

2) 見出し語（英語）の数

以上の結果、第三版に収録した英語見出し語の数は、67330語である。

第二版には、79302語が収録されていたが、約27700語を削除し、約15800語を新たに採用した。

3) シソーラスの考え方

複数の用語が、同義語、上位語、下位語などの関係にあることがわかるようにしたものをシソーラスという。本来、用語には上位、下位、同義の関係があるから、これらが明らかになっていることが望ましい。既に述べたMeSHは、シソーラスである。しかし、日本では、シソーラスの作成は早い段階からその必要性が叫ばれながら、今日まで医学用語の完全なシソーラスは存在していない。医学中央雑誌は、日本語版シソーラスを第6版まで編纂してきたが、その改訂頻度が頻繁でないことが惜まれる。

日本医学会医学用語管理委員会も、医学用語辞典をシソーラス化することが必要であることを十分認識しているが、この作業は、膨大な労力を必要とする。

このため、今回の改訂では、将来のシソーラスの編纂が少しでも行いやすくなるように、その準備段階とでもいふべきものを辞典の中にも含ませることとした。まず英語圏のシソーラスであるMeSH採用語については、その階層構造を示すTree No 最初の2桁を本辞典にも収録して、これを使えば少なくともMeSH収録の日本語については、MeSHでの所属カテゴリーがわかるようにした。また、各分科会の協力を得て、同義語も判定し、英語見出し語間の同義語関係を明らかにし、それに対応する日本語も同義語であることがわかるようにした。しかし、今回は上位語、下位語の関係まで示すことはできなかった。

将来、これらを含ませることができれば、この辞典全体がシソーラスとみなすことができるようになる。今回の改訂ではその第1歩を踏み出したというべきであろう。

4) 日本医学会分科会の用語集及び他の医学用語集との関係

本辞典は、日本医学会の既に述べた各分科会の編集している用語集を可能な限り取り入れるこ

とを方針とし、各分科会に専門用語のチェックを依頼した。ほとんどすべての分科会の協力を得られたことを感謝する。また、電子化された用語集が存在している学会はその用語集も参考にしたが、その数はそれほど多くはなかった。今後各分科会の用語集が電子化されることを期待したい。

将来の本辞典と各分科会の用語集との関係は、学会間で共通に使われる用語については同一の用語が日本医学会の辞典の分科会の用語集に収録され、各分科会の用語集は、それに分科会の中でのみ使われる専門用語を追加したものとなることが望ましい。しかし、現段階では学会間で意見が一致していない共通用語もあり、今後時間をかけて調整を進め、上記の理想に近づけていく必要がある。

一方で、医学用語は、医学界という閉じた世界で使われるだけではない。現在では、新聞にも、また医療行政にも医学用語は頻繁に現れる。従って、既に世の中に普及し定着している医学用語との整合性を保つことも必要である。この観点から以下の用語集については可能な限り参照し、一致できるものは一致させたが、現時点ですべての用語（日本語）を一致させることはできなかった。今後の課題としていく必要がある。

- 1) ICD10 対応電子カルテ用標準病名マスター 医療情報システム開発センター編集
- 2) ICD10 国際疾病分類第10版 厚生労働省 ICD室編
- 3) 医学用語シソーラス NPO 医学中央雑誌刊行会発行
- 4) 平成17年度版 医師国家資格試験出題基準株式会社 まほろば 発行
- 5) ICH 国際医薬用語集日本語版 (MedDRA/J) (Medical Dictionary for Regulatory Activities) 日本公定書協会 JMO 事業部
- 6) 文部科学省・日本医学会共編 学術用語集医学編 日本学術振興会発刊 丸善販売

(4) 用語の表記法に関する方針

1) 英語の表記法の揺れ

用語の表記法は、必ずしも簡単ではなく、英語においても、表2に示すように、ハイフンの使い方や、スペルのゆれ、大文字やイタリックの使い方など表記に揺れがある。本辞典では、MeSHなどを参考にしつつ、できるだけ英語圏で標準的と思われる表記法の見出し語とした。

表2 英語における複数の記述方法の例

英語	表記法の例	本辞典で採用した表記法
アイソトープ	I-131, I ¹³¹ , 131I	131I
ギリシャ文字	α, alpha	alpha
ハイフンの有無	acid base, acid-base	慣習による
一語か二語か	aftercare, after care	慣習による
人名のあとのアポストロフィー	Basedow's disease, Basedow disease	原則としてアポストロフィーsをつける
大文字の使い方	Southern blotting	固有名詞は大文字
スペルの揺れ	leukocyte と leucocyte	広く使われている米国式スペル
イタリックの使い方	<i>Staphylococcus aureus</i>	動植物、細菌、ウイルスなどの種はイタリック

これらの表記法の「揺れ」の中で、この辞典で採用したものは表2の一番右の欄に見るとおりであるが、人名のあとのアポストロフィーSの使い方については付記しておきたい。

アポストロフィーSは、英語圏の医学論文をみても必ずしも統一されていない。しかし、統計的に見ると、アポストロフィーSをつけた用例が圧倒的に多い。しかし、MeSHの基幹語 (Main Heading)にはアポストロフィーSはついていなく同義語 (Entry Term) の項 についたものが記載してある。ただ、これは、MeSHが索引用に作られたために、入力の手軽さからこうなっているようである。

アポストロフィーSが圧倒的に多く使われているとは言っても、二人の名前がついた場合にはアポストロフィーSはつけない。ただし、人名にあたる人が遠い過去の人であるような場合には、それが名詞の一部のようになってつかない場合もあるようである。これらの慣習は、英語圏に属さないわれわれには判断は非常に難しく、用例を参照しつつ定めた。

例

Stevens-Johnson syndrome

二人の名前であるので 's はない

Argyll Robertson's sign

一人の名前であるので 's をつける

Brown-Sequard's paralysis

間に「-」があるが一人の名前ある。

2) 日本語の表記法の揺れ

英語に比べ、日本語の表記法は、かなや漢字があるためにはるかに複雑である。これまでは、学会の考え方や慣習などによって表記法は必ずしも一定していないものもあった。しかし、標準的表記法がないままに、複数の表記法が無秩序に存在する状態は好ましくなく、意見を統一する努力は必要である。

本辞典では、日本医学会分科会の意見を集約して、できる限り「標準的な表記法」を示す努力をしたが、現段階では意見の相違もある。意見が相違した場合は、その違いを示すことによって問題点を明らかにし、将来意見交換によって「標準」が自然に定まることを期待した。

表3 日本語における表記法の揺れ

表記法が異なる理由	異なった表記法の例	本辞典で採用した表記法
カタカナの使用	タンパク質、蛋白質	物質の名称、動植物の名称
漢字、かな	うっ血、鬱血	字画が多くて意味が理解できる場合はひらかな うっ血
漢字の異字体	靱帯、鞅帯	正字を選定 靱帯
漢字の略字体	頸部、頸部	略字体がある場合は略字体 頸部
同音漢字による置き換え	交差、交叉	一般社会で用いられている用語、ただし医学的に意味が重要な場合を除く 神経の交叉の場合は「交叉」
漢数字	十二指腸、3次元CT	熟語となったものは漢数字
ローマ数字	第Ⅷ因子	慣用による
人名	バージャー病、ピュルガー病	慣習を重要視
異なった用語	副甲状腺、上皮小体	学会の意見を集約

第3版では、「標準的表記法」を示すために、次のような記載方法をとっている。

①推奨語及び廃語

日本語が複数ある場合には、できる限り優先的に使用すべき語を推奨語として最初に記した。推奨すべき言葉を強調する時は日本語の後に【奨】という記号を付した。また、過去に用いられたが現在は使われない用語は後に【旧】という記号を付した。

例

hyperparathyroidism [M] (C19) 副甲状腺機能亢進【奨】、上皮小体機能亢進
schizophrenia [M] (F 03) 統合失調症、精神分裂病【旧】、

②意見の異なる日本語

学会に推奨語とは異なった強い意見がある場

合には、学会の略称を語の後に付して併記した。
例

dementia 認知症【奨】、認知症（痴呆）
【精神・神経】、痴呆【旧】

③表記の「揺れ」の一覧表

日本語表記に相違あるものの一覧表及び人名表記の一覧表を作成し、辞典に収載した。これは、まだ完全なものではないが、今後常時改訂して日本の医学界で合意したものになることを期待したい。この一覧表は、最初に記載してある日本医学会のホームページに公開されており、誰でも医学用語管理委員会に意見を述べるので、ぜひご意見をいただきたい。

以下に、これらの表記法の揺れの中で、異なった用語、及び漢字とかなの使用法について本辞典のとった方針について付記する。

3) 異なった用語

① 上皮小体と副甲状腺

これまで医学界では、parathyroid organ に対して、上皮小体と副甲状腺という二つの用語が使われてきた。今回の辞典では、医学会分科会の意見を求めたが、幸い「副甲状腺」を推奨語とすることに強い異論はなかったために、副甲状腺を推奨語とした。

② 蜂巣炎と蜂窩織炎

phlegmone の日本語訳として、過去には「蜂窩織炎」が使われ、今でも一部で使われているが、本辞典では、「蜂巣炎」を推奨語としたが、まだ異論は残っている。

③ 新生物という用語

neoplasm という英語に対し、「新生物」という訳語が使われる場合があるが、日本には、古来から「腫瘍」という用語があり、「新生物」という用語は必要がない。したがって、本辞典では、病理学会、癌学会、癌治療学会などの意見を求め、neoplasm に対する日本語は、「腫瘍」を推奨語とした。

④ schizophrenia の訳語

「精神分裂病」に対し、「統合失調症」が日本精神神経学会などの努力で定着した感がある。「分裂」という言葉のつくその他の医学用語についても、この観点から見直しが行われた。精神分裂病言葉は「旧」という符号を付して残した。

⑤dementia の訳語

「痴呆」という用語を廃止し、「認知症」という用語を使用することを提案したのは、厚生労働省の有識者会議であった。しかし、日本精神神経学会及び日本神経学会は、現在でも、認知症

作られた漢字である。

3) 第三版の辞典の漢字とかなの使用について

第三版の漢字とかなの使用法については、既に述べた表記法の揺れの表を参照されたいが、この表の方針について若干のコメントを付記する。

前節(5-2)で述べた漢字の種類に基づいて、例えば「常用漢字、表外漢字以外はかな書きとする」というような原則をたてることは可能ではあるが、医学用語の場合に必ずしも適当ではない。それは、例えば、「橈骨」を「とう骨」と書けば、「頭骨」と区別できなくなるなど、「かな」を使ったために意味がわからなくなる用語が多数存在するからである。従って、ここでは次のような比較的柔軟な方針をたてた。

①表外漢字、JIS 漢字を含め、可能なものは漢字表記を優先した。特に、人体の一部を示す用語はかなで書くと意味がわからなくなる場合が多いので漢字を優先した。

②総画数の多い漢字で、「ひらかな」を使うことが慣習となっている用語は、「ひらかな」とした。例えば、「うつ病」、「せん妄」、「てんかん」などである。

ただし、下記のような用語は、「ひらかな」のみでは意味がよくわからないが、漢字の総画数があまりにも多いため、ひらかなとして辞典の上では漢字を併記してある。このような用語は将来日本の医学界が智慧を絞って別な漢字の表記法を考えるべきであろう。

carbuncle よう、癰、furuncle せつ、癤

③略字体の漢字

第2版においては、文部科学省学術審議会(学術用語分科会)の承認を得て、略字体漢字を第一選択とすることを提案してきた。この基本的な考え方は第3版でも変わらないが、実用性の観点からワープロで表現できない略字体は、正規の漢字を優先することとした。ただし、ワープロで表現できる漢字は時代と共に変化することが予想されるので、その時点での見直しは必要であろう。

この原則により、例えば、「頸」を使い「頸」は使わないこととした。しかし、痙攣の略字体はへんかつくりの組み合わせで作ることはできるが、ワープロで記載できないため、略字体は使わないこととした。

④異体字について

コンピュータでデータを取り扱う場合には、異なった字体の漢字は、まったく別な字として認識され、これを知らないと検索に支障を来す。特に、次の医学で繁用される漢字をコンピュータで処理する場合には異なった字体が存在している異なる文字コードが割り当てられていることに注意する必要がある。

本辞典では、上記の観点から、これら異体字のある漢字については、現在、国語関係研究者の間で最も広く使われている下記の漢和辞典で正字とされている字体を使うことを原則とした。(鎌田正・米山寅太郎 著 新漢語林 大修館書店)

特に注意すべき漢字として、「靱」(E8D5)と「靱」(9078)がある。第三版の辞典では「靱」を使うこととしている。また、「膺」にも異体字があり、「膺」(E454)及び「膺」(E453)で、いずれも正字とされている。表外漢字字体表では「膺」を採用しているため、一般社会では「膺」が多く使われているが、本辞典の第二版では、「膺」を採用しており、第三版でも字画数が少ない「膺」を用いることとした。

⑤同音の漢字による書きかえ

常用漢字の制定以来、同音の漢字による書き換えが一般社会で少しずつ浸透しつつある。これについては規範となる文書は存在しないが、本辞典では、社会慣習を考慮してできるだけ一般社会の漢字の使用法に合わせるようにした。(詳細は、WEB上の表「日本語表記の揺れ」参照)

しかし、次の用語は医学的な考慮から一般社会の使用法と異なる選択をした。

① 交差：一般社会では「交差」が普及しているが、本辞典では、神経が交叉するような場合には「交叉」を用い、その他の場合は「交差」とした。

② 線維：一般社会では「繊維」が使われている。しかし、医学関係者の間では、「線維」が早くから普及しているため、「線維」とした。

③ 弯曲：一般社会では、「湾曲」が使われているが、医学においては、「側弯」、「大弯」、など弯曲以外でも「弯」が多く使われる。従って、この場合に限っては、一般社会の用語と異なる「弯曲」を使用した。

以上は、本辞典で採用した方針であるが、同音の漢字による置き換えは、総画数の多い漢字を使うことが多い医学用語を将来簡素化するための有力な手段となる可能性がある。過去にも、包帯(縫帯)、奇形(畸形)、死体(屍体)、回虫(蛔虫)など、医学用語は簡素化されてきた。今後も、医学界は簡素化に努力し、新しい医学用語を世の中に提案してもいいのではないかと思う。

例えば、「疥癬」は「介癬」、「褥瘡」は「褥創」とするなどを医学界の意見が一致すれば世の中に提案していくことも可能であろう。漢字文化の国であったはずの中国では、思い切った漢字の簡素化が行われたが、その時は音が一致している漢字が多く選択された。漢字の意味を大事にすることも必要であるが、それが過度になると一般社会から受け入れられない場合もあり、先見性をもって世の中をリードしていくことも医学会の役割ではないかと思う。

4) 人名の表記法について

人名をカタカナで表現する時にどの発音に基づいて、どのように表記するかは非常に難しい問題であるが、本来の発音に近い表記法とすることは必要であるが、正確を期すことは不可能で、どの国でもその国の表記法が存在している。例えば、キリル文字のロシアの人名やウムラウトなどがつく人名は英語圏では、英語表記を考案して用いている。重要なことは、一つの国の中では表記法を一定にしておくことで、この点、政治家、芸術家、など一般社会で使われている人名はジャーナリズムが努力して標準的な書き方が確立している。

医学関係の人名については、これまでそのような標準的な表記法を定めようという努力がなかったために、混乱を招いている感があるが、今後医学関係者がその努力をするべきであろう。

今回の辞典の編纂にあたっては、その意味から人名表記法の案を一覧表として示した。(既に述べたWEBページを参照されたい。)

又、読み方以外では次の原則を用いている。

二人の名前の場合は、間に「・」を入れる。 ス
チーブンス・ジョンソン症候群
一人の名前の場合は、間にスペースを入れる
アーガイル ロバートソン徴候
一人の名前で間に「-」のある場合は「-」を入
れる ブラウン-セカール麻痺

WEB上の表はまだ不完全なものであるが、今後学会の意見を入れてすべての医学関係者に受け入れられる表記法になることを期待している。

5) 医学用語の「読み」について

医学用語の「読み」については、医学独自の読み方があり、一般社会と異なっているという批判を受けることがある。特に、「腔」は、「口腔(こうくう)」や「腹腔(ふくくう)」のように医学的には「くう」と読むが、漢和辞典には「こう」という読みが存在するのみである。

しかし、言葉は時代と共に変化するものであり、医学界からより自然な読みが提案され、それが定着していくことがあってもよく、医学界はもっと世の中に対してまともではっきりした意思表示をするべきではないかと思う。

その他にも、読みに関していくつかの医学特有の読みがあるがここでは省略する。

(6) 今後の課題

日本医学会・医学用語辞典英和第3版の編集方針について述べたが、今回辞典が出版されても、用語に関する課題がすべて解決されたわけではない。

この辞典で実現できていないものの一つに、上位語、下位語の整理がある。同義語については、各分科会の協力を得てかなり整理されているが、それぞれの語にどのような下位語があるかが明らかになれば、辞典としての価値は一挙に高まるのである。しかし、そのためには、医学的な概念(コンセプト)の整理が必要で、課題として残されている。

第4章 専門用語に関連した システムの研究

4. 1 日本語医学用語で MEDLINE を検索するシステム

4. 1. 1 はじめに

MEDLINE®は医学関連の論文を、MeSH® (Medical Subject Headings) [1]によって検索するシステムである[2]。これは各論文を MeSH という統制用語集で索引付けを行っているからこそ実現できることであるが、MeSH は英語であり、MEDLINE を検索するためには英語でキーワードを与えなくてはならない。これを英語の代わりに日本語で検索することができるならば、医療従事者に限らず多くの日本人が、母国語を用いて最新の医学文献を利用することが可能となる。これは EBM がますます重要になりつつある昨今、極めて有用なものであると考えられる[3]。英語以外の言語による MEDLINE 検索に関する研究[4, 5]は以前から行われてきた。しかしその検索結果は MeSH の英語と対になる各言語で記述された用語集の品質に左右され、用語集の単純な逐語訳が必ずしも正確な医学用語とはならないこともその要因としてあげられる。

平成 13 年度～15 年度の厚生労働省科学研究「UMLS と連携した医学用語ソーラスの作成に関する研究」では、医学中央雑誌の医学用語ソーラスを出発点として、日本語医学用語と MeSH との対応を検証するという作業を行い、結果として MeSH に準拠した日本語医学用語ソーラスを、そして MeSH を介して UMLS と連携する日本語医学用語ソーラスを得ている[6]。これを利用して、英語ではなく日本語の検索語によって MEDLINE の検索を実現した（以下これを旧システムと呼ぶ）。ここには日本語による文献検索を実現したこととは別の重要な側面がある。当時国内には、医学用語に関する用語集は多数存在したが、医学の各分野の専門領域の中だけで閉じた用語集であり、他の用語集との相互の関連性はなく、当然ながら欧米の用語集との関連性も不明であった。例外はわずかに標準病名集 (MEDIS・標準病名マスター) や MedDRA の日本語訳だけが、それぞれ ICD10 や MedDRA など欧米の標準的な統制用語集との関連を持っているだけであり、特定の専門領域にとらわれない医学全般にわたる医学用語に対して、MeSH (およびそれを介して UMLS との) 対応関係を明らかにした意味は大きい。またこの研究と関連・並行

して NLM では MeSH の各国語対応を充実させる努力を払い[7]、2004 年からは日本語版 MeSH が UMLS に加わっている。

しかしながらここで対象とした日本語医学用語は日本語の文献を索引付けするために用いられている語彙、すなわち文献検索用の語彙であった。これをもっと一般的な医学用語に拡張し、それら医学用語と MeSH および UMLS との対応関係を明らかにすることが望まれた。

本研究で対象とする医学用語は日本医学会の医学用語辞典英和版を出発点としている。この辞典は日本医学会が編纂していることからわかるように、国内の約 100 に及ぶ医学分科会が用いている医学専門用語を収集しただけではなく、同じ意味に対して異なる表記が存在する場合には医学会が音頭を取って標準的な表記を選択する、という調整作業を行った結果として得られた用語集である。その第 2 版は 2001 年に刊行されており、第 3 版が 2007 年に刊行される予定で改訂作業が続けられていた。この改訂では従来の調整作業に加えて新たに、すべての英語見出し語に対してその語彙が UMLS に存在するか否か、および MeSH に対応するか否かのチェックを行っている。これにより、個々の見出し語に UMLS および MeSH へのリンクを付すとともに、各見出し語が英語として正しいものであるかを確認している。これにより以前の医学用語辞典に含まれていた、いわゆる Japanese English やラテン語と英語が混じり合った日本特有の医学用語などを一掃するとともに、各見出し語と UMLS および MeSH との対応関係を明確にしている。本研究ではこの辞典の日本語を検索語として使い、MEDLINE 検索を実現することを目的とした（以下これを新システムと呼ぶ）。

旧システムでは、MeSH と対応する日本語だけを対象としたので文献検索には有利であるが、MeSH のみに準拠した用語集を出発点としているために MeSH 以外の UMLS に含まれている語彙との対応関係がわからなかった。また、文献検索に用いる日本語語彙は一般的な医学用語とは異なっている可能性がある。新システムでは一般的な医学用語辞典に記載されている日本語の中

で MeSH に対応するものを抽出し、さらに UMLS に記載されている統制用語集で MeSH 以外のものへの対応を抽出している点に特徴がある。

なお、MeSH2006 年版は 23,885 個の Main Headings とその同義語や関連語 (related terms) 約 151,140 語とを合わせた 175,025 個の英語医学用語から成る。ひとつの Main Heading に対して平均 6 語の同義語 (または関連語) が存在し、これは実際にはひとつの Main Heading に対して最小 0 個から最大 172 個に至る同義語 (または関連語) が存在する。因みにこの最大 173 個の同義語を有する語彙は、Acetylcysteine であり、同義語を持たない語彙は 2,582 個に及ぶ。

UMLS® (Unified Medical Language System®) は 1986 年に始まり、現在も継続している NLM (米国 National Library of Medicine) の project である [8, 9]。その目標は、生物医学に関連する様々な文書から機械が読める情報を検索・統合できるようにすることである。しかし様々な医学関連の用語集ごとに異なる語彙と定義の多様性があり、同時にそれらの関連性が明記されていないことは大きな障壁であった。そこで UMLS では異なる用語集にある語彙であっても、概念が同じ語彙には同一の概念コードを振ることにより、多種類の統制用語集を統合した。つまり統制用語集 (シソーラス) のシソーラスであることから、UMLS Metathesaurus® と命名されている。UMLS には他に医学で用いられる一般的な意味関係を記述した Semantic Network と英語の統制語情報を記述した SPECIALIST lexicon が含まれるが、ここでは Metathesaurus® だけを使用している。UMLS には統制用語集のひとつとして MeSH を含む。旧システムでは医学中央雑誌の医学用語シソーラスで利用されている MeSH のバージョンと適合させるために 2002AC を使用しており、ここには 15 カ国の言語 (日本語は含まれていない) で約 208 万の語彙と対応する 87 万概念が収められており、うち英語は 175 万語彙 (87 万概念) である。また新システムでは UMLS 2006AC を利用し、これは 17 カ国の言語 (日本語も含む)、約 620 万語彙 (135 万概念) を収載する。うち英語は 398 万語彙 (135 万概念) である。

4. 1. 2 方法

(1) 材料

日本医学会編纂 医学用語辞典英和第 3 版 (現時点では刊行直前のデータを利用) と UMLS 2006AC 版を材料として利用した。なお、本研究で開発されたシステムの評価を行うために、以前のシステム (平成 13 年度~15 年度の厚生労働省科学研究「UMLS と連携した医学用語シソーラスの作成に関する研究」) で開発されたシソーラスおよびそれを使った MEDLINE 検索システムを利用した。このシステムは UMLS 2002AC を使用している。

日本語医学用語と UMLS および MeSH との関連性を記述したデータベースには Linux 上の PostgreSQL を使用し、日本語医学用語と UMLS・MeSH との関連検索および MEDLINE 検索システムは、Linux 上で PHP を利用してユーザーインターフェイスを構築した。

1) 日本語医学用語と UMLS との対応付け

医学用語辞典英和 (第 3 版) の英語見出しを小文字に変換し、同じく小文字に変換した UMLS の英語表記が完全に一致するものを「UMLS 対応あり」とした。以前は英語文字列の正規化処理を行って両者の一致判定を行ったが、その場合には語順の違いや語尾変化の違いが吸収されることにより、軽微な表記の違いを含む多くの一致を得た反面、誤一致も多く得られてしまったため、今回はあえて正規化を行わず、確実に一致するものだけを対象とした。

2) 日本語医学用語と MeSH との対応付け

上記にて得られた「UMLS と対応する語彙」の中でさらに典拠が MeSH であるものを選択する方法も考えられたが、MeSH との対応関係は用語辞典として、および MEDLINE 検索において重要であることを考え合わせ、英語見出しと MeSH 語彙の両者の英語文字列を正規化して比較を行い、一致するものを対応候補とした上で、個々に対応関係を人手で確認し、誤った対応を削除して最終的な対応関係を得た。

ここで MeSH の全語彙数は 2006AC で 612,109、うち化学物質に関する用語は 437,084 であり、残りの 175,025 語彙 (23,885 概念) が医学に関連する用語である。医学用語辞典の性質から、化学物質に関する対応は不要と考えられるため、対象を医学用語に限ると同時に、

MeSH のカテゴリで最初が D (物質・薬剤) と Z (地理) であるものの多くを除いた。(いずれも医学用語辞典に必要と考えられるもののみを採用)。

3) 医学用語辞典の見出し語 (英語・日本語からの UMLS および MeSH 語彙の検索)

上記作業によって得られた、医学用語辞典の英語見出しと UMLS および MeSH との対応関係に加え、医学用語辞典の英語見出しと日本語訳との対応関係を加えることにより、医学用語辞典の英語検索およびそれに対応する日本語訳、MeSH および UMLS 情報の参照、医学用語辞典の日本語検索およびそれに対応する英語見出し語と MeSH および UMLS 情報の参照を行うユーザーインターフェイスを構築した。

4) 日本語医学用語からの MEDLINE 検索

医学用語辞典に記載されている日本語医学用語が MeSH と対応している場合には、直接 MEDLINE を検索することができる。そこで日本語医学用語から NCBI の Pubmed にリンクするユーザーインターフェイスを構築した。

5) 旧システムとの比較による新システムの評価

新システムでは日本医学会の医学用語辞典英和 改訂第 3 版の元データと UMLS 2006AC を、旧システムでは医学中央雑誌の医学用語シソーラスと UMLS 2002AC を利用している。そこで MEDLINE の検索において、新旧システムの性能を比較する指標として、検索に利用できる MeSH の概念数、対応する日本語の概念数と表記の数、および日本語概念に含まれる表記の平均を調べ、さらに MeSH カテゴリごとの概念カバー率を調べた。また両者に含まれる日本語医学用語で MeSH と対応の取れている語彙に関して、その異なり語彙数や共通語彙数などを調べた。

4. 1. 3 結果

(1) 日本語医学用語と UMLS との対応付け

医学用語辞典の英語見出しの数は 67,417 であり、UMLS の英語と完全一致した医学用語辞典の英語見出しの数は 43,329 であった。およそ 64%の語彙が UMLS と対応付けられている

ことになる。これは医学用語辞典の日本語異なり数 68,672 のうち 44,604 が UMLS と対応していることになる。また対応する UMLS の語彙数は 145,983、統制用語集別に同じ表記も異なる表記とみなすと 274,962、概念数にして 43,652 であった。概念数が元の英語見出しの数よりも多いことから、ひとつの英語見出しが複数の概念に対応していることがわかる。また UMLS は英語だけではなく多言語システムなので、ある概念に属する表記の数を英語以外の言語も含めて算定することができる。これを調べたところ、医学用語辞典の英語見出しと概念レベルで一致した UMLS 内の全言語の語彙数は 1,180,529 であった。

(2) 日本語医学用語と MeSH との対応付け

MeSH と対応する医学用語辞典の英語見出しの数は 23,717 であり、これは辞典の全見出し数の約 35.2%にあたる。これは医学用語辞典の日本語異なり数 68,672 のうち 24,578 であり、約 35.7%である。また対応分は MeSH の概念数にして 17,813 であり、これは MeSH の全医学概念数である 23,885 の約 74.5%をカバーしていることになる。

また MeSH と対応した日本語医学用語は MeSH の各カテゴリにどのように分布しているかを、対応する概念ごとに調べたものを表 1 に示す。MeSH は分類カテゴリに対して多階層構造を持ち、ひとつの概念が複数の分類にカテゴリライズされているため、例えば糖尿病は C18 (栄養疾患と代謝性疾患) と C19 (内分泌系疾患) の 2 つに分類される。そのためカテゴリ総数は概念数よりも多くなっている。方法で述べたように、医学用語辞典の見出し語には物質・薬剤 (D カテゴリ) と地理 (Z カテゴリ) の語彙はあまり採用されていないが、この結果を見るとその程度が把握できる。なお参考までに付録の表 4 に MeSH Tree 番号の上 3 桁ごとに集計した結果を示す。

表 1. 医学用語辞典の見出し語が MeSH と一致したのに関し MeSH カテゴリごとの概念数および MeSH 全体の概念数に対する採用割合
(カテゴリ分類ごとに集計)

分類	辞典	MeSH 全体	割合
A 解剖	1,848	1,895	97.5%
B 生物	2,371	2,594	91.4%
C 疾患	6,750	6,817	99.0%
D 物質・薬剤	4,127	10,395	39.7%
E 診断・治療・機器	2,527	2,554	98.9%
F 心理・精神	736	758	97.1%
G 生命科学・衛生・生理・薬理	2,315	2,385	97.1%
H 物理・数学・標準	507	522	97.1%
I 社会・教育	427	461	92.6%
J 工業・産業・農業・食物	266	298	89.3%
K 人文・宗教	146	176	83.0%
L 情報・分類・図書館・出版	324	362	89.5%
M 人間	174	187	93.0%
N 集団・人口・経済・医療業務・倫理	1,267	1,322	95.8%
V 出版タイプ	3	23	13.0%
Z 地理	1	371	0.3%

(3) 医学用語辞典の見出し語（英語・日本語）からの UMLS・MeSH 検索

システムの画面例を図 1a-d に示す。まず英語または日本語の部分文字列を検索語として与えることにより、医学用語辞典の中でその部分文字列を含む[見出し語・その辞典上の対応する日本語・英語・および UMLS もしくは MeSH に対応していることを示す記号(U・M)]を全て表示する。このリストの中で日本語をクリックすれば辞典から得られる付加情報とし

て同義語およびラテン語表記が存在すれば表示される。また UMLS もしくは MeSH に対応している場合には、英語をクリックすると、UMLS 上で対応する英語表記を表示する。また U または M 欄をクリックすると、対応する MeSH の Main Heading と UMLS での定義(英文)、その MeSH の同義語・関連語、そしてその MeSH 概念に一致する UMLS 上の他の統制用語集での(英語だけではないすべての)表記をすべて表示する。



図 1a. 日本語の検索語から辞書の日本語を検索し、左画面に辞書の内容がリストされる。ここで日本語をクリックすることにより、右画面に辞書の内容が(同義表現やラテン語があればそれも含めて)表示される。

医学用語辞典UMLS検索 - Windows Internet Explorer

http://aquamarine.cchur-tokyo.ac.jp/jumls_new/s_lookup.php

医学用語辞典UMLS検索

検索する日本語を入力してください 糖尿病 検索

95件が見つかりました。要素クリックで関連情報が右側に表示されま
す。

#	日本語	英語	UMLS
1	アロキサン糖尿病	alloxan diabetes	M
2	抗糖尿病薬	antidiabetic agent	M
3	糖尿病治療薬	antidiabetic agent	M
4	ブリットル糖尿病	brittle diabetes	U
5	不安定糖尿病	brittle diabetes	U
6	ブロンズ糖尿病	bronze diabetes	M
7	化学的糖尿病	chemical diabetes	U
8	化学糖尿病	chemical diabetes	U
9	糖尿病合併症	complication of diabetes	M+
10	糖尿病	diabetes mellitus	M+
11	糖尿病性の	diabetic	U
12	糖尿病の	diabetic	U
13	糖尿病性アシドーシス	diabetic acidosis	M
14	糖尿病アシドーシス	diabetic acidosis	M
15	糖尿病性筋萎縮	diabetic amyotrophy	M
16	糖尿病筋萎縮	diabetic amyotrophy	M
17	糖尿病性血管障害	diabetic angiopathy	M+
18	糖尿病血管障害	diabetic angiopathy	M+
19	糖尿病性関節症	diabetic arthropathy	U
20	糖尿病関節症	diabetic arthropathy	U
21	糖尿病性視力低下	diabetic asthenopia	
22	糖尿病性白内障	diabetic cataract	U
23	糖尿病白内障	diabetic cataract	U
24	糖尿病性昏睡	diabetic coma	M+
25	糖尿病性合併症	diabetic complication	M
26	糖尿病合併症	diabetic complication	M

MeSHの表記

Hemochromatosis

TTY	STR	CUI
PM	Bronze Diabetes	C0018995
EP	Diabetes, Bronze	C0018995
PM	Hemochromatoses	C0018995

この英語表記に対応するUMLSの全表記

- C0018995 -

LAT	SAB	STR
CZE	MSHCZE	HEMOCHROMATOZA
DUT	MSHDUT	Bronze diabetes
DUT	MSHDUT	Chromatose, hemo-
DUT	MSHDUT	Carrose, pigment-
DUT	MSHDUT	Hemochromatose
DUT	MSHDUT	Pigment-carrose
DUT	MDRDUT	bronzdiabetes
DUT	ICPC2ICD10DUT	bronzé, diabetes
DUT	ICPC2ICD10DUT	diabetes; bronzé
DUT	MDRDUT	hemochromatose
DUT	ICPC2ICD10DUT	ijzer; stapelingsstoornis
ENG	MSH	Bronze Diabetes
ENG	SNOMEDCT	Bronze diabetes
ENG	SNOMEDCT	Bronze diabetes (disorder)
ENG	MDR	Bronzed diabetes

ページが表示されました

インターネット 100%

図 1b. 日本語の検索語から辞書の日本語を検索し、左画面に辞書の内容がリストされる。ここで英語をクリックすることにより、右画面に UMLS および MeSH に関連する情報が表示される。ここでは MeSH に対応している語彙なので、MeSH での代表語と同義語およびもとの英語表記に一致する UMLS の全表記が表示される。

医学用語辞典UMLS検索 - Windows Internet Explorer

http://aquamarine.cc.hu-tokyo.ac.jp/umls_new/s_lookup.php

医学用語辞典UMLS検索

検索する日本語を入力してください 検索

95 件が見つかりました。要素クリックで関連情報が右側に表示されます。

#	日本語	英語	UMLS
1	アロキサン糖尿病	alloxan diabetes	M
2	抗糖尿病薬	antidiabetic agent	M
3	糖尿病治療薬	antidiabetic agent	M
4	ブリットル糖尿病	brittle diabetes	U
5	不安定糖尿病	brittle diabetes	U
6	ブロンズ糖尿病	bronze diabetes	M
7	化学的糖尿病	chemical diabetes	U
8	化学糖尿病	chemical diabetes	U
9	糖尿病合併症	complication of diabetes	M+
10	糖尿病	diabetes mellitus	M+
11	糖尿病性の	diabetic	U
12	糖尿病の	diabetic	U
13	糖尿病性アシドーシス	diabetic acidosis	M
14	糖尿病アシドーシス	diabetic acidosis	M
15	糖尿病性筋萎縮	diabetic amyotrophy	M
16	糖尿病筋萎縮	diabetic amyotrophy	M
17	糖尿病性血管障害	diabetic angiopathy	M+
18	糖尿病血管障害	diabetic angiopathy	M+
19	糖尿病性関節症	diabetic arthropathy	U
20	糖尿病関節症	diabetic arthropathy	U
21	糖尿病性視力低下	diabetic asthenopia	
22	糖尿病性白内障	diabetic cataract	U
23	糖尿病白内障	diabetic cataract	U
24	糖尿病性昏睡	diabetic coma	M+
25	糖尿病性合併症	diabetic complication	M

MeSHでの定義 (D006432)

C0018995
A disorder due to the deposition of hemosiderin in the parenchymal cells, causing tissue damage and dysfunction of the liver, pancreas, heart, and pituitary. Full development of the disease in women is restricted by menstruation, pregnancy, and lower dietary intake of iron. Acquired hemochromatosis may be the result of blood transfusions, excessive dietary iron, or secondary to other disease. Idiopathic or genetic hemochromatosis is an autosomal recessive disorder of metabolism associated with a gene tightly linked to the A locus of the HLA complex on chromosome 6. (From Dorland, 27th ed)

MeSHのカテゴリ

C16 先天性・遺伝性および新生児疾患と奇形
C18 栄養疾患と代謝性疾患

このMeSHに対応する概念のUMLS表記

- C0018995 -

LAT	SAB	STR
CZE	MSHCZE	HEMOCHROMATOZA
DUT	MSHDUT	Bronze diabetes
DUT	MSHDUT	Chromatose, hemo-
DUT	MSHDUT	Cirrose, pigment-
DUT	MSHDUT	Hemochromatose
DUT	MSHDUT	Pigment-cirrose
DUT	MDRDUT	bronsdiabetes
DUT	ICPC2ICD10DUT	bronzé; diabetes
DUT	ICPC2ICD10DUT	diabetes; bronzé

ページが表示されました

インターネット 100%

図 1c. 日本語の検索語から辞書の日本語を検索し、左画面に辞書の内容がリストされる。ここで M や U の記号をクリックすることにより、右画面に UMLS および MeSH に関連する情報が表示される。ここでは語彙が MeSH に対応しているため、MeSH での定義、MeSH のカテゴリ、そして UMLS でこの概念コードを持つすべての表記が表示される。



図 1d. 英語の検索語から辞書の英語を検索し、左画面に辞書の内容がリストされる。各要素をクリックすることで付加情報を参照する点は日本語の場合と同様である。

(4) 日本語医学用語からの MEDLINE 検索

システム概念図を図 2 に示す。日本語医学用語を検索語として与えることにより、医学用語辞典の中でその部分文字列を含み、かつ MeSH と対応の取れている語彙のリストが表示される。この語彙の中から MEDLINE で検索したいものをクリックすると、画面右側にその語彙に対応する MeSH term を Pubmed で検索した画面が表示される。

また、共出現検索を行うこともできる。共出現検索の欄にチェックを入れて、上記と同様に検索を行って表示された語彙リストからいずれかの語彙をクリックすると、その語彙に共出現した MeSH term (ひとつの文献上で共起した MeSH term) に対応する日本語医学用語が共起頻度順にリストされる。例えばブロンズ糖

尿病という医学用語 (MeSH での代表語はヘモクロマトーシス) に共出現する MeSH term は頻度順に膜タンパク質 (302 件)、クラス I 抗原 (257 件)、鉄 (121 件)、...、となる。そのリストの中から検索したい語彙、例えば「鉄」をクリックすると、「ブロンズ糖尿病」と「鉄」の両方を含む文献が Pubmed で検索され、その結果が画面右側に表示される。この様子を図 3a-b に示す。

なお、NCBI には、Pubmed の他に多くのデータベースがあり、それらを横断的に検索する画面も用意されている。そこでその横断検索に直接アクセスするインターフェイスを図 3c に示す。PubMed の他にも検索を行いたい場合には有用である。

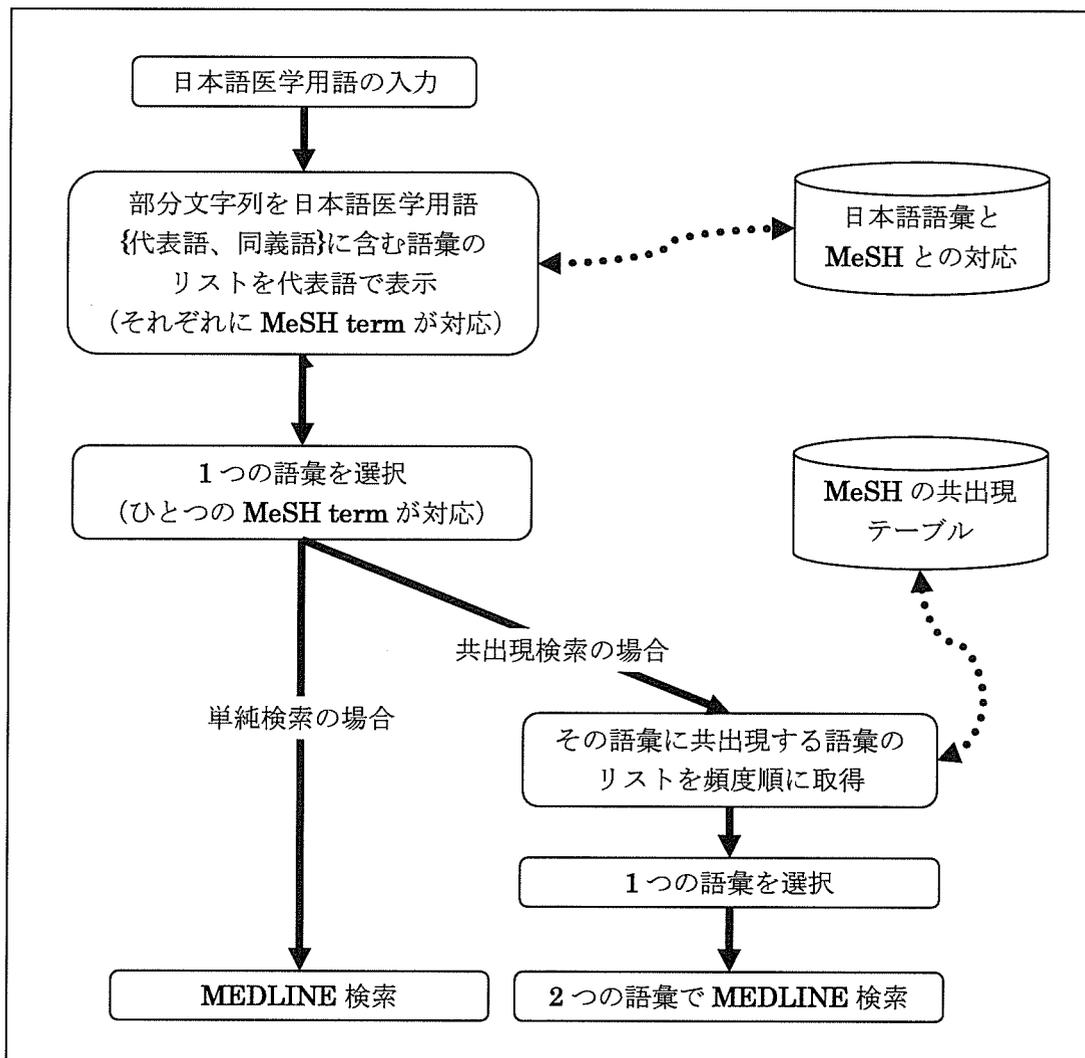


図 2. 日本語による MEDLINE 検索システムの処理図

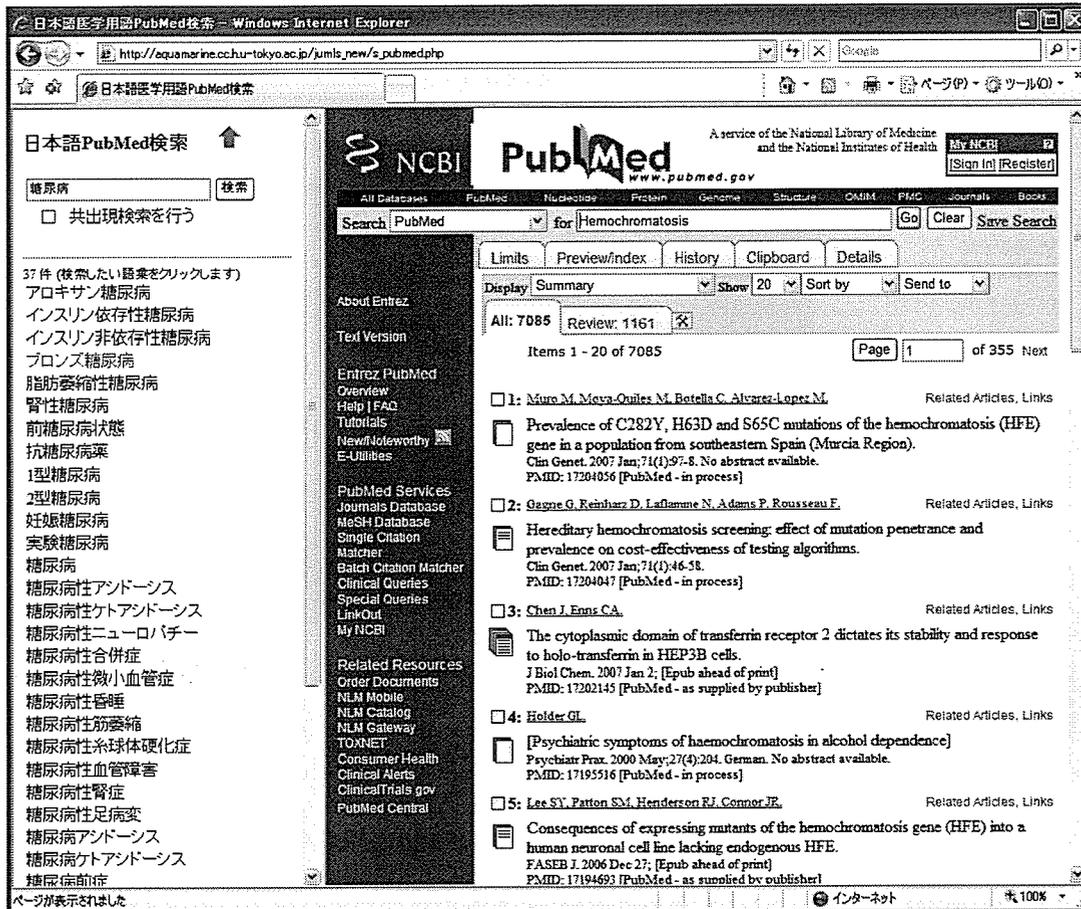


図 3a. 日本語の検索語から辞書の英語を検索し、左画面に検索された辞書の日本語がリストされる。この中の語彙をクリックすることにより、PubMed 検索が行われ、結果が右画面に表示される。

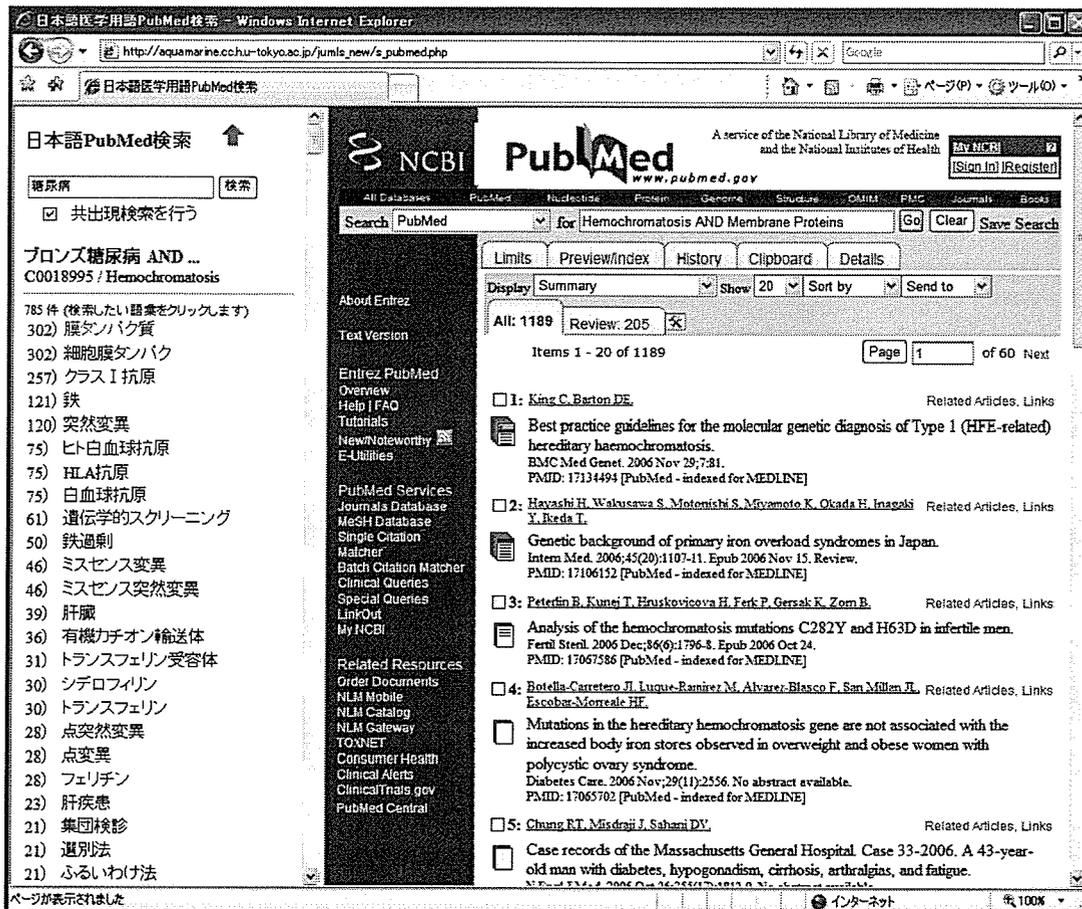


図 3b. 共出現検索欄をクリックして検索を行うと、図 3a と同じリストが画面左に表示される。そこでいずれかの語彙をクリックすると、その語彙に共出現する頻度の高い順に語彙がリストされる。ここでは「ブロンズ糖尿病」と共起する語彙が頻度順に 785 件表示されている。各語彙の左に示されている数字は、その共出現数である。この中の語彙をクリックすることにより、ふたつの語彙を両方とも含む文献が PubMed で検索され、結果が右画面に表示される。

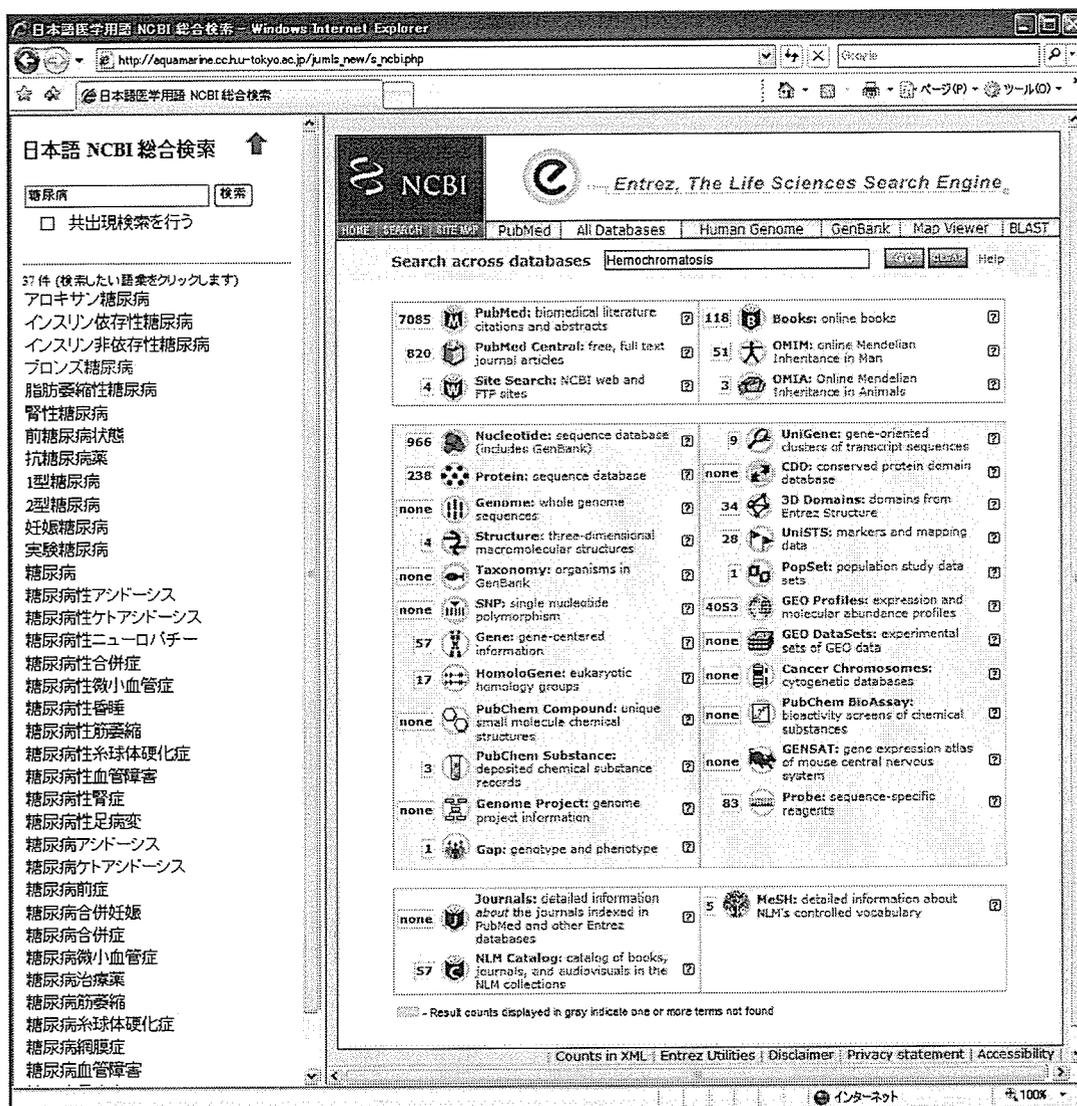


図 3c. PubMed 検索と同じく、NCBI の横断検索をおこなった画面。インターフェイスは PubMed 検索と同様であるが、右画面には NCBI の横断検索の結果が表示される。

(5) 評価 (旧システムとの比較による新システムの評価)

新システムと旧システムで、検索に利用できる MeSH の概念数、対応する日本語の概念数と表記の数、および日本語概念に含まれる表記の平均を調べた結果を表 2 に、また MeSH の概念カバー率 (システムで検索に使うことのできる語彙に対応した MeSH 概念の、全 MeSH 概念に占める割合) を MeSH のカテゴリごとに比較した結果を表 3 に示す。

また旧システムの日本語 (異なり数 45,706) と新システムで MeSH と対応している日本語 (異なり数 24,578) を比較したところ、総和は 57,471、共通は 12,813 であり、旧システムに

あって新システムにない日本語は 32,893、新システムにあって旧システムにない日本語は 11,765 であった。これを図 4 に示す。

表 2. 日本語から MEDLINE 検索を行うシステムの評価指標 (新旧システムの比較)

	旧システム	新システム
対応 MeSH 概念数/全概念数	18,919 / 20,743 (91.2%)	17,813 / 23,885 (74.5%)
対応する日本語表記の数	45,706	25,478
1 概念あたりの日本語表記数	2.42	1.43

表3 MeSH カテゴリごとの概念のカバー率

(カバー率とは、検索に利用できる MeSH 概念と全 MeSH 概念との比)
 (なお 2002AC にはカテゴリ V が存在していないため、それ以外を提示)

	旧システム	カバー率	新システム	カバー率
A	1,686 / 1,704	98.9%	1,848 / 1,895	97.5%
B	1,699 / 2,278	74.6%	2,371 / 2,594	91.4%
C	6,385 / 6,496	98.3%	6,750 / 6,817	99.0%
D	8,879 / 9,155	97.0%	4,127 / 10,395	39.7%
E	2,237 / 2,331	96.0%	2,527 / 2,554	98.9%
F	845 / 867	97.5%	736 / 758	97.1%
G	2,046 / 2,110	97.0%	2,315 / 2,385	97.1%
H	465 / 484	96.1%	507 / 522	97.1%
I	377 / 450	83.8%	427 / 461	92.6%
J	235 / 253	92.9%	266 / 298	89.3%
K	106 / 151	70.2%	146 / 176	83.0%
L	311 / 321	96.9%	324 / 362	89.5%
M	134 / 173	77.5%	174 / 187	93.0%
N	1,039 / 1,265	82.1%	1,267 / 1,322	95.8%
Z	1 / 368	0.3%	1 / 371	0.3%
計	26,445 / 28,406	93.1%	21,941 / 29,225	75.1%
DZ を除く	17,565 / 18,883	93.0%	19,658 / 20,331	96.7%

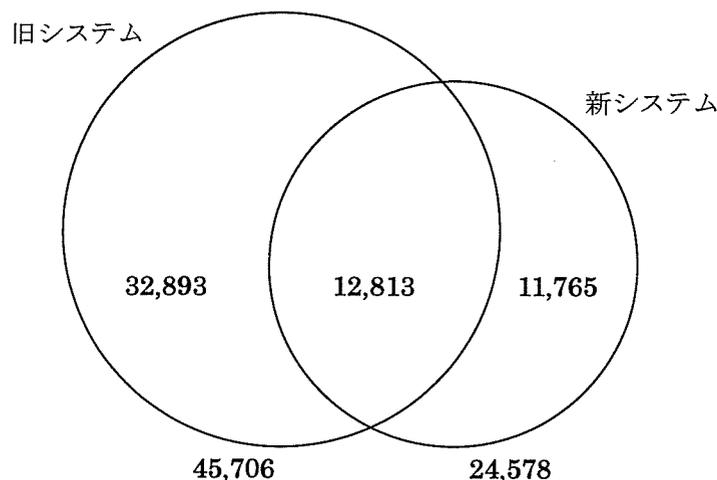


図4. 新旧システムにおける MeSH と対応する日本語語彙の異なり数。

旧システムと比べて新システムでは語彙数が約半分であるにもかかわらず、その中のおよそ半数は旧システムにはない語彙である。

4. 1. 4 考察

システムの評価として新旧システムの比較を行った表2の結果を見ると、本研究で開発した新システムは MeSH 概念のカバー率も、MeSH に対応する日本語の数も、旧システムに劣るように

見える。しかし表3でカテゴリごとに細かく見ると、その内容は旧システムとは大きく異なっていることがわかる。D や Z カテゴリの語彙は医学用語辞典には積極的に収載されていないので、これらのカバー率は低くなっているが、他のカテゴ