階層の状況は、元となる用語にある、②関連したコンセプトと関連性の型、③変化するソースのコンセプトを引き起こす、例)MEDLINE

メタシソーラスの限界

先ほども述べたが、ひとつの統一された用語より、用語集の倉庫以上になる。この異質さは、臨床のアプリケーションがメタシソーラス全体を使いそうもない。特異なアプリケーションは、個々の用語集には忠実である。例)臨床用途なら、SNOMED CT、文献インデキシングと知識のための、MeSH。異なる利用方法および現存する用語システムは単一の標準になりそうもない。

2. 3. 4 統制用語の限界

コーディングされたデータはデータの質で計って、あつめることができ、臨床の場では、意思決定支援になるというメリットがあるが、医師の時給は高く、コード化には抵抗がある。下位レベルのコードは解決され得るであろうが、時間を要し、臨床のエキスパートがいなければならぬ。アプリケーションは費用便益分析によって、操作されなければならない、ここに限界がある。

2. 3. 5 統制用語集の未来

SNOMED は臨床データのアプリケーションとして、巨大になり、LOINC は臨床試験においては、より詳細な価値が出るであろう。医薬品の標準化

は継続する必要がある、MeSH は良いものであるが、ナレッジベースのアプリケーションとしては、パーフェクトではない。ゲノミックスのアプリケーションは議論の必要がある。

2. 3. 6 まとめ

以上が、Dr. Hersh の講演内容である。臨床ベース、研究ベース、医薬品ベース、それぞれが、異なるシステムで、異なるシソーラスのような用語集を開発しているが、それぞれの場では有用であるが、結びつけるものがないことが、どうしてもネックになる。NLM の副館長である、Ms. Humphreys は、和解を求めていると述べたが、今後、どのように進んでいくのか。これは西洋医学の分野に限ったものであるが、アジアなどでは、用語や医学文献検索システムが共有され、それぞれの分野のエキスパートが医療の質を向上させるような、データベースシステムが開発されるのは、まだまだ、先の話と言える。本研究が、このような場で、どのように展開するかを、検討する必要はある。

第3章 日本における 医学専門用語

3. 1 医学中央雑誌のあゆみ

3. 1. 1 はじめに

医学中央雑誌刊行会は1903年(明治36年)、医学文献抄録誌として、市井の開業医であった尼子四郎によって、創刊された。その目的は、創刊号の謹告で述べられているように、我が国の医学の発展を願ったものであった。以来、100余年、幾多の困難を乗り越えながら、発行され続け、形態は冊子体、CD-ROM、Webと時代の変化とともに、変わったが、医師、医学研究者、歯科医師、薬学関係者(研究者、薬剤師)ばかりでなく、看護師、臨床検査技師、リハビリテーション分野の理学療法士、職業訓練士、栄養士など、医療従事者ばかりでなく、介護関係、報道分野、法曹界、患者およびその家族にまで、幅広く利用されている。

3. 1. 2 医学中央雑誌(医中誌)の歴史

1903年、尼子四郎の手によって、創刊され、その当時は、本郷区春木町(現在の文京区本郷)の南江堂内に、医学中央雑誌社として、事務所を構えていた。その後、本郷にあった尼子四郎の自宅に移った。創刊当時は、月1回の発行で、収録文献数は、1,886、掲載誌は64誌であり、B5版の縦組みであった。1928年(昭和3年)、尼子四郎が逝去する2年前に、子息である、尼子富士郎に代表をゆずり、その後戦災によって事務所を失い、現在の杉並区高井戸にある尼子富士郎の自宅に事務所を移した。1959年(昭和34年)、株式会社に組織形態が変更されたが、1964年(昭和39年)に任意団体に組織形態を変更、直接購読会員制を採用した。この間、1963年10月に、尼子富士郎は医中誌発行の功績により、紫綬褒賞を受賞。1967年(昭和42年)、4月には、日本科学技術情報センターより「日本科学技術情報センター賞」、同年11月、医学文献功労者として、日本医師会より最高功労賞を受賞した。1972年3月、尼子富士郎逝去。尼子宏子夫人が理事長代行を務め、1973年、村上元孝が理事長に就任した。

この頃、毎号末尾に最新文献情報、腫瘍文献が掲載されており、歯学版、外科学・整形外科・放射線学版といった分冊が発行されていた。1981年(昭和56年)、科学技術振興事業団(JST)に国内医学文献データベース作成への協力を開始。1983年(昭和58年)、4月、医学用語シソーラスの第1

版が発行され、電算処理編集による医中誌の発行が開始された。また、この年より、年間累積索引誌の発行も開始された。1986年(昭和61年)、JSTへ医中誌基本データベースの提供を開始。翌、1987年、医学用語シソーラス第2版を発行。年間累積索引誌にIndex Medicus方式を採用した。

1989年、村上元孝理事長が逝去し、篠原恒樹が理事長に就任した。1992年(平成4年)、医中誌CD-ROM版のサービスを開始。紀伊國屋書店、丸善、ユサコの3社が代理店となり、大学、大学附属病院、病院の図書館、製薬メーカーなどに展開された。同年9月には、第1回の医中誌ユーザー会を開催。医中誌利用者の声を生で聞ける機会とするばかりでなく、普段、質問等で問い合わせのある内容などを吟味し、より便利な使い方等を紹介することなどを行った。1992年、医中誌OA化検討委員会を設置し、コンピュータ処理による、編集作業の検討が始まった。1994年、医学用語シソーラス第3版を発行。1995年6月、社内LANによるOA化が完成、コンピュータによる編集作業が始まった。同年、分冊であった歯学版が終刊。また、科目別分類による編集を廃止、同時に抄録誌からIndex Medicus方式による、キーワード配列の索引誌に変更。翌1997年(平成10年)、第2回、医中誌ユーザー会を開催。医中誌Web開発の先駆けとなる医中誌ネットワーク準備委員会を設置。1999年(平成11年)、医学用語シソーラス第4版を発行。また、医中誌のホームページが立ち上がった。この年の9月には、第3回、医中誌ユーザー会を開催。医中誌Webの紹介が東京、大阪の2会場で行われた。

この数年前より、医学基本データベースを提供していたJSTより、思わぬ圧力がかかることになる。JSTはWebにてのサービス提供を既に始めており、医中誌が独自にWebサービスを始めることを快く思わなかった。この争いは現在でも続いている。2000年には、JSTへのデータ提供を中止。その2年後の2002年、JSTでの医中誌基本データベースの使用が終了となる。2000年4月より、待ちに待った、医中誌Webのサービスが開始となる。代理店にサンメディアが加わり、4社での展開となる。同年11月、日本電子計算より、医中誌@infostreamとして、サービスが開始され、

2001年(平成13年)、3月、第4回医中誌ユーザー会を開催。内容は医中誌 Web 全般に渡り、今後、予定されている開発の内容などが広く紹介されるものとなった。同年、英文タイトルの入力を開始、EBM 準備・検討委員会が設置される。2002年7月、公共性の高い業務の性質上、公益法人化が望ましいと判断し、特定非営利活動法人(NPO)に組織形態を変えた。

これと時を同じくして、時代の流れは医中誌 Web への移行を促進し、長年続いた冊子体は終刊となった。翌年には、年間累積索引著者名編も終刊した。2004年には、冊子体2003年、医学用語シソーラス第5版が発行され、同年、3月に「医中誌創立100周年」を迎えた。その年より、EBMへの要求に応えるべく、4つの「研究デザイン」タグの付与を開始。1999年作成データまで遡り、「メタアナリシス:Meta-Analysis(MA)」、「ランダム化比較試験:Randomized Controlled Trials(RCT)」、「比較臨床試験:Controlled Clinical Trials(CCT)」、「比較研究:Comparative Study(CS)」のタグが付与された。その後、「診療ガイドライン」のタグも付与されるようになり、凡その作成状況が把握できるようになった。

2005年、12月をもって、冊子体での唯一の提供媒体であった、年間累積索引件名編も終刊となり、冊子体での提供はその年をもって終了した。2006年2月、医中誌 Web も Version 4 になり、一次情報へのリンクができるようになった。同年3月に、CD-ROM 版のサービスも終了。この時をもって、Web サービスのみになり、インターネットの普及を改めて実感した。今後はますます、Web サービスの重要性が増し、Web を中心としたサービス形態の利点を生かし、利用者のニーズに応えていくことが、医療情報を提供し、医療の質の向上を目指す当会の、100年前に尼子四郎の思いを実現することであると、新たな医中誌の使命を自覚している。

医中誌 Web、Version 4 のサービスは、当初、法人契約のみのサービスであったが、今年、2007年2月より、パーソナル Web にも適用され、法人、個人という分け隔てなく、利用者への期待に応えるべき、更なる機能の向上、サービスの展開を目指している。

3. 1. 3 医中誌の掲載資料数及び収録文献数

1) 創刊から1945年(昭和45年)

① 創刊当時、掲載誌数、64誌、収録文献数、1,886件

② 1943年(昭和18年)、掲載誌数、465誌、収録文献数、21,452件
40年で、約11倍に増加

2) 1946年(昭和21年)から2002年

① 戦後、1947年(昭和22年)、掲載誌数、350誌、収録文献数、9,740件
350

② 2002年(平成14年)、掲載誌数、2459誌、収録文献数、282,785件

57年間で、約28倍に増加

* 1944年(昭和19年)、戦災のため、休刊

3. 1. 4 発行形態の変化

1) 創刊から1936年(昭和11年)まで、B5版、縦組、科目別分類

2) 1936年(昭和11年)、4月から1983年(昭和58年)3月まで、B5版、横組、科目別分類、各巻索引

3) 1983年(昭和58年)、4月から1995年(平成7年)、A4版に変更、12月まで、横組、各号索引、年間累積索引

4) 1992年(平成4年)、1月より、CD-ROM版、サービス開始

5) 1996年(平成8年)、1号から2002年(平成14年)、12号まで、A4版、科目別分類の抄録誌から、Index Medicus 方式へ変更。キーワード配列による索引誌に。抄録はCD-ROM版、医中誌 Web に掲載

6) 2000年(平成12年)、4月より、医中誌 Web サービス開始

7) 2002年(平成14年)、12号をもって、冊子体での提供は終了。但し、年間累積索引主題編は年間累積版として、引き続き発行。著者名編は2001年(平成13年)に終了

8) 2006年(平成18年)、3月、CD-ROM版のサービス終了

収録内容

1) 戦前・・・一般文献(科目別)以外に綜説(一般文献)、官報、新刊図書、最新文献、海外新着誌、法律、学位記事等を適宜掲載。

2) 戦後・・・殆ど一般文献(一次最新文献、新刊図書掲載)が主体

3) 最終分類科目…33科目

生理学、生化学、薬理学、薬学、解剖学、病理学、実験腫瘍学、衛生学・公衆衛生学、疫学、微生物学、血清学・免疫学、法医学、精神医学、内科学、小児科学、放射線学、外科学、整形外科学、形成外科学、眼科学、耳鼻咽喉科学、産婦人科学、皮膚科学、泌尿器科学、麻酔学、臨床検査、リハビリテーション、歯学、看護学、東洋医学、獣医学、医史学、社会医学

4) 戦前医中誌掲載一次文献綜説例

- ① 腎臓疾患の病態生理 稲田龍吉
(医中誌 38 巻、669-691 頁、1933 年、昭和 8 年)
- ② 生理学の基礎的概念の考察(1)
刺激、興奮、被刺激性及び興奮性
橋田邦彦 (医中誌 36 巻、735-757 頁、1933 年、昭和 8 年)

3. 1. 5 冊子体における文献検索

- 1) 創刊から 1995 年(平成 7 年)末まで
科目別分類表示文献をブラウジングする検索法と各号あるいは各巻掲載の件名および人名索引をキーとした検索法
- 2) 1996 年(平成 8 年)から 2002 年(平成 14 年)末まで
科目別編集による抄録誌を改め、キーワード配列による索引誌となり、キーワードのもとに各文献の書誌事項が掲載されている

3. 1. 6 医学用語シソーラス

検索および索引用のキーワード集である医学用語シソーラスは現在、第 6 版(2007 年 1 月発行)まで、作成されている。五十音順とカテゴリー別リストで構成されている

3. 1. 7 医中誌データベース作成の概要

医学中央雑誌データベース(医中誌DB)は、医学・歯学・薬学および看護学などの関連領域の定期刊行物を対象とした文献情報を蓄積したもので、データソースは、国内の大学・研究所・学協会・病院などから発行されている紀要や学会誌・会議録、医学系出版社が発行する商業誌の他、公共資料や講演集など、約 2400 誌。収録されている文献データは 1983 年から現在までで、5,954,704 件。データの内容は、いわ

ゆる書誌情報(文献のタイトル、著者名、著者の所属機関名、雑誌名と発行年月・巻号ページなど)と、キーワードおよび抄録である。

データ作成の大まかな流れは、1)資料収集と採択、2)資料の受入、3)文献の採択、4)書誌情報の入力、5)抄録の作成、6)インデクシング、となる。

(1) 資料収集と採択

「医学中央雑誌」は創刊以来、網羅的な資料収集を行っている。資料の半数ほどが、学会・大学・学校・病院・出版社からの寄贈で、半数は購入している。未収録の資料や新しく創刊された資料については、新刊案内・医学雑誌目録・学会案内リストなどをチェックし、取り寄せます。そして、採否を検討します。最近では、「医中誌Web」利用者から資料の自薦・他薦情報がメールで寄せられることもある。収録候補資料の採否検討は、複数の採択担当者によって行われている。この検討会は、以前は半年～1年に1度であったが、タイムラグ短縮への配慮により、2002年より1～2ヶ月に1回の割合で開かれている。

(2) 資料の受入

医学中央雑誌刊行会には、毎日 40～50 冊、年間約 1 万冊の資料(雑誌)が届きます。届いた資料に対して、まず書誌番号(雑誌番号)を付与し、データベースに登録する。書誌番号とは、届いた年月日と登録順から成る 10 桁の番号です。例えば、2004 年 6 月 2 日に届いた資料で、1 番目に登録される場合 2004060201 となる。

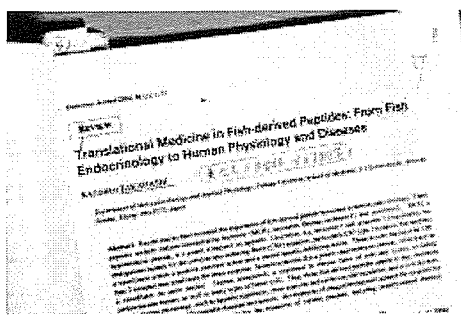
(3) 文献の採択

資料の登録後、掲載されている各文献について、データとして採択するかどうかを検討する。新しく発刊された雑誌に対しては、定期的に採択するか否かを検討している。手順としては、採択担当者が各文献をチェックし、医学に関連する文献はすべて、採択、化学物質の合成・分析研究、植物そのものの研究、野生動物の生態や病気の研究などは採択対象にはしていない。医学の周辺領域に関する文献の採択はかなり微妙なところもあり、採択担当者の経験に拠っている。また、インタビュー、対談、座談会、巻頭言、訪問記、書評、エッセイなどは採択し

ないというルールになっている。また、発表者が記載されていない文献も採択していない。採択された文献は、即座に、記事区分(論文種類)と抄録を付与するかどうかが決めている。原著論文に分類された文献は、原則として抄録を付与する指示が入り、抄録作成に取り掛かります。

(4) 書誌情報の入力

文献の書誌情報の入力は専門の入力業者に依頼しておりそのための、細かい指示が文献ごとに書き込まれる。まず、文献管理番号で、これは、10桁の雑誌番号に4桁の番号を加えたものである。たとえば、2004060201の雑誌の1番目に採択された文献は、2004060201-0001となり、これが文献管理番号すなわち文献のID番号となる。この他に、特集記事の指示や著者の所属機関等、どれを、どのように入力するのかを細かく指示がなされる。この作業は一見無駄のようにも見えるが、入力者が迷わずに済むように、きちんと細かく指示をする方が、結局は時間が短縮される。文献の書誌情報が一定のルールでデータ化されるように、この処理は必須となっている。ただし、1,000件以上の文献が掲載された会議録(学会抄録)などは、この作業に1週間以上要することもある。



(図1)

この前処理の指示に従って入力したデータ(書誌情報)がデータベースに流し込まれると、以降の作業は「JAMAS 2 System(医中誌データベース作成システム)」上で行われる。

入力された書誌情報のうち、著者名と所属機関名については、当会で収集した、それぞれの辞書とのマッチングを行い、自動的にヨミが振られる。著者名は複数のヨミがある場合でも、便宜上1つに決めてヨミを振っている。たとえば、「西原」という名字は「ニシハラ」「サイバラ」の両方の

ヨミがあるが、「ニシハラ」を採用しており、名前についても同様に1つの読みを採用している。マッチングで漏れた初出の著者名・所属機関名のヨミ振りは手作業で行うと同時に、各辞書に登録され、次回からは自動的にヨミ振りされる。このように、ある程度、統制された、ヨミ振りを行っている。

(5) 抄録の作成

抄録作成は、それを専門とする50名ほどの契約者に依頼し、第三者抄録として作成している。抄録者には文字数が、300字程度に収まるよう指示し、出来上がってくるまでに、平均2週間程度を要している。2004年から、この期間を短縮するため、著者抄録の利用を開始し、各学会、出版社に直接、交渉し、許諾を得るという作業を行ってきた。この許諾交渉は順調に進んでおり、昨年までで、約400の学会、約30の出版社から、著者抄録の使用許諾を頂き、今後も、増える予定である。昨年までの累計で、35,000件ほどになっている。

作成された抄録原稿は、電子ファイルの場合と手書きの場合がある。抄録原稿は数年前までは手書き原稿が主であったが、今ではファイルにしてメールで届けられることが多くなっている。手書き原稿はテキストファイルにする作業が入り、今でも達筆のあまり判読に苦勞する原稿がある。こうして作成された抄録テキストは、担当者により校正され、それが終了すると、データベースに追加される。

(6) インデキシング(索引作業)

書誌情報と抄録が入力された文献は、データ作成の最終段階であるインデキシングに回される。約30名(職員10名、契約者20名)のインデクサーで、毎月20,000~25,000件、年間約30万件もの文献にインデキシングしている。インデキシングとは、当刊行会作成の「医学用語シソーラス」に基づいてキーワードを付与し、さらに、副標目、チェックタグ、研究デザインタグを付与することを言っており、この作業によって、漏れの少ない検索やさまざまな観点からの絞り込み検索が可能になります。

インデキシングが終了すると、「医中誌Web」で提供されるデータとして準備が整ったことになる。これらは、改めて、その年度の通し番号(10

桁の「文献番号」)が付与され、毎月1日と16日に医中誌Webにアップされる。

以上の工程に、最短で1ヶ月、平均3~4ヶ月かかっており、この期間を少しでも短縮することがデータベース作成上の大きな課題となっている。

(7) インデキシングとマッピングを支える用語管理

「医中誌Web」の検索を補完する機能として、「同義語から統制語へのマッピング機能」がある。この機能は、もともとはインデクシングサポート機能として利用していた仕組みを検索に応用したもので、この仕組みを支えているのが「用語管理」である。インデクサーは、日々文献と接していくなかで新しい用語に出会う。病名のことでもあれば物質名のことでもあり、それらの新規語について、さまざまな辞書や資料を調べ、また文献で、どのように表現されていることが多いのかなどをチェックし、見出し語と同義語の関係を定めていく。

一昨年発生した「SARS」を例にあげますと、医中誌では一昨年、4月に、フリーキーワードとして登録したが、その際、この用語の見出し語を「重症急性呼吸器症候群」と定め、その同義語として「Severe Acute Respiratory Syndrome」「新型肺炎」「SARS (Severe Acute Respiratory Syndrome)」などを登録した。

このような処理により、いずれの言葉で表現されていても、インデクシングでは間違いなく「重症急性呼吸器症候群」が付与される。そして、検索の際には、たとえば「新型肺炎」と入力すると、自動的に、「重症急性呼吸器症候群」が付与された文献がヒットすることになる。

このような日々の用語管理作業を通して収集された用語数は現在45万余語にのぼっており、これらが、「医中誌Web」の検索機能に生かされ、利用者はシソーラスをそれほど意識することなく、検索できるように工夫されている。

(8) EBM対応データ作成への取り組み

「医中誌基本データベース」では、1999年以降の作成データのうち、原著論文について「研究デザイン」という視点からチェックし、該当する文献に「メタアナリシス: Meta-Analysis (MA)」「ラ

ンダム化比較試験: Randomized Controlled Trials (RCT)」「比較臨床試験: Controlled Clinical Trials (CCT)」「比較研究: Comparative Study (CS)」の4種類の研究デザインタグを付与している。

① 比較臨床試験

「比較臨床試験」は、臨床試験の対象者を複数のグループに分ける際の方法が正確なランダム割り付けではないが、それにかなり近い方法で行われている場合にタグが付与されている。したがって、非常に狭い定義で付与していると言える。この定義からすれば、「準ランダム化比較試験」とした方が良いのではないかと、というご意見もある。

② 比較研究

「比較研究」は、診療ガイドラインを作成する際などに有益なデータを収集しやすいようにとの配慮から設けられ、臨床文献のうち、比較分析を行っている研究に付与しているもので、次の2つの方法で研究されている場合が代表的なものである。一つは、対象者を複数のグループに分けて比較している場合。もう一つは、治療や診断などの方法が複数あり、それを比較している場合である。たとえば、花粉症の患者にある薬物を投与して、効果をみている場合、その薬物を投与する患者群と投与しない患者群をつくり、結果を比較している場合が前者です。後者は、花粉症の治療方法として、薬物療法とレーザー治療の、どちらが有効かを、実際に患者に行って、比較するなど、あるいは過去のデータを収集して比較検討している場合などを指している。さらに、疾患のリスクを調べる「コホート研究」や、健常者と患者とを比較する「症例対照研究」にも付与されている。「比較研究」は、2003年以降の作成データのみには付与されているが、2004年7月1日現在の付与データ数は、13,311件。他の3つの研究デザインは1999年まで遡って付与されているにも関わらず、付与総数は2,427件であるので、それを大きく上回っている。しかし、疾患ごとに検索して「比較研究」で絞り込むとそれほど多くはなく。前述の花粉症の比較研究は60件である。ちなみにRCTは12件。「比較研究」は膨大なデータを絞り込むツールとして一つの有効な方法ではないかと思っている。