データベース検索技法の習熟と限界 その3

情報検索技法の向上の検討

鈴木重量*

Practice of information retrieval technique and limitation of database usage (3)—Improvement in retrieval techniques—

SUZUKI Shigekazu

[著者抄録] データベース検索技法の習熟の程度によって、結果の良否のみでなく限界も明白になると指摘した第1報、および検索の限界と代行検索について考察した第2報に続く第3報で、検索技法の向上に関する諸検討を述べた。検索システムの中でエキスパート化が可能な部分について具体的に8種類の機能を挙げ、それらの機能を、実際の検索を想定して、依頼書の受注から検索結果の処理までを4段階に分け説明した。また検索へエキスパートシステムを導入した報告側の文献調査の結果に触れ、エキスパートシステム導入によって情報検索技法の向上が期待できると結んだ。

[著者付与キーワード] 情報検索, 検索技法, エクスパートシステム, 検索システム, キーワード, オンライン検索, 検索代行, AI辞書

[Author Abstract] In the preceding two papers, the following were mentioned; 1. Searchers will realize their limitations after making improvement in their retrieval techniques, 2. Searchers should ask information brokers for better results in some cases. In this paper, possible improvement in retrieval techniques is described. Eight functions of the retrieval system are available through the EXPERT system. By example of actual retrieval, the usage of these eight functions is explained in four stages from acceptance of request to delivery of results. By showing the results of a literature search obtained by retrieval using EXPERT system, the author shows how the EXPERT system could improve information retrieval techniques.

[Keyword by Author] information retrieval, retrieval techniques, EXPERT system, retrieval system, keyword, online retrieval, retrieval by information broker, AI dictionary

* (株) AI ビジネス (〒542 大阪市中央区西心斎橋2-9-33-310) Tel. 06(213)9278
(9-33-301, 2-chome, nishi-shinsaibashi, Chuo-ku, Osaka-shi, 542)

1. はじめに

「データベース検索技法の習熟と限界」は、第1報¹⁾(1985)を本誌に発表し、その中で、検索技法の習熟の程度(第一世代から第三世代以降)によって、検索結果の良否のみならず、検索の限界についても明白になることを指摘した。

データベース提供者(抄録作成と索引の付与)とそれを利用する側(検索者)とが別人であることから、必要情報を徹底的に求めることは不可能であることを、事例を列挙して説明した。

第2報²⁾(1988)は、「検索の限界と代行による検索についての考察」に言及した。

オンラインネットワークの基盤(インフラストラクチャ)が発達し、データベースからの検索が容易にできるようになったその結果として、検索代行業(サーチャー)という新しい業種が出現した。

端末を設置すれば、簡単に検索内容の一部分を表すキーワードやコードをキーインするだけで、必要な情報が検索されてくる。設備投資も少なくて済むので、代行業者が多く輩出した。

しかし、検索する行為そのものは容易であるが、代行者が検索する内容について門外漢である場合は、検索結果に正当な評価ができなくなる。

そのため、企業に所属する検索グループが行う検索結果に抗すべくもなかった。またしかるべく人容を擁して検索代行を行う機関にも後れを取り、その結果、当時多く出現した代行者は厳しい淘汰を受けることになった。

1報¹⁾および2報²⁾で云わんとしているのは検索の恐しさを知ること、すなわち検索の限界を知ることであり、それを十分わきまえて検索に対処し、今一度研究の姿勢に戻り、検索代行の役割の重要性を再認識し検索代行の地位を向上させることである。

そこで本報においては、その願いをかなえるものとしての検索技法の向上に関する諸検討を述べたい。

2. コンピュータ検索とコンピュータ(ハードウェア、ソフトウェア)性能との関係

コンピュータ検索は、如何に精緻に検索式を組み立てて行っても、最終的な検索結果の良否(正否)の判定は、ヒトに待たねばならない。

かつて、「ペニシリンを投与して尿路感染症を治療した」報告の文献検索でヒトによる調査とコンピュータ検索との違いを対比した³⁾。

ヒトが「書きもの」を読みながら必要な情報を検索する場合、コンピュータでは不可能な「考える」機能を包含している。「考える」をさらに細分すれば、「読んでいる書きものの内容理解」と「読んだ書きものがどうかの判断」に分かれる。

コンピュータは「考える」機能が欠如した単純な「読み」「書き」の一連の行為をプログラムで行うが、検索した情報は前述のようなヒトによる最終的な判定が必要となる。

ヒトが行う「読んでいる書きものの内容の理解」は、コンピュータ(プログラム)では「比較する」に相当する。さらに「読んだものがどうかの判断」は、コンピュータが比較した結果、「イコールのものを書き出す」の一連のプログラムの処理で行う(図1)。

コンピュータは「考える」とは程遠い処理でこの仕事を代行するので、当然ヒトの調査とは全く異なる。検索する内容の「ペニシリンを投与して尿路感染症を治療した」文献を求める場合、ペニシリンと腎臓のキーワードがドキュメント上に併記されていればよいことになる。そうすると以下のような内容の文献が混入することになる。

例1:「ペニシリンを投与中に腎臓に副作用が発症した。

例2:人工腎臓を用いた腎不全患者の透折中に雑菌による感染予防のためペニシリンを投与した。

当然ヒトの場合は、このような内容は内容判断によって除去できる。一方コンピュータはキーワードの併置によって適と判断するプログラムのため、除去は不可能である。

図 1 検索の流れ図(フロー)

現行のコンピュータ検索では検索結果から以下のような修正を加えることで必要情報のみを残し、ノイズのみを除去することは可能であろうか。

例1：副作用と明記されている文献をキーワードでNOTにして除去。

例2：人工腎臓適用例を同様にNOTで除去。

例1では、副作用と臓器の病的状態を表すキーワードが類似していること、例2では感染予防と感染治療のキーワードが明確に識別された索引加工上の配慮が少ないことから、NOTによるノイズの除去は、時には必要文献までも除去することになる。単純にNOTを用いて除去できるものではないことがわかるだろう。ヒトの調査のような行為をコンピュータに期待することは、現行の性能では不可能であろう。

コンピュータの高速性を利用して、ヒトでは到底不可能な情報量を検索し、予備調査(Pre-Search)として、ヒトの判断を待つことになる。予備調査をどの程度正確にかつ効率よく行えるかが、これからのコンピュータ検索の課題となる。

3. コンピュータのソフトウェアおよびハードウェアの進歩

最近、「人工知能」「エキスパートシステム」等が、従来のプログラミングとは、ランクアップされた考えとして定着し始めている。

ジョセフソン素子などのような処理速度の高速化、論理演算回路の2進化法から、ノイロコンピュータ方式による複雑演算方式の実現化の試みなどの動きが始まっている。記憶容量においても一つのチップで1メガバイトの実用化競争も盛んである。

またソフトウェア（厳密にはソフト&ハードウェアシステム）でも前述の人工知能、エキスパートが積極的に導入されつつある。

これらの進歩は、従来のコンピュータ検索を、ヒトに近い検索までに向上させることを期待している。問題は、このようなコンピュータのソフトウェアおよびハードウェアが、いつ市場に出回る

ようになるかということであろう。

素材の探索、プログラム処理上の革命的な向上は、技術突破の格好のテーマであり、有期限的に解決できるであろう。一方、人工知能やノイロコンピュータは、その擬態的な方法は行われつつあるが、本質には迫っていない。これは大脳生理上の新事実の発見に待たねばならず、またその発見は、偶発的に起るため、タイムスケジュールが立て難い。

そこで筆者は、情報検索をプログラムから始め今年に至る20有余年にわたり携わってきた経験を基にコンピュータ利用のシステムデザインを見直し、エキスパート的手法で検索システムを開発してみようと意図した。コンピュータは、マルチジョブが可能なMX/DOSが望ましいが、MS/DOSでも十分対応できるパソコンを用いて検討した。

4. エキスパートシステムの導入

エキスパートシステムは、前述のテクノロジーブレークスルーを必要としなくとも導入は可能である。

エキスパートシステムのプログラムは、従来のプログラム進行が、一つのステップを終れば次へ進むといういわば直線型の進行(サブルーチンやテーブルルックアップも考えようによっては、直線型)に対し、ステップの要所に選択肢を複数個保持させていることと、テーブルルックアップ機能が、エンド部分がオープンされていて絶えず新しい情報(条件の拡大、データ・選択肢の追加等)が加えられ、それが直ちに利用される(学習機能)ように設定されている。この考えを情報検索のコンピュータ処理上のシステムデザインに適用してみる。

5. 検索システムの中でエキスパート化が可能な部分

従来の検索システムの中でエキスパート化をした方が、易操作性の面で、また洩れのないキーワード検索(コード検索も同様)の可能化の面で、さ

らに、検索しようとする場合の標的キーワードのキーイン操作の省略の面で、また難易度の高い内容の検索の面で、さらに適正検索を可能化する面で有利になることが多い。

5.1 社内・外データベースオープン機能

自社のシステムで運用する自社独自のデータベース(プライベートファイル)、あるいは外部のインフォメーションベンダー(Information Vender)である JOIS III, DIALOG, BRS, および ORBIT などのデータベースを利用する場合、それらの識別、パスワード参照等を予めプログラム化して自

動的にセットできるようにする。検索者は端末をオープンするだけでどのシステムでも利用可能な態勢になっている。

5.2 コマンド汎用化機能

Information Vender ごとに独自のオンライン検索システムを所有している。コマンド名がそれぞれ固有で準備されているが、機能的には全く同じであるものが多い。これらをまとめ、コマンド翻訳機能を付して、一つの共通コマンドで、各システムごとのコマンドをキーインしなくても利用できるように図る。

表 1 日英の両方のキーワードが用いられる例

sulfonic acid	徐放性製剤
surfactants	交感神経遮断薬
sustained release preparation	交感神経興奮薬
sympatholytic drug	シンターゼ
sympathomimetic drug	タキフィラキシー
synthetase	タンニン酸
tachyphylaxis	催奇形性物質
tannic acid	催奇形性物質
teratogen	催奇形性試験
teratogenic agent	テルベノイド
teratogenicity test	粘度測定法
terpenoid	テストステロン
test for viscosity	テトラサイクリン
testosterone	治療学的同等性
tetracycline	トレオニン
therapeutic equivalence	トロンビン
threonine	血栓溶解薬
thrombin	トロンボプラスチン
thrombolytic drug	甲状腺ホルモン
thromboplastin	サイロニン
thyroid hormone	甲状腺刺激ホルモン
thyronine	チロキシン
thyrotropic hormone	トコフェロール
thyroxine	トルブタミド
tocopherol	トキシプラズマ
tolbutamide	トランキライザー
toxoplasma	トランスアミナーゼ
tranquilizer	トランスフェラーゼ
transaminase	トリグリセリド
transferase	トリプシン
triglyceride	トリブタミン
trypsin	トリプトファン
tryptamine	紫外吸収スペクトル
tryptophan	紫外線吸収
ultraviolet absorption spectrum	不飽和化合物
ultraviolet absorption	
unsaturated compound	

各システムのみには特有のコマンドはそのまま残すが、その場合どのシステムに特有であることを明示する。この汎用化コマンドを用いることによって、システム固有のコマンドがどうであったかという細かい雑事にわずらわされずに検索に集中することができる。

5.3 データベース優先使用指示機能

検索する内容から主要と思われるキーワード(前出の「ペニシリンを投与して尿路感染症を治療した」報告では、ペニシリンか尿路感染症)をキーインし、共通コマンドで次々とデータベースを呼び出し、検索を行い、頻度を確かめる。最も多い頻度のデータベースから3番目位までをディスプレイ上に表示する。一応頻度の最も多いものを最優先データベースとみなす。

この機能は、検索者が予め検索内容からどのデータベースを用いればよいかがわかっている場合は使用する必要はない。なお最近の医学情報は、以前と違い、診断、検査(臨床検索)、培養、生物工学(バイオテクノロジー)へと拡大し、これに、医療機器、医用電子工学なども多く含まれている。これらのうち後者のエレクトロニクス分野の境界領域情報を求める場合は、医学領域のみならず、エレクトロニクス関係のデータベースも用いなければならない。

5.4 標的キーワードの日→英翻訳機能

キーワードは、日英両語で検索する場合が多い。AIDSのような省略語はそのままの綴り(英文字)とカタカナのエイズ(ヅ)の2通りが用いられている。疾病は、日本語データベースの場合は日本語および英語(時にはドイツ語、ラテン語)が混在しているので、日英の両方で検索する必要がある(表1)。

この場合の日英の語彙数の多少が翻訳機能の充実性を決定する事になる。全くの経験的な数であるが、医学・薬学分野では、内科、外科、小児科薬理および医薬品名(治験番号→一般名→商品名)をカバー(網羅に非ず)すればほぼ実用使用が可能である。その数は40,000～50,000タームであろう。

翻訳機能の数量は、医学・薬学領域では以下の

ように対象が拡大すればそれに伴いタームや語彙も増加する。

- (1) 40,000～50,000：一応出発時の語彙数。
- (2) 60,000～100,000：各種医学辞書(例 ステッドマン、南山堂の医学辞書など)から追加する。
- (3) 110,000～130,000：SNOMED, WHO Disease Code などの疾病分類コード類、市販データベースが準備しているシソーラス類(MEDLINEのMeSH, Excerpta MedicaのMALIMETなど)。
- (4) 130,000以上：細菌名 各種学会が使用している学術用語(慣用語)などを追加する。

我々の経験では、60,000～100,000のタームが有ればほぼカバーされているようである。

5.5 検索用キーワードの拡大機能

検索用のキーワードは1タームで済むことはない。そのキーワードを中心に上位語、下位語、同意(義)語などに展開される。

ペニシリン一つを取ってみても、ペニシリンを上位語とすると下位語(同義語に分類している場合もある)は多くある(表2)。

なお市販データベースには、それぞれ特有のシ

表2 ペニシリンと下位語(同義語)

Penicilein	ペニシリン
Ampicillin	アンピシリン
Amoxicillin	アモキシシリン
Aspoxicillin	アスポキシシリン
Benzylpenicillin	ベンジルペニシリン
Carindacillin	カリングシリン
Carbenicillin	カルベニシリン
Cyclacillin	シクラシリン
Fluchloxacillin	フルクロキサシリン
Lenampicillin	レナンピシリン
Piperacillin	ピペラシリン
Sulbenicillin	スルベニシリン
Ticarcillin	チカルシリン
⋮	⋮

ソーラスを有している。

MEDLINE	: MeSH
Excerpta Medica	: MALIMET
	: EMTREE
国内医学文献ファイル	: JICST語学技術用語 シソーラス

しかしこれらのシソーラスは、それぞれのデータベースに合わせて設定しているので、絶対的なシソーラスの機能を備えているとは限らない。

特に表2のペニシリンや前報²⁾の図4で示した胃癌の hedge 検索例でわかるように、同義語的なタームへの網羅的な配慮が欠けている。MeSH や、EMTREE では階層コードが付与されているので、一応上位のコード検索でこれらの hedge のある程度はカバーされているようである。

しかし以下のように、索引の付与（換言すれば検索に用いる索引）が判然としない。

消化器癌 (Digestive System Neoplasms)

└→ 胃腸癌 (Gastrointestinal Neoplasms)

: C 4,588, 274,476

└→ 胃癌 (Stomach Neoplasms)

: C 4,588, 274,476,767

胃の疾病 (Stomach Diseases) : C 6,405,748

└→ 胃癌 (Stomach Neoplasms)

: C 6,405,748,789

(以上 MeSH より)

上位の概念は消化器癌と消化器疾病であり明らかにその違いは識別できる。しかし、具体的な胃癌ではその事象は全く同じにも拘らず、異なった分類(コード体系)に属している。こうなると検索も両体系をカバーするように配慮する必要がある。

しかし一方、キーワードは同じであるから、コードを無視してキーワードを検索に用いれば、そのような配慮は不要になる。上位概念的な検索も、NeoPLASMS (前方一致) のみで十分カバーできる。これらは手検索用に準備された体系的分類と、どこからでも検索できるコンピュータ索引が混然として同居しているせいであろうか。

今後は、データベースによって異なる上位語や下位語の由来にとらわれているよりも、同義語(同

位語) リストの充実が、コンピュータ検索では重要になるかもしれない。

検索用キーワードの拡大機能も、これまでの数千回以上のテストの結果、コードよりもキーワード優先、上位語、下位語の関係よりも同義語(同位語)優先を採るように考えている。

本機能は、検索内容から一つないし二つ程度の検索用標的キーワード(前述のペニシリンの例を参照)のみのキーインであっても、直ちに多くの同義語(同意語)または hedge に展開してくれるので、網羅的検索は可能になる。

一つのキーワードから多くの同意語などに展開するには、その分野の知識(専門知識)と数多くの検索を積み重ねる必要がある(少なくとも数千から1万件程度の検索)。

まさにエキスパートでなければできない検索であろう。ただし、hedge の単なる増加は、オンラインサーチの所要時間を多く必要とするので、やみくもな増加は考えものである。

次に、hedge と検索効率を考えてみる。前報²⁾の図4で胃癌に関する膨大な hedge 検索例を示した。しかしその数が膨大ゆえに無制限に hedge を増やし、網羅検索を徹底的に追求するのも効率上は賢策とは思えない。

図2は、癌の同義語に類する頻度をプロットしたものである。CANCER を含め14種のタームの頻度と累積を行ったが、CANCER および CANCERS で264,900の頻度で全体の87%を占めている。これに Tumour(s) を合わせるだけで96%を占めることになる。このように徹底的に hedge を追求することは、時には効率的でない。

5. 6 拡大キーワードの自動キーイン機能

ペニシリンの同義語展開で示したように多数のキーワードがディスプレイ上に表れる(表2)。

エキスパートシステムは、外部の検索システム(JOIS III, DIALOG など)とは別の独立したシステムで運用されるので、展開した同義語類は、逐一別途にキーインして再入力することになる。キーイン動作はタイプライターに不慣れな我々には

Set	Items	Description
S1	264708	CANCER
S2	7999	CANCERS
S3	264900	1+2
S4	105837	TUMOR
S5	283632	3+4
S6	90304	TUMORS
S7	292167	5+6
S8	69148	CARCINOMA
S9	294287	7+8
S10	10763	CARCINOMAS
S11	294446	9+10
S12	6596	ADENOMA
S13	296009	11+12
S14	3084	ADENOMAS
S15	296152	13+14
S16	9071	ADENOCARCINOMA
S17	296333	15+16
S18	2237	ADENOCARCINOMAS
S19	296344	17+18
S20	36316	NEOPLASM
S21	302873	19+20
S22	24806	NEOPLASMS
S23	303494	21+22
S24	23554	METASTASIS
S25	303604	23+24
S26	12464	METASTATIC
S27	303914	25+26

図 2 hedgeの増加と頻度の増加
(EMBASE 82-90年)

大変な作業になる。その上多くの同義語をキーインするとなると長時間を要し、オンライン使用経費も嵩む。

そのため外部のシステムを利用してもキーボード上の「改行キー」あるいは「実行キー」をキーインするだけで、逐次キーインしなくとも検索できるように図る必要がある。

キーイン作業の簡略化は、我々日本人とりわけ40歳台以上の男性にとっては、極めて有用な配慮である。また検索代行や、企業内で検索を専業とする者たちのようにキーボード操作に慣れている場合はよいが、研究者や医師が自分の必要とする情報を直接キーインして検索（理想とする検索のあり方）する場合は、この機能は便利な操作になることだろう。

5.7 検索用キーワードの学習機能(AI辞書へのupdate)

5.5で記述した「検索用キーワードの拡大機能」へのニューターム(New Term)のupdateを行う機能である。キーワードの拡大機能は、プロ

グラムのステップ上で説明すれば、検索の内容からまず選定したキーワードをキーインした時、そのキーワードの有無を別のディスクファイルに記録した中から走査する。

例えば前出のペニシリンの場合であれば、ペニシリンをまず走査して見つけ、ペニシリンに付属する同義語をディスプレイ上に展開する。もしこの場合ペニシリンがタームとしてディスク上に記録されていなかったら、表2で示すような同義語の展開はできないことになる。

ここでヒトによるエキスパートシステムへの教育(システムの学習)が行われる。ヒトが各種の参考書からペニシリン誘導体で実際に使われている薬物名(一般名や商品名)を調査し、AI辞書にインプットする。この動作によってキーワード拡大機能は新たな学習を受けたことになり、その後はペニシリン検索を行ってもキーワードの拡大が行えることになる。

5.4で前出の、標的キーワードの日⇄英翻訳機能もこの学習によって語彙が増加したことになる。語彙数は40,000~130,000以上と推定している

SSNO	件数	検索式
[1]	3,695	ウツビヨウ+ウツジヨウタイ+ヨクウツ&
[2]	309	テイネン+テイネンタイシヨク&+タイシヨク
[3]	6	1 * 2
[4]	42	モエツキシヨウコウグン+BURNOUT&
[5]	6	1 * 4
[6]	16,437	ロウドウシヤ+シヨクバ&+サギヨウカンキヨウ+セイカツカンキヨウ +シヨクシユベツ&
[7]	142	1 * 6
[8]	0	TI=&ニオロシ&
[9]	15,725	セイカツ+セイカツカンキヨウ+ニンゲンカンケイ+ロウドウセイリ+ストレス +カイホウカン+カイホウ+シヨクギヨウ+カイシヤ
[10]	345	1 * 9
[11]	0	TI=&ニオロシ&
[12]	12	3 + 5
[13]	3,695	1
[14]	1	TI=&ニオロシ&+&RETIREMENT&

図 3 荷おろしうつ病の検索式
(国内医学文献ファイルを使用)

がその一つ一つはすべてこれらの過程によって蓄積される。その経費と絶え間ない蓄積の努力は無視できない程大きい、その結果が優れたエキスパートシステムとなって成果をもたらしてくれる。

5. 4の項で語彙数が40,000~50,000で一応できると記述したのは、利用する側の帰属する分野によってかなり違うと考えられることによる。

医学・薬学分野の領域でも、医学側に偏向する医師の場合と薬学側になる製薬会社とは、別途の語彙の蓄積が40,000~50,000の基本的な語彙をもとに始まって行く。

医師の場合は、疾病や、解剖学用語、医用機器方面の蓄積(学習)が増大するであろう。一方製薬会社では、薬理、研究開発に関係するターム、化学物質名などの蓄積が増大する。

また医学の領域であっても、内科や外科領域に一応基本的なタームがカバーされているとは考えられるが、耳鼻咽喉科、眼科、歯科(口腔外科)、臨床検査、医用電子工学では、それぞれ特有のタームと語彙の拡がりが見られるであろう。

即ち、エキスパートシステムの主要な機能であるAI辞書部分は、共通の目的を持ったグループごとにそれぞれ独自に学習を行わなければ十分な効果は発揮できない。

40,000~50,000に次いで60,000~100,000, 110,000~130,000そして130,000以上と増加するに従って、カバーする領域も拡大する。しかしそのグループ単位ごとの特有のキーワードまでカバーできたかどうかはわからない。したがって運用者が自ら教育(インプット)し、学習させる努力を怠らないことが肝要である。

5. 8 適正検索例の蓄積と参照および学習(適正検索例のupdate)機能

前報^{1), 2)}でしばしば引き合いに出したAIDSのように、疾病名が統一されなかった頃の検索は、容易ではなかった。また検索者の知識的な背景の相違によって、出力する回答にもかなり差が認められた。科学技術情報を検索する場合は、このような統一された名称になる以前の検索が多い。

この場合検索する側が、検索した結果出力した内容が中しているのかどうかを判断するのは容易ではない。この場合複数の検索者が協議し、検索結果をその分野の専門家に見せ、専門家の見解も盛り込んだ検索式を、一応確立しておく必要がある。そしてそのような検索式を適正検索例ファイルに蓄積(学習)すると、次の検索は追試するだけで済む。

検索式の設定は、容易に考えると、オンライン検索の下では手軽に行える。しかし、果たして検索した内容が、医師や研究者の必要とする内容であったかどうかを考えると、深刻にならざるを得ない。

かつて検索者達の管理運用を行っていた頃、私は、検索スタッフ達に「たかが検索、されど検索」という言葉を口ずさんでいた。

適正検索例の成立とその検索の蓄積を例示する。

「荷おろしうつ病」という病気が最近紹介された(日本経済新聞 平成2年4月21日付)。

定年退職した人や、役職からはずされたサラリーマンらが重責を果たしホッとした時や、目標を失い、心にポッカリ穴が開いた時に陥るケースと記されていた。

いまこの内容を文献検索で求めてみた(図3)。データベースからは、荷おろしうつ病そのものがタイトル中に記載された文献は検索されなかったが、それに類する文献は検索された。たかが検索といった考えでは、この時点で検索は終わったかもしれない。

しかし、うつ病と停年、停年退職の組み合わせで6件の文献が検索されている。その分野の専門知識のない検索代行者でもていねいに検索すればこの時点までの検索は可能であろう。しかし、これとよく似た疾病で、「燃えつき症候群」があり、さらに職業病として作業環境や、職場の配置替え、転勤などによる生活環境の変化などによるうつ病のあることまで配慮し、検索できるか否かは疑問であろう。

これらは精神(心)神経医学のうつ病の起因、あるいは誘起まで考えを及ばせたならば、検索が可

図 4 検索のエキスパートシステム想定図

能となる。しかし、まだ図3に例示の検索例は適正検索例に一応蓄積はするが、まだ学習行為は継続する。

まだ「荷おろしうつ病」をタイトルに明記した文献が検索されてないからである。

今後も検索を追試し、的中文献が検索されたなら、その文献の索引代与をたどり、それらを基に再び検索を行い、その回答が満足すべきものであったらここで初めて正式に登録する。

(1). 適正検索例ファイルの他の用途

本ファイルはSDIのプロファイルにも転用が可能である。従来のSDIプロファイルはupdateからの選別のため、検索する情報量も、遡及検索と比べ少ない。そのため検索式もラフに設定され、多少のノイズの混入も容認するようになっている。

しかし、本ファイルの検索例は量の多少に関係なく厳正に設定した検索例であり、その上最新の情報に即して修正されるので、効率のよいSDIサービスが行える。

また、SDI提供者の氏名と関連づけがあるので、医師や研究者の研究動向や、テーマ変更も追跡すれば把握できる。

5.9 各種エキスパート機能のシステム図

エキスパートシステムによる検索には、従来の検索システム以外に既に紹介した諸機能が加えられる。

- ①社内外データベースオープン機能
- ②コマンド汎用化機能
- ③データベース優先使用指示機能
- ④標的キーワードの日↔英翻訳機能
- ⑤検索用キーワードの拡大機能
- ⑥拡大キーワードの自動キーイン機能
- ⑦検索用キーワードの学習機能
- ⑧適正検索例の蓄積と参照および学習機能

①～⑧のこれらの機能を、検索時を想定して段階的に説明する(図4)。

ステップ1：依頼書の受理

依頼書を受け付け、それに書かれている内容をまず読む。前出の「ペニシリンによる尿路感染症

の治療例」が依頼として書かれていると仮定する。

まず①「社内外データベースオープン機能」を利用する。そうすると2～3秒以内にどのシステムとも連結できる即応(Ready)態勢に入る。

ステップ2：検索準備

依頼書の文面から、まずペニシリン、尿路感染症を標的キーワードとみなし、キーインする。次いで④「標的キーワードの日↔英翻訳機能」を利用して以下のように日英の両タームで検索できるように準備する。

ペニシリン	Penicillin
尿路感染症	Urinary tract Infection

次いでこれらの日英標的キーワードを用いて③「データベース優先指示機能」を呼び出す。

外部からアクセスできるデータベースは数百種に達しているの、予め検索者側で複数のデータベースを指示し、そこから第1次、第2次利用と選択して行くのがよいだろう。

依頼内容から、これらの情報を最も多く収録するデータベースは、EMBASE(EM)、MEDLINE(MED)、およびJ-MED(国内医学文献ファイル)の3種であることが、医学・薬学領域のデータベースの比較検討の研究の結果わかっている^{4,5,6,7,8)}。

一応これら3種のデータベースの頻度は以下のようなになる。

(1989～)	EM	MED	J-MED
ペニシリン	:		51
Penicillin	: 2,036	1,392	
尿路感染症	:		1,420
Urinary tract Infection(s)	: 397	1,078	

③の機能を利用した結果、国内文献は、J-MED、外国では一応MEDを選んだことにする。

ステップ3：検索段階

検索に先立ち、この依頼内容が先行検索例に記

録されているかどうかを参照する。

標的キーワードに用いたキーワードが併置された先行検索例の有無を確認するために、⑧「適正検索例の蓄積と参照および学習機能(ファイル)」を利用して走査する。走査の結果、この依頼は先行例がなかったので、検索を行う。標的キーワードのペニシリン、尿路感染症を⑤「検索用キーワードの拡大機能」を用いて、同義語展開を行う。

ペニシリンについては表2にあるが、一方尿路感染症は、拡大すると腎盂炎：Pyelitis, 腎盂腎炎：Pyelonephritis, 腎炎：Nephritis, 膀胱炎：Cystitis…などがあるが、MEDでは, Urinary tract Infection(s)で殆んど網羅されているので、拡大せずに標的キーワードのみで検索式を設定する。

ペニシリンの表2の集合体と Urinary tract Infection(s)を AND で結び⑥「拡大キーワードの自動キーイン機能」を利用して、MED の検索を行う。

検索結果の回答出力をディスプレイ上から査読し、的中文献の可否を判定する。大体 10 文献程度を査読すれば、的中文献の数から検索が正鵠を射ているかどうか分かる。当然出力した回答の中には、前出の人工腎臓の透析や、腎毒性などの副作用に関する文献も混入している。ただしこれらの文献が単純にノイズとみなされ難い場合もあるので、人工腎臓や、腎毒性を除去用のキーワードとして用いることができない場合がある。

尿路感染症の治療報告例には、多数症例の報告があり、その中には、的中症例もあれば、人工透析例の症例もある。また治療中に発症した副作用もある。これらの症例のキーワードが索引として単純に併記されているので、単純に NOT 処理するとそれと一緒に的中症例も除去されてしまう。この問題は、今後、索引の付与の方法などで検討する課題となるだろう。

抄録は、多数症例の報告を文脈を追って報知できる程の長文ではない。単純なキーワードの羅列であっても、同じタームは以下のように 1 タームに集約されてしまう。

1. 尿路感染症／腎盂腎炎 ペニシリン／アモキ

シリン

2. 尿路感染症／膀胱炎 ペニシリン／クロキサシリン 副作用／腎糸球体炎

3. 腎不全 人工透析, 人工腎臓, ペニシリン, 感染予防

など普通少なくとも数症例, 多いものは 50 症例以上に達する症例報告が 1 文献に記載されている。

これらの報告は索引としては、症例ごとに区別されず、論文単位に以下のように集約されてしまう。

尿路感染症, 腎盂腎炎, 膀胱炎, 腎不全, ペニシリン, アモキシリン, クロキサシリン, 副作用, 腎糸球体炎, 血液透析, 人工腎臓, 予防効果

またデータベースによってはキーワード付与数の制限もあり、すべてのキーワードが索引化されるとは限らない。市販のデータベースを利用する限り索引加工上の制限があり、今後この問題について検討される時が来るであろう。

ステップ4：検索結果処理

最初に用いた標的キーワードはいずれも AI 辞書の中での同義語の拡大で済んだので、⑦「検索用キーワードの学習機能」にインプット(学習)する必要はなかった。

検索式は回答の出力を査読し判定の結果、一応⑧「適正検索例の蓄積と参照および学習機能」へインプットすることになる。

以上が、エキスパートシステムを用いた検索の想定である。

6. 検索へエキスパートシステムを導入した報告例の調査(文献調査)

5. 1～5. 8 の各機能(5. 9 の①～⑧)に言及した文献を調査した。ただしエキスパートの報告に関する検索は、どのキーワードを標的にするかの判定が困難であり、網羅的な検索ではない。

また検索の途次、情報検索には関係しない、人工知能全般の参考書、ファジー理論、ノイロコンピュータ、バイオ記憶素子の報告、PROLOG, Lisp のみの報告は除去した。こうして検索し、オリジナルを収集したのは 200 余文献であった。

これらを査読した結果の要約を列挙する(調査

対象期間 1985～1989.8)。

1. ①～⑧のすべてを包含したシステムの報告例は認められなかった。今回査読した文献の中では、田中の報告⁹⁾が最も充実しており、①②⑤⑥に関しては、実施済みであり、我々のシステム設計に最も多くの示唆を与えてくれた。今回紹介のシステムとは趣を異にしたクローズドシステムではあるが、検索、知識データベース、検索対象情報を一括パッケージ化したシステムで「超高速文献検索システム」があり、注目したい^{10,11)}。

高雄ら¹²⁾が「知的図書館 ILIAS」を報告している。閲覧者が、自分が読みたい内容が漠然としている場合、ILIAS を用いて段々と内容が具体化し、そして特定化をし、その時点で特定化した内容を備えた本にたどりつくシステムで、目的は異なるが、参考にしたい。

伊藤ら¹³⁾が同様な考え方で「知識型情報検索システム NIRS」を報告している。

エキスパートシステムではないが、「英日機械翻訳システム DUET EJ」の紹介¹⁴⁾があった。

この中での分野別 EJ 辞書は、④と関係がある。

岡村ら¹⁵⁾、生駒ら¹⁶⁾は、化学構造検索用 DB のエキスパート手法によるコマンド変換システムを報告している。

なお研究報告やシステム開発の報告ではないが、システムの考察を行ったものとしては、近藤ら¹⁷⁾、小口ら¹⁸⁾はが知識ベースの管理運用に有用な、知識とデータの相違点に関する基本的考え方¹⁷⁾、および知識ベースとデータベースの融合方式¹⁸⁾を述べている。

総説や、動向に関する報告は数多くあるが、その中では、斉藤¹⁹⁾の人工知能とドクメンテーションの接点について、大塚の図書館とエキスパートシステム²⁰⁾、ドクメンテーションへの応用化の可能性²¹⁾、中村²²⁾のハイテク時代のデータベースなどは、参照する価値がある。

此処に紹介した文献は既に 10 か月を経過したものばかりで、最新の文献までを査読して別途に紹介したい。

7. エキスパートシステムの開発

これらの文献を参考にした結果、現時点で可能なエキスパート機能(①～⑧)をシステム化した開発は、既に④⑤の一部を残し、一応終了している²³⁾。

④標的キーワードの日⇄英翻訳機能および⑤検索用キーワードの拡大機能は、既に蓄積量 40,000～50,000 を上回り、60,000～100,000 の域に達している。これ以上のタームあるいは語彙が必要かどうかを検討中である。

また対応機種もパソコンレベルによる stand alone 使用、パソコン LAN、および大型機種によるオンライン使用の拡大対応も検討している。

あらゆるメーカーの機種にも対応するために、C 言語で書くつもりである。

8. エキスパートシステムの利点

情報検索技法の向上と検討を、エキスパートシステムの実現とその利用に置き換えたのは、理由がある。

「ヒトによる調査」は情報の増大と調査対象が多様化する昨今の研究では、従来のように体系づけられた目録や、手検索用二次資料からの調査では、多数を動員してこれに当たるが、さもなくば、少人数で長期間の調査を行うかのいずれかでなければ不可能である。前者はコストの問題で後者はタイミングを失するために良策とは考え難い。

最近、特許情報や、文献情報を単なる閲覧のために収集することから一步踏みこみ、動向把握、企画、市場調査に用いる試みが行われている^{24,25,26,27)}。また新製品の開発の有力な手掛けとして用い、成功した例もある²⁸⁾。このような利用の仕方においても必要とする情報の収集期間が律速となる。

一方において、情報の収集はヒトによる調査は無論、従来方法のコンピュータ検索でさえ欲する情報が意のままに集められなくなった。この問題に対し、エキスパートは、効果的な働きをするで

あろう。

検索需要はますます増大の一途をたどっている。それに呼応し、企業内での検索代行業も少数者から漸増し、グループ化し、組織化されるようになった。ここに配属されるスタッフは女子社員や、高齢者の割合が従来から存在する課に比して多い。依頼受付や、検索操作は女性が、一方、難解な依頼や、その場合の検索結果の回答の内容を確認するのはベテラン男性が当たる。

そして女子社員は、検索を通じて徐々にベテランに、同様に男子社員は、検索という新しい職種に納得し、生甲斐を感じ、やがて親切な受付の対応と検索操作にも馴染んで行く。

早いもの（女子社員に多い）で半年から1年、遅いもの（男子社員に多い。殊に専門分野の蘊蓄がある者）で1年～2年で、いわゆるベテラン（エキスパート）検索者へと生れかわる。

この間に数千件から1万件程度の検索を経験する。困ったことにせつかくベテランに育った頃、女子社員は退社し、一方男子社員は、円満になった人柄と技能を買われ転属することが多い。いわゆる入れ替えの激しい職場となり、絶えず、未熟者が一定の割合で存在することになる。このような場合、エキスパートシステムは、新しく配属されたスタッフが検索してもベテランに伍する結果をもたらすことになる。

さらに付言するならば、エキスパートシステムで用いる日英辞書（④標的キーワードの日英↔翻訳機能）や、⑤検索キーワードの拡大機能は、日↔英の自動（機械）翻訳の有力な専門分野別辞書と成り得る。

学習させれば、学習させる程、検索において、

そして翻訳において効果が上がることになる。

9. お わ り に

データベース検索技法の習熟と限界に対し、筆者自身が応える情報検索技法の向上と検討は、検索のエキスパート化をもって回答とした。

ただし、本報は、備えるべき機能を中心に展開し、なんら、実験側を開陳していない。5.9の項のうち、④と⑤に対し、これでよしといったピリオドを未だ着けられずにいるためである。一応100,000～130,000程度のタームあるいは語彙に達した時（8月中）に区切りをつける予定である。エキスパートシステムで複雑かつ難解な他の機能は終えているのに対し、最も簡単で、ただコツコツたゆまず入力すればよい部分が、最も遅れてしまった。改めてデータベースの重要性を教えてくれた。

検索のエキスパートは、④と⑤を分野別ごとに確保すれば、医学・薬学領域以外でも、利用できる。

今回のエキスパートは、検索に重きを置いたが今後は次の2点を検討するつもりでいる。

（1）検索した内容の中で標的キーワードの位置、前後のタームとの関係などから、ヒトが調査し、必要と認める行為に一步でも近づいた内容判定のためのAIの導入。

（2）英文データベースを検索し、その回答を自動翻訳の機能を介して、日本文で出力する。

エキスパート化に着手したため、また新たな問題と取り組むことになったが、今後もその経過や成果を公表し、諸兄姉から、御意見の賜わることを期待してやまない。

参 照 文 献

- 1) 鈴木重量. データベース検索技法の習熟と限界. 情報管理. 27(10)897-904(1985)
- 2) 鈴木重量. データベース検索技法の習熟と限界 その2—検索の限界と代行による検索についての考察—. 情報管理. 31(2)117-126(1988)
- 3) 鈴木重量. 電子計算機による情報検索:TIS-COM (Takeda Information System by Computer) の紹介. 生活衛生. 18(2)43-53(1974)

- 4) 鈴木重量, 平林和夫, 酒井満, 二次資料の効果的利用法, *Excerpta Medica*, 薬学図書館 24(314)160-190(1979)
- 5) 鈴木重量, 平石繁雄, 平林和夫, 寺浦一正, 日比野啓, 武田薬品中央研究所における文献検索センター活動 *Excerpta Medica* および Ringdoc の紹介ならびにそれらの検索利用, *武田研究所報*, 35(314)264-322(1976)
- 6) 鈴木重量, 平林和夫, 平石繁雄, 角田喜治, エクセプタ・メディカ (*Excerpta Medica*), メドラス (MEDLARS) リングドック (Ringdoc) の検索比較, *医学図書館*, 23(4)204-218(1976)
- 7) 北口正, 野尻務, 鈴木重量, 吹田忠純, 川名敏夫, 市販後医薬品監視における文献データベースの評価, *情報管理*, 26(3)189-198(1983)
- 8) 北口正, 野尻務, 鈴木重量, 吹田忠純, 川名敏夫, JMEDICINE (JMED) および JAPICDOC (JD) を用いた安全性情報の雑誌別出現頻度リストの作成と考察, *情報管理*, 26(9)736-744(1983)
- 9) 田中征義ほか, 索引・検索支援のための用語管理システムの作成, 第25回情報科学技術研究集会発表論文集, 155-162(1988)
- 10) 超高速文献検索システム(日立), *日刊工業新聞*, 日本経済新聞 1989.7月26日号
- 11) 超高速文献検索システム, 日立 News Release 1(6)89-108(1989)
- 12) 高雄保嘉ほか(CSK 総研), 知的図書館システム ILLIAS の開発, *情報処理学会第34回 (S.62年前期) 全国大会 6J-1*, 1425-1426
- 13) 伊藤秀昭ほか, 知識型情報検索システム NIRS におけるデータベース結合の検討, *情報処理学会 (S.63年前期)* 38-11419-1420
- 14) シャープ英日機械翻訳システム パンフレット
- 15) 岡野博之ほか, 文献データベース 知識検索システム, 第26回情報科学技術研究集会発表論文集, 43-49
- 16) 生駒哲一ほかエキスパートによる JOIS-STN コマンド変換システム, 第26回情報科学技術研究集会発表論文集, 51-60
- 17) 近藤秀文ほか知識ベース管理システム—知識とデータの相違点に関する基本的考え方—, *情報処理学会第31回 (S.60年後期) 全国大会 6M-2* 10001-1002
- 18) 小口琢夫ほか知識ベース管理システム—知識ベースとデータベースの融合方式—, *情報処理学会第31回 (S.60年後期) 全国大会 6M-3* 1003-1004
- 19) 斉藤孝, 人工知能 (AI) とドクメンテーションの接点をさぐる, *ドクメン研究* 36(1)3-8(1986)
- 20) 大塚敏夫, 図書館とエキスパートシステム 現代の図書館, 23(2)93-97(1985)
- 21) 大塚敏夫, ドクメンテーションへの応用の可能性, *ドクメン研究* 36(2)63-67(1986)
- 22) 中村幸雄, ハイテク時代のデータベース, *ドクメン研究* 36(6)261-272(1986)
- 23) 鈴木重量, 知識データベース (医学・薬学分野) を用いた文献検索のエキスパートシステムの開発, 第27回情報科学技術研究集会で発表予定
- 24) 角田喜治, 鈴木重量, 平石繁雄, 野本吉広, 野口俊作, 医薬品に関する情報科学的研究 II—ALARM による情報科学的分析 (Ringdoc および CPI を用いての事例報告—, *情報管理*, 21(6)443-461(1978)
- 25) 角田喜治, 鈴木重量, 平石繁雄, 野口俊作, 同上 III—一時系列展開リストを用いての情報科学的分析—, *情報管理*, 21(8)592-606(1978)
- 26) 角田喜治, 鈴木重量, 開道巧, 野口俊作, 同上 IV—研究ライフサイクルに関する情報科学的研究—, *情報管理*, 21(9)653-662(1978)
- 27) 角田喜治, 鈴木重量, 開道巧, 平石繁雄, 野口俊作, *情報管理*, 21(9)663-680(1978)
- 28) 鈴木重量, 調査活動に関する一考察, *薬学図書館* 34(2)103-111(1989)