紹介

田代 朋子*

新しいシソーラス管理システム「Word Clearing」 を利用した病名辞書の作成

Compilation of disease name thesaurus using a novel thesaurus management system "Word Clearing"

TASHIRO Tomoko

Introductory article

[著者抄録] 日本で初のシソーラス管理専用システム「Word Clearing」を利用した病名辞書(病名 T 辞書)の作成を紹介する。病名 T 辞書は同一概念の病名に同一コードを付与し、そのコードによって用語の関係付けを行っている。「Word Clearing」はこのようにコード管理体系から用語展開を行い、用語関係(同義語、上位語、下位語、関連語)を 1 画面に表示する機能のほか、階層関係のチェック、検索式としての出力などシソーラス管理に必要な機能を備えている。エンドユーザがシソーラスを簡単に利用する必要性が高まる中で、利用価値が高いシステムである。最後に、シソーラスもデータベースの所有物ではなく、汎用性を持つべきであるとの意見を加えた。

[著者付与キーワード] シソーラス, シソーラス管理システム, Word Clearing, 病名辞書, エンドユーザ

[Author Abstract] Disease name thesaurus has been compiled by using "Word Clearing", probably the first thesaurus management system in Japan. It assigns a unique code on individual disease concept and the relation between disease names can be understood from the code. Some of useful functions of "Word Clearing" are (1) displaying a hierarchial relation between terms visually in one frame (synonyms, broader terms, narrower terms and related terms); (2) checking of hierarchial relation of terms; (3) automatically formation of search statements. It has high value since there is a growing need toward user-friendly thesaurus. Author indicates a direction of future thesaurus: thesaurus should not be a slave of a database, and must have wide applicabilites to varius database management systems.

[Keywords by Author] thesaurus, thesaurus management system, Word Clearing, disease name thesaurus, enduser

* (株)ティ 辞書企画 (〒981 仙台市泉区南光台 2-12-13) Tel. 022(233)6401
T-JISYO KIKAKU (12-13, Nankoudai 2-chome, Izumi-ku, Sendai-shi, 981)

1. はじめに

「データベースで検索する」といえば、ひと昔前までは専門のサーチャーが検索するのが主であった。しかし、近年、システムやソフトの急速な進歩と共にインターネットがパソコンによる検索をより身近なものとし、“誰もが検索できる環境にある”という体制が確立しつつある。

しかし、この簡便さの中での落とし穴が気になる。〈リンゴジュース〉で検索し、5件ヒットさせ、「私でも簡単に検索ができるんだ!」と思っている人が〈りんごジュース〉の10件や〈アップルジュース〉の15件の情報を見逃し、その見逃した情報の中に、重要な情報が含まれていることがないか? 文字の検索は実にシビアであり、なんの工夫もないデータベースでは一文字違っていても検索してはくれない。全文データベースとなれば、なおさら文中の文字の統一は困難を極め、〈Appleジュース〉や〈リンゴ・ジュース〉までもが無造作に使われて、検索モレしてしまう可能性がある。

1989年に副作用データベース「ARIS」(小田島作成)の副作用用語辞書を作成した時に、このような検索モレがないようにと、同義語、下位語を整理した副作用用語辞書を作成した。いわゆる副作用用語のシソーラスである。作成時、念頭に置いたことは、“簡単に使えて、的確な検索ができるシソーラス”である。“簡単に使う”ためには同義語を多く収録し、思いついた用語からでも検索できるような整備が必要であったし、また、“的確な検索”のためには、その思いついた語に含まれる概念をまとめて検索できるような下位語の整備が必要であった。

医学用語の分類や階層は大変複雑で、一つの用語が複数の概念に属することが頻発する。また、複合語も多く、一つの小概念の中でさえ、複雑な上下関係が存在することもある。既存の辞書類を利用しようと調べたが、英語のシソーラスを元にしたものは、日本語に翻訳すれば同じ用語となる英語に別々のコードが付与しており、また、他の辞書も重複分類となると、細かいところまでは

フォローされていなかった。そこで、結局それらの辞書類を参考としてオリジナルシソーラスを作成することになった。その際、複雑な用語の関係をコードを使って整理したが¹⁾、コードで関連付けしたものを即座に用語に展開できるシステムがなかった為に、作成した辞書のチェックや印字には多大な労力を費やした。

その後、この副作用シソーラスの構造をさらに改善して独自の病名辞書(以下、病名T辞書と称する)を作成したが、用語展開と印字物作成の問題点は残されたままであった。

そのような状況下で登場したのが「Word Clearing」(開発・販売: 小田日本データベース開発)である。このソフトは、まさにシソーラスを管理するための専門ソフトであり、今までに存在しなかった種類のソフトである。そこで、用語の整理に光明を与えてくれたこのソフトを病名T辞書の作成を例として以下に紹介する。

なお、以下「辞書」という用語を、用語の関係を整理してあるシソーラスという意味で用いることとする。

2. 病名T辞書の構造

病名T辞書は、大分類も含めすべて同一体系のコード付与と、用語関係付けもそのコードを用いて行っていること、そして区分による同義語の整理を特徴とする。

2.1 コードによる関連付け

病名T辞書では一つの病名に対し一つのユニークコードが付与されている。正確に言えば、同義語にも同一のコードが付与されているので、一つの病気概念に対し一つのコードが振られているといえる。これは、同義語も含め、同じ病名はどこにおいても同一コードに置き換えて使えるように考えたからである。コードは一つの体系を成しており、例えば循環器系疾患はS21、呼吸器系はS22と決まっているが、付与されたコード以外の分類にも属する場合には「他系上位コード」とい

う項目の中に、そのコードを入れてある（従来は補助上位コードと呼んでいたが、本稿では「Word Clearing」で使われている他系上位コードという用語を使用する）。関連語についても同様に、「関連語コード」の項目にコードで入れてある。

図1に病名T辞書の項目一覧を示し、図2にコード体系例を示す。コードはSから始まり、各階層2桁、最高6階層（12桁）である。

2. 2 同義語を整理する区分

さらにそれぞれの用語には区分を設けた。区分とは同じ概念の用語の別表記を整理するためであるが、別表記にも種類があるので、以下の5種類に分けた。

- 区分1－統制語
- 区分2－同義語
- 区分3－準同義語（検索同義語）
- 区分5－英語同義語
- 区分6－略号

区分3の準同義語は、正確な意味では同義ではないが、検索上は一緒に検索できた方が好ましい（と考えられる）語である。また、区分4は準々同義語のような語が発生することを考えて空けておいたが、最終的には使用しなかった。これらの区分の実例を図3に示す。

3. 病名T辞書の「Word Clearing」への取り込み

「Word Clearing」は病名T辞書の構造をそのまま取り込むことができるWindows 3.1および95用のソフトである。病名T辞書の場合、すでに別のパソコンソフトで構築中であつたので、それをCSV形式で出力し、さらに「Word Clearing」の編集用エディタの機能を使って、SDAファイルに変換した。CSV形式のファイルから「Word Clearing」の辞書登録までの工程を図4に示した。なお、基本元用語として登録する時に、用語関係

項 目	内 容 例
コード	S1317
区 分	1
病 名	進行性筋ジストロフィー症
ヨ ミ	シヨウセイキンジストロフィーショウ
英 語	PROGRESSIVE MUSCULAR DYSTROPHY
説 明	遺伝性筋症。進行性の筋の消耗と衰弱を主体とし、デュシェンヌ型、ベッカー型、肢体型、顔面肩甲上腕型などがある。
補助上位コード	S542;S1381 ※複数可
関連語コード	S113418 ※複数可

レコード数：8,637件

図1 病名T辞書の項目と内容

の矛盾などを自動的にチェックし、エラーリストが出力される。チェック機能については次項であらためて説明するが、病名 T 辞書の取り込みでもエラーリストを見ながら元データを修正する作業が必要であった。

4. 「Word Clearing」の基本機能

取り込んだ後にその機能を利用して病名 T 辞書の見直しを行った。特に便利な機能は用語の展開機能と、統制語から下位語へ、下位語から関連語へと自由に渡り歩きその関係が 1 画面で一目で確認できる表示機能であった。その他にも辞書専用

図2 コード体系

コード	区分	病 名	ヨ ミ	英 語
S141112	1	躁状態	ソウゾウタイ	MANIC STATE
S14111211	1	躁病	ソウビョウ	MANIA
S14111211	2	そう病	ソウビョウ	
S14111211	3	躁病性興奮	ソウビョウセイコウフン	MANIC EXCITEMENT
S1411121112	1	苦悶性躁病	クモンセイソウビョウ	ANXIETY MANIA
S14111212	1	饒舌	ジョウセツ	LOGOMANIA
S14111212	2	多弁	タヘン	TALKATIVENESS
S141115	1	躁うつ病	ソウウツビョウ	MANIC DEPRESSIVE PSYCHOSIS
S141115	2	交代精神病	コウタイセイシンビョウ	FOLIE ALTERNE
S141115	2	感情精神病	カンジョウセイシンビョウ	AFFECTIVE PSYCHOSIS
S141115	5	躁うつ病	ソウウツビョウ	MANIC DEPRESSIVE IRRESEIN
S141115	6	M D I	MDI	MDI

図3 区分の実例

図4 Word Clearing への病名 T 辞書の登録作業

ソフトならではの配慮が随所にみられるので、病名T辞書作成例を図で示しながら、その機能を簡単に紹介する。

4. 1 検索・表示機能

検索はコード、用語、カナ、英語の4項目からワイルドカードを用いて、前方一致、完全一致、中間一致検索が可能である。検索結果表示スタイルは2種類あり、スタイル1ではさらに2種の表示フォームがあり、スタイル2では複数の辞書(最高三つ)を呼び出して、検索・表示することができる。どの表示スタイルでも画面上に表示されている上位語、下位語、関連語のいずれにも自由に移動し、ダブルクリックで移動した用語を中心とした用語関係を見ることができる。図5に「進行性筋ジストロフィー」の表示形式例(表示スタイル2)と、さらにその上位語の一つである「ミオパチー」に移動した時の画面例(表示スタイル1)を示した。

4. 2 用語関係の展開機能

コードで関連付けを行っている辞書の展開が複雑であることは容易に想像されよう。他系上位コードとして付与したコードにさらに他系上位コードがついている場合などは、二重、三重に展開しなければならぬ。この点、「Word Clearing」は直近の上位語さえ登録すれば、あとは自動的に上位語を展開してくれる。図6に「慢性痛風性関節炎」での他系上位コード付与ならびに「多発性関節症」について関連コード付与の実例を示す。

4. 3 辞書メンテナンス

用語の修正、追加、削除など一般的なメンテナンスはもちろん可能であるので、特殊なメンテナンスを二つ紹介する。

4. 3. 1 優先語削除・コード変換

優先語を削除する場合には関連する下位語などをどのように取り扱うかを決めなければならない。「Word Clearing」はこれを会話形式で処理できる。その処理も何通りかに分岐し、削除するの

ではなく、コードを変換することもできる。コードを変換した場合には、そのコードに関係する用語関係のコードが一括で更新処理される。

4. 3. 2 用語複写

用語複写は別々に構築していた用語辞書からの取り込みができる機能である。複数の辞書を作業対象としておいて、複写元の辞書から用語を検索・選択し、複写先の辞書に一括登録できる。この時一つ一つコード付けを行わなくても、最上位のコードを決定すれば下位語は自動的に付与する機能もある。二つの辞書のコード体系が異なった場合でも問題はない。図7にコード体系の異なる辞書からの用語複写画面を示す。

4. 4 辞書体系および文字種のチェック機能

用語関係の矛盾や文字種などは、基本元用語として登録する時に自動的にチェックされる。文字種はユーザがニーズに合わせ、全角だけにするのか半角も許すのか等の指定ができる。以下にチェックされるエラー例をあげる。

- ・優先コードが存在しない
- ・階層構造がループする
- ・階層構造に誤りがある
- ・優先語が重複する
- ・他系コード内に重複がある
- ・用語に半角文字が含まれる
(全角指定をした場合)

4. 5 出力

出力は、画面、プリンタ、ファイルへの3通りが可能であり、出力形式も項目選択、何階層まで出力するかを選択などができる。図8に出力様式の設定画面と出力例を示す。

4. 6 検索語出力

実際に「Word Clearing」で管理した辞書をオンライン検索などに利用する方法として、検索語のクリップボードへの出力機能がある。この時に、検索時にそのまま使えるように任意の演算子を含んだ検索式として出力できる。

<表示スタイル2>

↓ ミオパチーを中心とした表示に変更

<表示スタイル1>

図5 検索結果表示画面

コード	用語	付与した関係付けコード
S16	結合組織・運動器疾患	
S1613	関節疾患	
S161311	関節炎	関連語：S161312;S161313 (関節症;関節痛)
S16131112	慢性関節炎	
S16131116	化膿性関節炎	他系上位：S6111 (細菌感染)
S16131117	多発性関節炎	関連語：S16131212;S16131311(多発性関節症;多発性関節痛)
S16131171	外傷後関節炎	他系上位：S16131271 (外傷)
S16131172	痛風性関節炎	他系上位：S3127 (痛風)
S1613117211	慢性痛風性関節炎	他系上位：S16131112 (慢性関節炎) ←
S161312	関節症	
S16131212	多発性関節症	←
S161313	関節痛	
S16131311	多発性関節痛	他系上位：S7611 (疼痛)

↓
自動展開後

<p>S1613117211 慢性痛風性関節炎</p> <p>上位語 (B T)</p> <p>S16131172 痛風性関節炎 ◎</p> <p>S16131112 慢性関節炎 ○</p> <p>・S161311 ・関節炎 ◎</p> <p>・S3127 ・痛風 △</p> <p>・S77121112 ・高尿酸血症 △</p> <p>・S1613 ・関節疾患 ◎</p> <p>・S31 ・代謝・栄養障害 △</p> <p>・S16 ・結合組織・運動器疾患 ◎</p> <p>・S771211 ・高窒素血症 △</p> <p>・S7712 ・窒素化合物の異常 △</p> <p>・S77 ・検査値異常 △</p>	<p>S16131212 多発性関節症</p> <p>上位語 (B T)</p> <p>S161312 関節症 ◎</p> <p>・S1613 ・関節疾患 ◎</p> <p>・S16 ・結合組織・運動器疾患 ◎</p> <p>関連語 (R T)</p> <p>S16131311 多発性関節痛 △</p> <p>S16131117 多発性関節炎 △</p> <p>◎ 直系上位 ○ 他系上位 (コード付けによる) △ 自動付与</p>
--	---

図6 上位語の自動付与例

図7 用語複写画面

4.7 ユーザインタフェース

「Word Clearing」の作業実態はコードにより用語を管理・展開しているが、インタフェースを良くするために主な作業は用語で行うことができるようになっている。例えば、図9は上位コードを追加する画面であるが、現在の辞書登録用語から選択して上位コードを追加できる。これはコードによる入力ミスの防止にもなる。なお、エンドユーザが利用する場合にコードを気にしなければならないことは全くないといってよい。

4.8 セキュリティ

「Word Clearing」はスタンドアロンでもネットワークでも使用できる。その際気になるのがセキュリティであるが、ユーザレベルに4階層あり、一般の検索者に辞書内容の変更などはできないようになっている。一番上の階層は「システム管理者」であり、その下の3階層のユーザの設定およ

び辞書の登録・削除はこのシステム管理者しか操作できない。次の「管理者(階層)」は各辞書への用語の登録ができ、さらにその下の「管理者」は用語メンテナンスまでができ、末端の「検索者」は検索および印刷のみの利用となっている。また、はじめから辞書の検索ソフトとしてだけ使用したいという場合には検索機能のみの簡易版も提供されている。

5. 病名T辞書の応用例

このように、「Word Clearing」の機能は必要に応じて分岐しており、多様なニーズに合うように設計してある。辞書は本来ユーザのニーズに応じて更新していくものであるが、病名辞書の類は“何が元であり何を追加したか?”を問われる場合がある。そこで、最後に既存の病名T辞書の用語に独自の用語を追加した例を紹介する。

印刷様式の設定

<div>階層のインデント有無 ☒ する ☐ しない </div> <div>階層のインデント文字 ☒ 中黒 ☐ スペース </div> <div>印刷階層レベル 階層レベル指定 <input type="text" value="6階層まで"/> </div> <div>印刷帳票タイプ 帳票タイプ指定 <input type="text" value="A4縦型"/> </div>	<div>印刷対象項目の選択 (見出し語)</div> <div>印刷対象項目の選択 (関係語)</div> <table style="width: 100%;"> <tr> <td>☒ コード</td> <td>☒ コード</td> </tr> <tr> <td>☒ 用語</td> <td>☒ 用語</td> </tr> <tr> <td>☐ カナ読み</td> <td>☐ カナ読み</td> </tr> <tr> <td>☐ 英語</td> <td>☐ 英語</td> </tr> <tr> <td>☒ 説明</td> <td>☐ 説明</td> </tr> <tr> <td>☐ メモ</td> <td>☐ メモ</td> </tr> </table>	☒ コード	☒ コード	☒ 用語	☒ 用語	☐ カナ読み	☐ カナ読み	☐ 英語	☐ 英語	☒ 説明	☐ 説明	☐ メモ	☐ メモ
☒ コード	☒ コード												
☒ 用語	☒ 用語												
☐ カナ読み	☐ カナ読み												
☐ 英語	☐ 英語												
☒ 説明	☐ 説明												
☐ メモ	☐ メモ												

↓

出力

S1613
 アルトロパチー
 説明：関節の病変の総称として用いられる。
 優先語 (USE)
 S1613
 関節疾患

S1613
 関節疾患
 同義語 (UF)
 S1613
 アルトロパチー
 上位語 (BT)
 S16
 結合組織・運動器疾患
 下位語 (NT)
 S161317
 関節拘縮
 S161318
 関節強直
 S161322
 脱臼
 S161312
 略

図8 出力様式の設定画面と出力結果（関節疾患について実行）

図9 上位語追加操作画面

S241111	ウイルス性肝炎
・J01	・急性ウイルス性肝炎
・J0101	・・A型急性肝炎
・J02	・麻疹肝炎
・S24111111	・A型肝炎
・J0101	・・A型急性肝炎
・S24111112	・B型肝炎
・J03	・・B型劇症肝炎
・J04	・・輸血後B型肝炎
・J05	・・B型肝炎ウイルスキャリア
・S2411111211	・・B型慢性肝炎
・S2411111212	・・B型肝炎ウイルス母子感染
・S24111113	・C型肝炎
・S24111121	・輸血後肝炎
S241112	アルコール性肝炎
S241113	胆汁うっ滞型肝炎
・S24111311	・胆汁うっ滞型慢性肝炎
・S24111312	・胆汁うっ滞型急性肝炎
・S24111321	・中毒性細胆管炎
・S24112111	・中毒性閉塞性肝炎
S241121	中毒性肝炎
・S24111321	・中毒性細胆管炎

※ Jで始まるコードは後から追加した用語

図10 病名 T 辞書利用例

図10の一覧でJから始まるコードは後から追加した用語である。コード的には元辞書の分類の中に組み込まずに、他系上位コードを利用して関連付けを行った例である。先ほどの図9は、この時の「B型劇症肝炎」に対する上位語追加画面である。このようにすれば、検索上は追加した用語も利用できるが、辞書内ではコードが分離しており、追加語だけを簡単に取り出すこともできる。

6. お わ り に

用語の整理という仕事はどちらかというと裏方の仕事であり、手作業が多い地味な仕事である。しかし、データベースが出来た当初からずっと課題とされ続け、研究が続けられている分野であり、データベースの検索結果の是非を決める大切な要因となっている。

従来、辞書はデータベースの付随物として作成され、その結果、データベースごとに異なる辞書が使われるために、度々横断検索上の壁となってきた。どんなに使いやすいシステムが開発されて

も、必要な情報が検索できなければなんの意味もない。誰もが簡単にデータベースを検索できる時代、さらに全文データベースが急増する中で、用語整備の必要性は高まっているといえる。「Word Clearing」の出現を契機として、用語整備の重要性が見直され、さらに整備された辞書が一つのデータベースの所有物としてではなく、検索に必携の標準辞書として広く普及することの必要性が再考されればよいと願っている。最近では1996年10月の第33回情報科学技術研究集会で、「汎用シソーラスブラウザの試作」という演題が発表され²⁾、エンドユーザにおいてのシソーラス利用の重要性が再確認された。ブラウザの次には、汎用シソーラスこそ必要であると思われる。海外ではすでに、複数のシソーラス類を統合したシソーラスや、同義語類を多数集めたシステムなども提供されはじめている。作成した病名T辞書は現在「Word Clearing」のオプション辞書（または辞書単体）として提供されているが、実用的シソーラスの単独での普及の小さな一歩となればよいと思う。

参 考 文 献

- 1) 田代朋子. 同義語・関連語処理機能付き副作用用語シソーラス「HYTARD」. 情報の科学と技術 41(11)873-882(1991)
- 2) 下村央人. 汎用シソーラスブラウザの試作. 第33回情報科学技術研究集会予稿集 45(1996)