JMEDICINEファイルの医中誌記事への JICSTシソーラス語の自動付与

鈴木 博*

Automatic addition of descriptors of JICST thesaurus to the ICHUSHI subfile of JMEDICINE file

SUZUKI Hiroshi

〔著者抄録〕 JICST・医中誌国内医学文献ファイル(JMEDICINE ファイル)は、二つの部分ファイルで構成されている。一つは JICST シソーラス用語が付与される部分で、もう一つは医中誌作成の医学用語シソーラス用語が付与される部分である。JICST では、医学用語シソーラスの用語に JICST シソーラスの用語を対応させた医中誌-JICST シソーラス用語対応辞書を作成し、JMEDICINE ファイルの医中誌分記事に JICST シソーラスの用語を自動付加した。この対応辞書の詳細、自動付加システム、検索上の改善点について述べた。

〔著者付与キーワード〕 医学情報、文献データベース、シソーラス、索引語、情報検索、JMEDICINE ファイル

〔Author Abstract〕 JICST-ICHUSHI domestic medical literature file (JMEDICINE file) comprises two part files. One is indexed according to JICST thesaurus terms and another is indexed according to ICHUSHI's medical term thesaurus terms. JICST made corresponding dictionary for ICHUSHI-JICST thesaurus terms and added automatically JICST thesaurus terms to the ICHUSHI part of JMEDICINE file. This paper describes the details of the corresponding dictionary, automatic adding system of the dictionary terms, and the benefit which were improved for searching the file.

〔Keywords by Author〕 medical information, literature database, thesaurus, index term, information retrieval, JMEDICINE file

* 日本科学技術情報センター (〒100 千代田区永田町 2-5-2) Tel.03(3581)6411
JICST (5-2, Nagatacho 2-chome, Chiyoda-ku, 100)

1. はじめに

JICST・医中誌国内医学文献ファイル(JMEDICINE ファイル)は、日本で発行される資料に掲載された医学文献を年間約23万件収録する文献データベースで、1981年から1994年5月までの総蓄積件数は約250万件となっている。作成・提供機関は日本科学技術情報センター(JICST)である。

JMEDICINE ファイルは1981年10月からJOISでのサービスが開始され、当初はJICSTのシソーラス語と分類を付与した記事のみを収録していた。その後、1986年10月から医学中央雑誌刊行会(医中誌)が作成する医中誌基本データベースよりJICSTで作成していない記事を抽出したものを加え、国内の医学文献を網羅するJICST・医中誌国内医学文献ファイルとしてサービスされるようになった。

このような経緯で作成されたため、JMEDICINE ファイルは二つの部分ファイルで構成されていることになる。一つはJICST科学技術シソーラス(以下JICSTシソーラス)とJICST科学技術分類表により索引される部分で、JICST作成分と呼ぶ。もう一つは医中誌作成分と呼ぶ部分で、医中誌基本データベースからJICST作成分と重複する部分を削除したもので、医中誌作成の医学用語シソーラスと医学中央雑誌科目別分類表に基づいて索引されている。

JICSTシソーラスは科学技術の全分野を網羅する総合シソーラスであり、一方医学用語シソーラスは医学領域の専門シソーラスであるなど、両シソーラスはその体系、思想、収録用語数などに大きな違いがある。そのため、従来JMEDICINE ファイルを検索する際には、JICSTシソーラスの用語と医学用語シソーラスの用語を調べた上、両者が異なる場合は両者を入力し、論理和を作るなどといった処理が必要であった。このことがJMEDICINE ファイルの利用上の大きな問題点であった。

そこで、JICSTでは、医学用語シソーラスの用

語にJICSTシソーラスの用語を対応させた医中誌-JICSTシソーラス用語対応辞書を作成し、JMEDICINE ファイルの医中誌分記事にJICSTシソーラスの用語を自動付加することにより、JMEDICINE ファイルを利用しやすく改善した。

2. 自動付加システムの概要

自動付加システムの概要を図1に示す。まず、医中誌基本データベースからの抽出原データに医中誌-JICSTシソーラス用語対応辞書によりJICSTシソーラス用語を付与する。ここで原データにある医学用語シソーラスに基づくキーワードの漢字表記データを辞書マッチングキーとして利用する。原データにはカタカナ読みのデータも存在するが、カタカナ読みのデータをマッチングキーとして利用した場合、同音異義語を区別できないためである。

また、複数の医学用語シソーラスの用語が同一のJICSTシソーラスの用語に対応している場合もあるため、マッチングによって発生したJICSTシソーラス用語について重複削除処理を行う。

さらに、JOISのファイル作成上キーワードの総数の上限は39個であるので、もともと医中誌基本データベースからの抽出原データにある医中誌キーワードと、マッチングによって発生したJICSTキーワードがこの上限値を越えないよう、必要であれば削除処理が行われる。

3. 対応辞書

3.1 対応辞書の作成とファイル作成

JICST作成分に適用されるJICSTシソーラスは科学技術全分野を対象としており、1981年版、1987年版、1993年版と改訂されてきた。改訂ごとに医学関連の用語数が大幅に増加している。1993年版の収録語数は50,798語でそのうち索引に使用されるディスクリプタは39,046語である。医学関連用語はディスクリプタが約12,000語である。なお、バックファイルメンテナンス処理により、現

在は JICST 作成分の全記事が1993年版に基づいて索引が行われている。

一方、医中誌作成分は、1983年発行の医学用語シソーラス第一版、1987年発行の第二版、1994年発行の第三版により索引されている。しかし、バックファイルメンテナンス処理は行われていないので、1983年から1986年の記事は第一版で、1987年から1993年の記事は第二版で、1994年以降は第三版で索引されている。

医学用語シソーラスは、名前が示すごとく医学とその関連領域を対象としており、医学関連用語数についてみると、JICST シソーラスより収録語数が多くなっている。また、医中誌基本データベース作成に実際に利用されるのは医学用語シソーラスの磁気テープ版であり、冊子体のデータに新語の追加、若干の修正などの更新が毎年行われている。

対応辞書の作成は、医学用語シソーラス1990年版の用語に JICST シソーラス1987年版の用語を対応させる作業を、医学中央雑誌刊行会の協力を得て1990年に開始した。

当初の対応辞書の内容は、医学用語シソーラス磁気テープ1990年版に収録されている用語25,689語中、医学中央雑誌の巻末索引および年間索引の見出し語となる約17,700語に JICST シソーラス1987年版のディスクリプタ(索引に使用する用語)を対応させたものであった。また、索引の見出し語とならない約7,900語に対しては、"see" 参照先の医学用語シソーラス用語に対応する JICST シソーラス用語を対応させることとした。

その後、JICST シソーラスが1993年版に改訂された後、この対応辞書をもとに1990年版医学用語シソーラス磁気テープ版・1993年版 JICST シソーラスの対応辞書を作成した。この対応辞書により1994年1月に医中誌作成分の全記事に対する JICST シソーラス用語の付与を行った。

さらに、1994年に医学用語シソーラスの第三版が完成した後に、第三版医学用語シソーラス(1994年版)・1993年版 JICST シソーラスの対応辞書を作成し、第三版医学用語シソーラスで索引されて

いる医中誌作成分の全記事への JICST シソーラス用語付与を1994年5月に行った。第三版医学用語シソーラス(1994年版)・1993年版 JICST シソーラス対応辞書は、見出し語約14,300語および非見出し語約8,400語に対する対応関係を収録している。

なお、医中誌作成分記事の一部に JIS 第一水

図1 自動付加システム概要

準, 第二水準語以外の外字が使用されているため, 漢字部分が空欄になっているものが少数だが存在する。これについては対応辞書にないため, JICST シソーラス用語は付加されない。

表 1 に現在 JMEDICINE ファイル作成に適用されている, 二つの対応辞書とその適用範囲を示す。

3. 2 対応辞書の詳細

医学用語シソーラスと JICST シソーラスは用語数も体系も異なるため, 単純に用語の対応関係が付いているわけではない。用語の対応関係のパターンと例を表 2 に示す。

1 対 1 の対応になっている場合でも, (a)完全に同型の場合, (b)読みは同じで表記が異なる同音異形語の場合, (c)医中誌用語が倒置形の場合, (d)医

表 1 対応辞書統計

	90医中誌-93JICST 対応辞書	94医中誌-93JICST 対応辞書
収録語総数	25,689	22,705
ディスクリプタ	17,768	14,298
非ディスクリプタ	7,921	8,412
対応 JICST 用語なし	178	103
適用範囲	1981.01-1993.11	1993.12以後

注 1 副標目, チェックタグはディスクリプタに含めてある。

注 2 適用範囲はファイルレンジであり, 1993.11とは1993年度の11号で, これは1994年2月更新分にあたる。

表 2 対応辞書のパターン

1 対 1 対応の例		
パターン	医学用語シソーラス語	JICST シソーラス語
a 完全同形	視神経	視神経
b 異形語	馬	ウマ
c 倒置形の語	肝炎-B 型	B 型肝炎
d 英語名	Penicillamine	ペニシラミン
e 同義語	生活の質	クオリティオブライフ
上位語が対応する例		
パターン	医学用語シソーラス語	JICST シソーラス語
f 1 語が対応	中足指節関節	指関節
g 2 語が対応	喉頭麻痺	喉頭疾患, 麻痺性疾患
h 3 語が対応	肉芽腫症-リンパ腫様	肉芽腫, 前癌状態, リンパ系疾患
対応関係が付かない例		
死亡診断書		
良性 (副標目)		
代謝・吸収・排泄 (副標目)		

表3 医中誌副標目-JICST シソーラス用語対応リスト

	医中誌副標目	JICST 用語		医中誌副標目	JICST 用語
I 群	解剖学・組織学	—	III 群	微生物学	微生物学
	細胞学	生物学		寄生虫学	生物学
	超微構造	微細構造		傾向	—
	神経支配	—		病因	病理
	類似体・誘導体	—		病原性	病原性
	組成・成分	組成		有害作用・毒性	毒性
	分離・精製・同定	—		培地・培養・模型	—
	生合成	生合成		拮抗物質・阻害物質	—
	化学合成	化学合成		抗原・抗体・補体	—
	分析・検出	分析		免疫学	医学
	測定・定量	計測		病態生理学	病態生理
	計測・検査	計測		病理学	病理学
II 群	生理学	生理学		疫学	疫学
	生化学	生化学		統計	統計
	酵素学	—		死亡・死亡率	死
	薬力学・薬理学	—		分類	分類
	動力学	動力学		調査・実態	調査
	分泌	分泌		心理学	心理学
	代謝・吸収・排泄	—		実験	実験
	発生学・遺伝学	—		診断	診断
	抗菌性・抗腫瘍性	—		X線・超音波診断	画像診断
	脳脊髄液	髄液		放射線核種イメージング・画像診断	画像診断
	血液	血液		診断的利用	診断
	性状・分布	—		基準	標準
	尿	尿			
	成長・発育	成長			
	薬物影響	薬理作用			
	放射線影響	放射線効果			

I 群 解剖学的なものおよび化学物質などの分析・測定に用いられる

II 群 生理・生化学的な性状およびその影響・効果などに用いられる

III 群 疫学・免疫学・診断学的なものに用いられる

中誌用語が英語標記の場合、(e)同義語の場合があり、JICST シソーラス用語にこの(a)から(e)にあてはまるものがない場合は、上位概念にあたる用語を対応させた。上位語対応の場合、一つの医中誌用語に複数の JICST シソーラス用語を対応させたケースもある。

例えば、表2の(g)の「喉頭麻痺」については、JICST シソーラス用語にあてはまる用語がないため、上位概念の語を対応させた。上位概念の用語としては「喉頭疾患」と「麻痺性疾患」の二つがあるのでこの両者を対応させた。

医学用語シソーラスの約20%の用語に対して二

(表3 続き)

	医中誌副標目	JICST 用語		医中誌副標目	JICST 用語	
IV 群	異常・病変・症状	症状	V 群	薬物療法・投与量	薬物療法	
	良性	—		救急療法	救急療法	
	悪性	悪性疾患		精神療法	精神療法	
	実験的	病態モデル		免疫療法	免疫療法	
	急性	—		外科的療法	外科術	
	慢性	—		移植	移植【動物】	
	先天性	—		理学的療法	理学療法	
	後天性	—		併用療法	併用療法	
	原発性・特発性・本態性	—		運動療法	運動療法	
	家族性・遺伝性	—		食事療法	食事療法	
	化学的誘発・副作用	副作用		治療	治療法	
	伝染性・流行性	—		予後	予後	
	合併症・後遺症・侵襲	合併症		看護・リハビリテーション	—	
	奇形・位置形態異常	奇形		予防・抑制	予防	
	損傷・外傷	外傷		対策・方法	—	
	退行性障害	—		衛生学	衛生	
	機能障害	機能障害		治療の利用	治療法	
	欠損症・欠乏症	—		器具・装置・機械化	—	
	代謝性障害	代謝性疾患		VI 群	組織・管理・体制	—
	過敏症・中毒	—			法制・法学・行政学	—
炎症・感染・感染症	—	経済学	経済学			
腫瘍・腫瘍	腫よう	人的資源・倫理学	—			
		学問・研究	研究			
		教育・指導	教育訓練			
		情報・歴史	—			
		学会	—			
IV群	症状および疾患・障害などに用いられる					
V群	治療・予防・対策および器具・装置などに用いられる					
VI群	医学関係に限定されない一般的なものに用いられる					

つ以上の JICST シソーラス用語が対応している。さらに、医学用語シソーラス用語には上位概念を考慮しても JICST シソーラス用語を対応できないものもある。これらの語については無理に対応関係をつけないほうがよいと判断した。

医学用語シソーラスには独特の用語として医中誌副標目と医中誌チェックタグがある。これらについてもできるかぎり JICST シソーラス用語との対応を付けた。表 3 に医中誌副標目-JICST シソーラス用語対応表を、表 4 に医中誌チェックタ

グ-JICST シソーラス用語対応表を示す。

医中誌副標目については、「良性」のように適切な JICST シソーラス用語がなく対応関係のついていないものがある。また、「代謝・吸収・排泄」のように複数の概念をひとまとめでした医中誌副標目についても、JICST シソーラス用語を対応できない。この場合、JICST シソーラスには「代謝」や「生体吸収」や「排せつ」というそれぞれの用語はあるが、そのうちどれを対応させるべきかを判断できないためである。

(表3 続き)

1985年までに用いられていた副標目
(ただし医中誌分の1986年以降の記事の中でも使用されている場合があります)

医中誌副標目	JICST 用語	医中誌副標目	JICST 用語
分離・精製	—	分析	分析
計測	計測	薬力学	—
代謝	代謝	発生学	生物学
遺伝学	遺伝学	性状	—
有害作用	毒性	毒性	毒性
死亡率	死亡率	実態	—
X線診断	ラジオグラフィー	放射線核種 imaging	シンチグラフィー
異常	—	病変	—
症状	症状	原発性	—
本態性	—	合併症	合併症
奇形	奇形	欠乏症	栄養障害
中毒	中毒	感染症	感染症
腫瘍	腫瘍	薬物療法	薬物療法
供給・分布	—	薬物投与・投与量	薬物療法
看護	看護	対策	—
方法	科学	組織・管理	—
人的資源・情報	—	教育	教育訓練
歴史	歴史		

4. 検索への適用

図2に JICST シソーラス用語が付加される以前(図2 A)と付加された以後(図2 B)の医中誌作成分の記事の一例を示す。JICST 作成分の記事と同様に、図2 Bの記事には表示されている JICST シソーラス用語に加えて、それらの JICST シソーラスにおける上位語も付与されている。そのため、検索時に JICST シソーラスの階層関係を利用して網羅的な検索が簡単にできる。

図3に対応辞書による自動付加が行われる以前の検索式と、自動付加実施後の現在の検索式を示す。前者では医学用語シソーラスの用語と JICST シソーラスの用語が異なる場合は両者を入力する必要がある。さらに、医学用語シソーラスでの下位語を一つ一つ入力し、論理和をとる必要がある。後者では JICST シソーラスの用語だけでよいので簡単でわかりやすい。

また、例えば「特定の薬物投与の副作用としての腎臓疾患」などといった検索テーマでは、従来「腎臓疾患」について医中誌作成分については、医学用語シソーラスでの「腎臓疾患」の下位語をすべて入力しないと検索もれの可能性があるため、網羅的な検索がほとんど不可能であった。現在では、JICST シソーラスの用語にある「腎臓病」を入力すれば、医中誌作成分を含めて網羅的な検索結果が得られる。

なお、医中誌作成分記事の FT 欄には医学用語シソーラスの見出し語、非見出し語の他に、医学用語シソーラスにない語がフリーキーワードとして付与されている。ここで見出し語については対応辞書により JICST シソーラス用語に変換され、非見出し語については”see” 参照先の医学用語シソーラス用語に対応する JICST シソーラス用語に変換される。しかし、フリーキーワードについては対応辞書にないため JICST シソーラス

表4 医中誌チェックタグ-JICST シソーラス用語対応表

	医中誌チェックタグ	JICST 用語	医中誌チェックタグ	JICST 用語
I 群	人*	ヒト*	鶏胚	—
	犬	イヌ	モルモット	モルモット
	猫	ネコ	ハムスター	ハムスター
	牛	ウシ	鼠	ネズミ類
	馬	ウマ	ラット	ラット
	豚	ブタ	蛙	カエル
	緬羊	ヒツジ	鳥類	鳥類
	猿	サル	魚類	魚類
	家兎	ウサギ	動物	動物
	鶏	ニワトリ	植物	植物

	医中誌チェックタグ	JICST 用語
II 群	妊婦	人間
	胎児	胎児
	新生児	新生児
	乳児 (一2)	乳児
	幼児 (3—5)	幼児
	小児 (6—12)	子供**
	青年期 (13—18)	青少年**
	成人 (19—44)	成人**
	中年 (45—64)	成人
	老人 (65—)	中年期 老人

	医中誌チェックタグ	JICST 用語
III 群	男	男性
	女	女性
	in vitro	in vitro 実験**
	症例報告*	症例報告*
	治療記録	病歴

	医中誌チェックタグ	JICST 用語
IV 群	古代	—
	中世 (江戸前)	—
	近世 (江戸期)	—
	近代 (明治一)	—
	現代 (戦後)	—

*は二次検索だけに使用でき、CK=ヒトあるいはCK=ショウレイホウコクのように検索する。

**は94年3月更新分(9312号)以降の記事にのみ付与されている。

用語への変換は行われない。

例えば「抗腫瘍薬の妊娠中の投与例」などといった検索テーマでは「抗腫瘍薬」で網羅的な検索ができるが、新薬名についてはフリーキーワードとして付与されており、対応辞書に収録されていないので、「抗腫瘍薬」での検索結果に含まれない。また新規の病名も対応辞書に収録されていない。

なお、前述したように適切な JICST シソーラス用語を対応できない医学用語シソーラス用語が、少数だが存在する。この場合は検索に医学用語シソーラス用語を用いることになる。さらに、

JICST シソーラス用語に対応していない医中誌副標目を使用するほうがよい場合もある。その一例を図4に示した。

5. お わ り に

JICST と医中誌ではシソーラスが異なるだけでなく、索引規則やファイル作成手法が異なる。そのため、対応辞書による JICST シソーラス用語の付加により JMEDICINE ファイル検索上の課題のすべてが完全に解決するわけではない。しかし、対応辞書による JICST シソーラス用語の

付与により、一般の利用者にとってJMEDICINE
ファイルの検索は非常に容易になったと思う。今
後対応辞書の不適切なところについて改善を加え

て行くと共に、他の面での改良を加えより一層利
用しやすいファイルとすることが次の課題であ
る。

A JICST シソーラス用語付加以前

#000001 JICST COPYRIGHT

CN 94D0008498

TI インターフェロン投与により精神障害が出現したB型慢性肝炎の2例

AU 服部宜裕(庄原赤十字病院(5076));外9名

JN W1409 日消病会誌

VN VOL. 90, NO. 臨増 PAGE. 1161 1993

CI (C)(J1)(JA)(JPN)

BD (94008498) 34歳男, 42歳男

CC 14C6

FT [肝炎-B型; Interferon; 精神障害; JAMAS; 慢性; 免疫療法; 有
害作用・毒性; 治療的利用; 化学的誘発・副作用; 人; 成人(19~44); 男]

B JICST シソーラス用語付加後

#000001 JICST COPYRIGHT

CN 94D0008498

TI インターフェロン投与により精神障害が出現したB型慢性肝炎の2例

AU 服部宜裕(庄原赤十字病院(5076));外9名

JN W1409 日消病会誌

VN VOL. 90, NO. 臨増 PAGE. 1161 1993

CI (C)(J1)(JA)(JPN)

BD (94008498) 34歳男, 42歳男

CC 14C6

KW B型肝炎; インターフェロン; 精神障害; 免疫療法; 毒性; 治療法; 副作用; ヒト;
成人; 男性

FT [肝炎-B型; Interferon; 精神障害; JAMAS; 慢性; 免疫療法; 有
害作用・毒性; 治療的利用; 化学的誘発・副作用; 人; 成人(19~44); 男]

図2 医中誌作成分記事の例

検索テーマ：B型肝炎のインターフェロンによる治療における副作用

A 自動付加がない従来の検索式

(医学用語シソーラス用語と JICST シソーラス用語を併用する)

U: Bカ`タカンエン+カンエン-Bカ`タ

[1] S: 9,441 Bカ`タカンエン+カンエン-Bカ`タ

U: インターフェロン+INTERFERON&

[2] S: 11,820 インターフェロン+INTERFERON& (注)

U: フクサヨウ+カガ`クテキユウハツ フクサヨウ

[3] S: 102,212 フクサヨウ+カガ`クテキユウハツ フクサヨウ

U: ト`クセイ+ユウガ`イサヨウ ト`クセイ

[4] S: 50,042 ト`クセイ+ユウガ`イサヨウ ト`クセイ

U: 1*2*(3+4)

[5] S: 271 1*2*(3+4)

(注) 医学用語シソーラスには、INTERFERONの下位語としてINTERFERON 1やINTERFERON 2などがあり、これらを網羅するために前方一致検索を行っている。

B 自動付加後の現在の検索式 (JICST シソーラス用語だけでよい)

U: Bカ`タカンエン

[1] S: 9,441 Bカ`タカンエン

U: インターフェロン

[2] S: 11,772 インターフェロン

U: フクサヨウ

[3] S: 102,212 フクサヨウ

U: ト`クセイ

[4] S: 50,042 ト`クセイ

U: 1*2*(3+4)

[5] S: 271 1*2*(3+4)

図3 JMEDICINE ファイルの検索例

検索テーマ：ラットにおけるカルニチン代謝

- U:ラット
[1] S: 112,846 ラット
- U: カルニチン
[2] S: 854 カルニチン
- U: タイシャ ← J I C S T シソーラス用語
[3] S: 76,063 タイシャ
- U: タイシャ キュウシュウ ハイセツ ← 医中誌副標目 (注)
[4] S: 15,173 タイシャ キュウシュウ ハイセツ
- U: 1*2*(3+4)
[5] S: 65 1*2*(3+4)
- U: ¥P A/I

#000001 JICST COPYRIGHT

CN 93D0083335

TI 糖尿病心及び虚血心における心筋細胞内カルニチン代謝の検討 細胞内分布とL - カルニチンの効果

AU 小林明 (浜松医大第3内科 (0 3 2 6 / / 3 4 6)); 外3名

JN W1458 日内会誌

VN VOL. 80, NO. 臨増 PAGE. 132 1991

CI (C) (J1) (JA) (JPN)

BD (93083335)

CC 14F0

KW 心臓病; 診断; 病態モデル; 糖尿病; 冠状動脈疾患; カルニチン; 心筋; イヌ; ラット

FT [心臓疾患; 診断; 実験的; 糖尿病; 心臓疾患 - 虚血性; Carnitine; 代謝・吸収・排せつ; 心筋; JAMAS; 犬; ラット]

(注) 「代謝・吸収・排せつ」は対応する J I C S T シソーラス用語がない

図4 医中誌副標目を使った検索例

参 考 文 献

- 1) 黒田雅子. JICST・医中誌国内医学文献ファイル(JMED)における索引語の自動付加. 第13回医療情報学連合大会発表論文
- 2) JMEDICINE ファイルの医中誌作成分へのJICST シソーラス登録語の自動付加. JOIS ニュース. No.83 p. 4-5 (1993)
- 3) 医学用語シソーラス改訂によるJMEDICINE ファイルのKW欄自動付加への影響. JOIS ニュース. No.84 p. 5 (1994)