

200835022A

厚生労働科学研究費補助金
(地域医療基盤開発推進研究事業)

テキストマイニングによる薬物有害事象の自動抽出を目的とした
オントロジー構築とシステム開発に関する研究

平成20年度 総括研究報告書

Annual Report

Grant-in-Aid for Research on Medical Safety and Medical Technology Evaluation

Supported by the Ministry of Health, Labor and Welfare, Japan in 2008

(Chief Researcher: Shuji KANEKO, Ph.D.)

平成21年(2009)年3月

主任研究者 金子 周司

厚生労働科学研究費補助金
(地域医療基盤開発推進研究事業)

テキストマイニングによる薬物有害事象の自動抽出を目的とした
オントロジー構築とシステム開発に関する研究

平成20年度 総括研究報告書

Annual Report
Grant-in-Aid for Research on Medical Safety and Medical Technology Evaluation
Supported by the Ministry of Health, Labor and Welfare, Japan in 2008
(Chief Researcher: Shuji KANEKO, Ph.D.)

平成21年(2009)年3月

主任研究者 金子 周司

目 次

I. 総括研究報告

テキストマイニングによる薬物有害事象の自動抽出を目的としたオントロジー 構築とシステム開発に関する研究	1
金子周司	

II. 分担研究報告

医療情報解析のためのテキストマイニングエンジンの開発	7
奥野恭史	

III. 研究成果の刊行に関する一覧表

11

IV. 研究成果の刊行物・別刷

13

厚生労働科学研究費補助金（地域医療基盤開発推進研究事業）
総括研究報告書

テキストマイニングによる薬物有害事象の自動抽出を目的とした

オントロジー構築とシステム開発

主任研究者 金子周司 京都大学大学院薬学研究科生体機能解析学分野

研究協力者：大武 博（京都府立医科大学医学研究科），藤田信之（製品開発技術評価機構）

[研究要旨]

本年度はこれまでにライフサイエンス辞書（LSD）から構築したシソーラスに用語を補充し、ツリー状に整理した 2.5 万語の統制語に日英約 18 万語の専門用語を割り当てた LSD シソーラスを完成させた。次に、これら語彙資源を有害事象の自動抽出に応用するため、FDA が公開している副作用報告システム AERS に収録された世界中の医薬品名についてほぼすべての名前解決を行える辞書を制作した。最後に、医薬品添付文書のテキスト解析によって、自由記述される医療情報から正しく医薬品名および疾患・症状名を抽出できるかを詳細に検討した。この LSD シソーラスを利用したテキスト処理は、医療文書からの有害事象の検出に極めて有用な手段であるのみならず、医療情報の解読や入力エキスパートシステムに応用できる優れた方策になると考えられる。

A. 研究目的

本研究は、ゲノム科学における情報科学的手法として発展・応用されつつあるテキストマイニング技術を医薬品の副作用（有害事象）のレポートや医療情報の解析に最適化し、日本語と英語を網羅する医療関連の用語オントロジーをテキスト解析エンジンに実装して、その評価を行いつつ、実効性のある情報解析システムを開発することによって、情報電子化時代を迎える医療における効率良く確かな安全体制の実現を、情報技術的に支援することを目的とした。

本研究は、過去 15 年にわたって広範な生命科学の論文や教科書で用いられる英語と日本語をそれぞれ用語の出現頻度を重要性の目安として

選んできた対訳辞書であるライフサイエンス辞書（LSD）を構造化してオントロジーとするとともに、テキストマイニングを簡便に行うためのプログラムの開発を中心に据えて行ってきた。

最終年度にあたる 20 年度は、構築したシソーラス辞書について、副作用情報の宝庫である米国 FDA の Adverse Event Reporting System (AERS) として公開されている膨大なデータを用いて、構築した辞書の網羅性を確認するとともに、我が国で市販されているすべての医薬品の添付文書を整理した財団法人日本医薬情報センター-JAPIC の医療用医薬品解説の全データをテキストマイニングの材料として実際に解析を行い、解読の実効性

と残された課題およびその解決のための方策について検討した。

B. 研究方法

1. LSD シソーラス構築

LSD には、生命科学の学術論文で用いられる専門用語が英語、日本語の見出し語がそれぞれ 9.5 万および 10.5 万語収録されている (2009 年 3 月現在)。このうち、医療情報を解読するために重要な意味をもつ病名および症候名、医薬品・化合物名、生体内分子名、解剖・発生部位名、生物学名、方法・手法、尺度、現象や概念を意味する日英あわせて 14 万語について、基本的には自作の Perl スクリプトによる集計と手作業での修正によって同義語 synonym を整理し、代表的表記としての統制語 descriptor を定め、統制語の上下関係をツリー状のシソーラス thesaurus に整理し、統制語の関連性に基づくオントロジー ontology の構築を目指した。シソーラスとしては対象分野の広さから NLM が定期的に更新している MeSH に準拠することにした。その際、LSD に収録されている単数形で自然テキスト順の用語と、MeSH に多く見られる複数形および階層性を残した語順の表記を一致させるため、MeSH 用語を標準テキスト順に戻し (例: leukemia, acute → acute leukemia)、かつ複数形を単数形に揃える Perl スクリプトを制作し、自動処理を施せるようにした。

2. オントロジー構築

医薬品に対する適応症と有害事象は、ともに病名・症候名によって記述される。これらを区別するためには、医薬品の薬効分類についての知識を医薬品名称と対応させることが必要である。そこで LSD シソーラスの統制語を用いて、医薬品の作用点データベースを作成した。また、医薬品名を WHO ATC 薬効分類および MeSH Pharmacological Actions に関連づけ、各薬物に薬効分類および化学

構造タグを付与した。なお、既存の病名分類である ICD-10 準拠標準病名マスターおよび ICH 国際医薬用語集 MedDRA/J と関連づけた。

3. AERS による辞書の評価

米国 FDA が公開している世界規模の医薬品有害事象データベース AERS では、薬物処置を施した症例の適応症と有害事象の記述に MedDRA の Preferred Term (PT) が用いられて一貫性を保っているが、医薬品の名称には商品名や一般名など多様な表記のゆれを含むため情報の統計的な処理が難しい。そのため、LSD シソーラスから医薬品名を抽出し、4 万種類以上の名称を約 4,500 種類の統制語にマッピングした医薬品名称解決のための辞書を制作した。

次に 2004 年から 2008 年第二四半期までに公開された 4 年半分 150 万件の AERS レポートをリレーショナルデータベースに再構築し、医薬品名称の解決を試みた。なお、適応症や反応事象の解決は MedDRA ver.11 を用いた。

4. JAPIC 医薬品情報のテキストマイニング

本研究の目的は電子カルテなど、医療現場で実際に発生する医療文書からの知識抽出であるが、特定の症例に基づく実証を行うまでに、網羅的な知識抽出が行えるかどうかの検証を行う必要がある。そこで適応症や副作用の記述に特別な取り決めが無く自由記述がなされている医薬品添付文書を材料にテキストマイニングを行ってみた。

まず LSD シソーラスから、関係抽出のための辞書を試作した。英文および和文テキストに対して文中の専門用語を認識して自動的に統制語タグを付与し、タグづけされた統制語の頻度と統制語同士の共起頻度をカウントするスクリプトを制作した。

その上で、2008 年版の JAPIC 医療用医薬品データベースに収録された 18,517 件の医薬品添付文

書テキストに対して、LSD シソーラスと自作スク립トを用いて専門用語へのタグづけを行った。用語頻度の集計を行い、さらに一部分を抽出して適合率と再現率の評価を行った。

C. 研究結果

1. LSD シソーラス構築

LSD に収録された日英 20 万語のうち、医療情報の解読に必要な概念としては、病名および症候名、薬物および生体内分子名、解剖・発生部位名、生物学名、方法や研究技術、現象や概念を意味する用語が必要十分であると考えられた。そこで本年度は最新の MeSH 2008 年版とリレーショナルデータベースで動的に関連づけた。また、公開されている病名分類や薬効分類へのリレーションを設けた。この結果、MeSH からの追加を含め、約 18 万語の専門用語を約 2.5 万語のツリー状に整理した統制語に割り当てた LSD シソーラスが完成した（詳細については総合報告書で解説する）。しかしながら一方で、LSD に収録されながら MeSH と照合できないため統制語に帰属されない用語が英語で 1 万語以上も存在することが明らかになった。特に、病名・症候名や解剖部位名においては、国内で用いられている標準病名マスターや MedDRA にも収録されながら、MeSH に帰属できない用語が数多く残された。また、医薬品としては国内医薬品において収録されていない用語が若干、存在した。今後は、これらの語句を帰属させるために統制語を拡張するとともに、複数のツリー体系に対応できるよう、データベース仕様を改良していく必要が示された。

2. オントロジー構築

医薬品と適応症との対応関係は JAPIC「添付文書記載病名集」などによって整理されつつある。本研究ではより分子作用に立脚した医薬品オントロジーを目指し、医薬品と生体の相互作用を数

少ない相互作用様式とともに三項関係によって記述する新たなデータベースを構築した。

このデータを制作するため、すべての統制語について PubMed より代表的な学術誌に掲載された 10 年分の論文抄録 (600 M バイト) を題材として、テキスト中に共起する統制語の頻度を収集し、関連概念としてデータベース化した。その中には医薬品の作用点あるいは薬効分類に属する用語が数多く集まっていたため、これらを材料にして手作業によるオントロジー化を行った。

LSD シソーラスに収録された薬理活性のある化合物の生体作用について、受容体、酵素、膜輸送タンパク質などの分子レベルでの相互作用が記述できる場合は、作用点となる生体分子を統制語あるいは RefSeq、酵素 EC 番号を用いて特定した。その相互作用様式については、阻害、活性化など少数の記述子を用いて三項関係として表現した。標的分子が明らかでない場合は、細胞ないし組織レベルにおいて報告されてきた知識を同じく統制語と作用様式を表す記述子を用いて整理した。

これらの結果、ほとんどの化合物について、薬物・標的分子（あるいは標的細胞ないし組織）・相互作用様式の三項関係を記述することができた。さらに薬物を表す統制語は WHO ATC 分類および MeSH Pharmacological Actions と関連づけ、各薬物に薬効分類および化学構造タグを付与した。

一方、病名に関しても複数存在する病名の階層化に対応するため、既存の病名分類である ICD-10 準拠標準病名マスターおよび ICH 国際医薬用語集 MedDRA/J と関連づけた。

3. AERS による辞書の評価

AERS は米国 FDA が公開している医薬品の有害事象データベースである。自発報告であるが適応症と反応事象に対しては MedDRA による用語の統制がとれているため、医薬品名の解決が可能に

なれば有害事象のデータマイニングに有用な研究資源となる。そこで本研究では構築した LSD シソーラスでの医薬品の網羅性を検証するため、AERS の名前解決を試みた。

AERS レポート 150 万件に出現する医薬品を解析したところ、18 万種類の異なる名称が検出された。しかし 45,000 種類の医薬品名を 3,900 語の統制語に整理した LSD シソーラスとの関係づけを直接、加工せずに行った段階でも 80%以上のレコードで薬物の有効成分名を解決することができた。関係づけできなかったレコードの多くは、複数の医薬品名が記載されたり余分な注釈が記載されたレコードであったため、AERS の元データに適切な二次加工を施すことで対処が可能であり、最終的に AERS に出現する医薬品レコードの 96% (Primary Substance に限定すると 98%) が同定できた。また、すべての医薬品について、先に述べたように分子作用点、薬効分類、構造分類の属性が付与されているため、例えば、「ベンゾジアゼピン系催眠薬」、「フェノチアジン化合物」、といった、ある一群の医薬品を指す用語を用いても AERS の検索が可能になった。

なお、本研究で制作した AERS 医薬品名解決辞書はすでに JAPIC に導入され、有害事象検索サービスとして実用化されている。また、京都大学薬学部において制作している学生教育用医療データベース (薬学統合ナビゲーションシステム) においても本辞書を実装し、医薬品情報や教材に出現する医薬品名から AERS の有害事象を検索できるシステムを医療薬学教育に用いはじめたところである。

4. JAPIC 医薬品情報のテキストマイニング

本研究の最終目標である医療文書のテキストマイニングによる有害事象の発見のためには、テキスト中のすべての専門用語が適切にもれなくタグづけられ、医薬品と病名・症状との共起関係

が抽出される必要がある。

この処理を行うため、本研究では専門用語を最長一致で発見し、LSD シソーラスを用いて統制語タグを付与する Perl スクリプトを開発した。また、処理速度や内容の質を評価するため、本研究は文書の題材として英語用は世界標準の薬理学教科書 Goodman & Gilman's The Pharmacological Basis of Therapeutics 10th edition の全テキスト (8 M バイト)、日本語用は JAPIC 医療用医薬品集 (56 M バイト) を用意した。

スクリプトによる処理は、テキスト中に出現する専門用語に属性を持つ統制語 XML タグを付与し、さらに医薬品名と病名・症候名が 1 文中で共起した場合にそれらを抽出するようにした。これを用いた場合、市販のパソコン (Apple Mac Pro) でも 1 分間に約 7 M バイトもの大量のテキストから、医薬品と適応症や副作用にタグ付けを行い、それらの出現頻度や共起関係を計数することが可能であり、十分な速度が得られた。

次に、タグ付けテキストの内容をブラウザで確認しながら、曖昧性の排除と統制語の最適化を行った。この過程において、テキストでの一致のみによって統制語への変換を行う場合、曖昧性を排除するために多義性のある略語や商品名等、一部のシノニムをタグ付け辞書から除外する必要性が生じた (約 200 語)。また、「ヒト human」、「病気 disease」、「酸 acid」等のように、非常に大きな概念は関連するキーワードとして不必要あるいは不適切と考えられたため、それら (約 360 語) もタグ付け辞書から除外した。

以上の最適化辞書を用いて、20 種類の医薬品添付文書に記載された副作用 (有害事象) のテキストマイニングを行った結果、適合率、再現率ともに 90%以上の成績が得られた。

しかしながら、この解析の結果、医薬品添付文書やインタビューフォームに記載された相互作用や副作用情報は規制用語である MedDRA や標

準病名等とは必ずしも一致せず、数多くの異表記を含むうえに、読み手の専門知識を前提とした記述で満ちていることが明らかになった。例えば、相互作用に注意すべき医薬品の呼び名として、「カルシウム拮抗剤」や「Ca 拮抗剤」のような表記のゆれとともに、「ジヒドロピリジン系薬剤」や「ニフェジピン等」などの表現があり、これが商品名「アダラート」を初めとする一群の「ニフェジピン」を中心とする一群の「降圧薬」との併用に注意を喚起している文章であることは、これらの関係を正確に解釈できることを前提としている。同様な表記の多様性は適応症や副作用についても指摘できる。幸いにも LSD シソーラスではそれらに関連づけるデータを有しているため、実用レベルのテキストマイニングにおいては名前解決を図るだけでなく、階層（粒度）の異なる用語を場合に依りて展開しながら解析を行う必要がある。

D. 考察

医療情報化社会において医療等の安全を達成するためには、市販後の医療情報や調査データを解析することによる有害作用の知識発見を早期に、確実かつ網羅的に行う必要がある。そのような医療情報のほとんどは、文章（テキスト）として記述される。医療現場において今後、急速に電子化が推進され、大量のテキスト情報が発生すると予想できる。しかし医療情報を記述する用語については、病名、医薬品名などで国際的協調によって表記の統一化の努力が続けられているが、実際に FDA 等で公開されている医療テキストを解析すると、様々な用語が統一されずに用いられていることが本調査研究から分かってきた。我が国において状況はさらに深刻であり、医薬品副作用報告で MedDRA/J などの規制用語が用いられているものの、医薬品添付文書を初めとして、それ以外の医療文書では英語以上に多種多様の日本

語が使われているのが実態であった。これまでの用語集は表記や分類を統一する方針で制作されており、網羅性に問題がある。医療情報の解析を行うためには、自然言語処理によって英語と日本語の語彙を網羅し、かつ事物（医薬品等）や概念（病名等）の同義性や関連性をツリー状に整理したオントロジーを構築し、常に最新の状態で維持することが最優先の課題である。

テキストマイニング技術をゲノム科学に応用し、遺伝子と発現プロファイルや代謝パスウェイとの関係から創薬標的の発掘や遺伝子の機能推定を行おうとする研究は情報科学の領域で盛んに行われ、一部は商品化もされている。しかし、テキストマイニングを副作用情報の発見に応用しようとする試みはほとんどなかった。機能が不明な遺伝子の機能推定とは異なり、薬品名や症候名は（表記は統一されていないものの）限られた数の語彙から構成されており、本来コンピュータによるテキストマイニング処理には適した材料である。しかしながら、過去に誰も着手しなかった最大の原因は、あらゆる文書での日本語・英語を網羅する「辞書」が存在しなかったためと考えられる。事実、医薬品名、遺伝子、病名の各用語はそれぞれ独自に国内外で規定されているが、これらの相互関係を多様なボキャブラリを含めて網羅的に記述したデータベースは今なお皆無である。専門領域に特化したオントロジーの構築は、検索エンジン技術を発展させる研究として情報科学でもきわめて注目されており、本研究で構築した LSD シソーラスは、医学オントロジーのプロトタイプとして有用な資源であると考えられる。

本研究によって、臨床現場から発生する大量の電子化された生の文書を早期に定量的に分析し、有害事象の早期発見を可能にするシステムの開発が可能性を帯びてきた。すなわち、様々な医療情報から有害事象が疑われるレポートを自動抽出し、人間による最終的な知識発見を支援する実

用システムが完成されよう。

しかし日本語は表記のゆれが著しいため、日本語解析のためのスクリプトを形態素解析と融合させることによって改良する必要がある。また、シソーラス辞書は適合率、再現率 90%以上の性能を有しているが、医療情報の解析において検索漏れは許されない。シソーラス制作時に 1 万語以上の該当漏れが生じたことから明らかなように、今後さらに網羅性の高いシソーラスを完成させる必要がある。

本研究は、抽出される薬物 (A) と有害事象 (B) との関係抽出に留まらず、DNA アレイを用いた他の研究における薬物 (A) と遺伝子 (C) との関係と論理的に組み合わせることによって、有害事象 (B) と遺伝子 (C) との関係が示唆されることになり、副作用メカニズムを実験科学で立証するための着眼点が提案される。これらの諸観点から、本研究の成果は ICT 医療時代にあつて極めて高い実現性と波及効果が期待される。

E. 結論

本研究によって、医薬品と疾患、症状に加えて、関連する技術や方法、解剖部位や生物名までを網羅した頑強な医学オントロジーがほぼ完成した。また、英語および日本語テキストについて文中でのキーワード共起解析を高速かつ簡便に行うための処理プログラムを開発し、テキストマイニングからデータマイニングへの橋渡しが可能であることを示した。

今後はさらに解析結果のフィードバックからシソーラス辞書の網羅性を高めると共に、共起解析の結果を二項関係から関連性の解読へと発展させることで医療文書の解読や入力エキスパートシステムの構築に向けて製品開発が期待できる。その際、本研究で制作した医薬品と薬理作用点データベースに、JAPIC が制作した医薬品と適応症のデータベースを組み合わせることで、有害

事象の判別が可能になると考えられる。実際の医療文書の解析と評価を繰り返すことによって、十分な実用性と有用性を有する医療情報システム設計が可能になると結論できる。また本研究は、電子カルテやオーダリングシステムにおいて医薬品添付文書の記載に基づく相互作用や禁忌症など使用上の注意に対する警告ないし助言を可能にする等、医療情報システムのインテリジェント化を推進するためにも有用な資源となると期待できる。

なお、本研究で構築した LSD シソーラスは、無料検索サービスとして京都大学サーバで公開している。今後、情報ポータルとしても有用性を高めていく予定である。

F. 研究発表 (学会発表)

1. 医学用語シソーラスに基づく効率的医療情報検索システムの開発 (論文 PDF, 金子周司, 鶴川義弘, 大武博, 河本健, 竹内浩昭, 竹腰正隆, 天野博夫, 藤田信之, 第 28 回医療情報学連合大会, 2008 年 11 月, 横浜)
2. 医薬品名称の同義語解決による有害事象データベース AERS の情報活用 (金子周司, 第 18 回医療薬学会年会, 2008 年 9 月, 札幌)
3. 国内外の医療用医薬品名を網羅する同義語辞書の制作 (金子周司, 第 11 回日本医薬品情報学会学術大会, 2008 年 7 月, 東京)

G. 知的財産権の出願・登録状況 (予定も含む)

無し

実用新案登録

無し

厚生労働科学研究費補助金（地域医療基盤開発推進研究事業）
分担研究報告書

医療情報解析のためのテキストマイニングエンジンの開発

分担研究者：奥野恭史（京都大学大学院薬学研究科・システム創薬科学講座）

[研究要旨]

本研究は、医薬品の副作用（有害事象）のレポートや医療情報の解析・評価に、テキストマイニング技術を適用し、薬物有害事象の情報解析システムを開発することにより、IT 時代を迎える医療における効率良く確かな安全体制の実現を情報技術的に支援することを目的とする。

最終年度となる本年度の分担研究としては、「XML データベースの構築とテキストマイニングエンジンの開発」、「医療情報とゲノム情報の連携システム（GEM-TREND）の開発」の 2 点の研究開発を行った。

「XML データベースの構築とテキストマイニングエンジンの開発」としては、昨年まで開発してきた JAPIC 添付文書記載病名集データベースに加え、米国 FDA が提供する副作用情報データ AERS (Adverse Event Reporting System) をコンテンツとし、テキストマイニングエンジンの改良として全文検索ツール Hyper Estraier の実装を行った。また、「医療情報とゲノム情報の連携システム（GEM-TREND）の開発」としては、これまで開発してきた医療情報テキストマイニングシステムとゲノム情報の連携を図るために、米国 NIH/NCBI が提供する遺伝子発現データベース（GEO: Gene Expression Omnibus）への独自検索システム GEM-TREND (Gene Expression Data Mining Toward Relevant Network Discovery) を開発し、Web 公開を行った。

A. 研究目的

本研究は、医薬品の副作用（有害事象）のレポートや医療情報の解析・評価に、テキストマイニング技術を適用し、薬物有害事象の情報解析システムを開発することにより、IT 時代を迎える医療における効率良く確かな安全体制の実現を情報技術的に支援することを目的とする。

B. 研究方法

1. XML データベースの構築とテキストマイニングエンジンの開発

薬物有害事象の自動抽出を目的としたテキスト

マイニングエンジンのデータ素材として、(財) 日本医薬情報センター（JAPIC）の添付文書記載病名集に加えて、米国 FDA が提供する副作用情報データ AERS (Adverse Event Reporting System)、および医学用語辞書 MedDRA (Medical Dictionary for Regulatory Activities)、UMLS (Unified Medical Language System) を新規に用いた。テキストマイニングエンジンには、昨年までに用いていた有料の NeoCore から、無償ツールの全文検索エンジン Hyper Estraier への実装変更を行った。さらに、検索のユーザビリティ向上を目指し、研究代表者が開発するライフサイエンス辞書

(<http://www.pharm.kyoto-u.ac.jp/channel/ja/dictionary/index.html>)との連携を図り、検索した単語の関連語・共起表現機能を実装した。

2. 医療情報とゲノム情報の連携システム (GEM-TREND) の開発

ヒトゲノムが解読された今日、ゲノム情報と医療や医薬に用いる試みが精力的に研究展開されている。そこで、これまで開発してきた医療情報データベースおよびテキストマイニングシステムとゲノム情報の連携を図るために、米国 NIH/NCBI が提供する遺伝子発現データベース (GEO: Gene Expression Omnibus) への独自検索システム GEM-TREND (Gene Expression Data Mining Toward Relevant Network Discovery) を開発した。本ツールは Linux server 上に、MySQL、PHP および R を用いて実装し、遺伝子ネットワークの可視化には Java Applet graphical user interface を用いた。

なお、本研究は計算機によるシステム開発であり、倫理面に関する問題は一切無い。

C. 研究結果

1. XML データベースの構築とテキストマイニングエンジンの開発

複雑・多岐に渡るデータの構成に柔軟に対応し、かつ大量のデータを高パフォーマンスで処理できる仕組みとして、全文検索エンジン Hyper Estraier と XML データベースとの連携を行った。データベースに登録する文書データには、昨年までに実装した JAPIC 医薬品添付文書記載病名集に加え、米国 FDA が公開している大規模副作用症例データベース Adverse Event Reporting System (AERS) を新たに実装した。JAPIC 医薬品添付文書記載病名集は、日本における医薬品の効能効果、禁忌情報、副作用情報などが記載されているのに対し、米国 FDA の AERS は世界各国における副

作用症例が集積されたデータベースであり、薬物有害事象のテキストマイニングの対象素材としても最適なコンテンツである。

また、開発したテキストマイニングシステムに、研究代表者が開発するライフサイエンス辞書や医学用語辞書 MedDRA (Medical Dictionary for Regulatory Activities)、UMLS (Unified Medical Language System) を組み込み、検索した単語の関連語・共起表現を日本語・英語で示すことにより、検索のユーザビリティ向上を図った。

2. 医療情報とゲノム情報の連携システム (GEM-TREND) の開発

近年のマイクロアレイ技術の発展に伴い、大量の遺伝子発現データが GEO (gene expression omnibus) に代表される公共データベースに蓄積されてきている。我々は、キーワードや独自の遺伝子発現データを問合せ情報として、これら遺伝子発現データベースから、遺伝子発現データを検索するツール GEM-TREND (Gene Expression Data Mining Toward Relevant Network Discovery) を開発した。

GEM-TREND では、任意のキーワードに加え、gene expression signature をベースにしたマイクロアレイデータの検索および得られたデータのネットワークを可視化する機能を実装している。gene expression signature をベースにした検索では、nonparametric、rank-based pattern matching approach を用いて、query とデータベース内に蓄積された GEO 発現データに対応した sample ごとの遺伝子ランクデータの signature を比較し、得られた similarity によって検索結果を決定する評価系を用いている。また、描画画面の遺伝子ネットワークの構築には Pearson の相関係数と K-means clustering を用いて行った。開発したシステムは、以下の URL で Web より公開している。
<http://cgs.pharm.kyoto-u.ac.jp/services/network/>

D. 考察

1. XML データベースの構築とテキストマイニングエンジンの開発

今回改良導入を行った新規のテキストマイニングエンジン Hyper Estraier は、高速全文検索を可能にする特徴は言うまでもなく、無償ツールであることからシステムの公開も可能であり、汎用性の高いシステムとして大きな利点を有する。

また、JAPIC 添付文書に加え、AERS の導入により、国内に止まらず、海外の副作用事例を加味した情報抽出、検索が可能となり、医薬品の安全性に大きな寄与を示すものと期待される。

2. 医療情報とゲノム情報の連携システム (GEM-TREND) の開発

今回開発した GEM-TREND を用いて、薬物名によるデータベース検索を実行し、その検証を行ったところ、高い精度で薬物投与と遺伝子発現データの取得が確認できた。今後は、GEM-TREND データベースと、上述、テキストマイニングデータベースとの直接的な連携を図る。

E. 結論

1. XML データベースの構築とテキストマイニングエンジンの開発

JAPIC 添付文書記載病名集データベースに加え、米国 FDA の大規模副作用症例データベース Adverse Event Reporting System (AERS) を新たに実装したテキストマイニングツールの開発を行った。全文検索アルゴリズムとして、無償ツールである Hyper Estraier の実装に成功した。

2. 医療情報とゲノム情報の連携システム (GEM-TREND) の開発

これまで開発してきた医療情報テキストマイニングシステムとゲノム情報の連携を図るために、米国 NIH/NCBI が提供する遺伝子発現データベー

ス (GEO: Gene Expression Omnibus) への独自検索システム GEM-TREND (Gene Expression Data Mining Toward Relevant Network Discovery) を開発した。これらは、

<http://cgs.pharm.kyoto-u.ac.jp/services/network/> より公開している。

F. 研究発表

1. 論文発表

1. van der Horst, E., Okuno, Y., Bender, A. and Ijzerman, A.P., "Substructure mining of GPCR ligands reveals activity-class specific functional groups in an unbiased manner", *J. Chem. Inf. Model.*, **49**, 348-60, 2009
2. Tsuchiya, S., Tachida, Y., Segi-Nishida, E., Okuno, Y., Tamba, S., Tsujimoto, G., Tanaka, S. and Sugimoto, Y., "Characterization of gene expression profiles for different types of mast cells pooled from mouse stomach subregions by an RNA amplification method", *BMC Genomics*, **10**: 35, 2009
3. Ruike, Y., Ichimura, A., Tsuchiya, S., Shimizu, K., Kunimoto, R., Okuno, Y., and Tsujimoto, G., "Global correlation analysis for micro-RNA and mRNA expression profiles in human cell lines", *J. Hum. Genet.*, **53**, 515-23, 2008
4. Kawanishi, H., Matsui, Y., Ito, M., Watanabe, J., Takahashi, T., Nishizawa, K., Nishiyama, H., Kamoto, T., Mikami, Y., Tanaka, Y., Jung, G., Akiyama, H., Nobumasa, H., Guilford, P., Reeve, A., Okuno, Y., Tsujimoto, G., Nakamura, E. and Ogawa, O., "Secreted CXCL1 is a potential mediator and marker of the tumor invasion of bladder cancer", *Clin. Cancer Res.*, **14**, 2579-87, 2008
5. Takano, H., Nakazawa, S., Okuno, Y., Shirata, N., Tsuchiya, S., Kainoh, T., Takamatsu, S., Furuta, K., Taketomi, Y., Naito, Y., Takematsu, H., Kozutsumi, Y., Tsujimoto, G., Murakami, M., Kudo, I., Ichikawa, A., Nakayama, K., Sugimoto, Y. and Tanaka, S., "Establishment of the culture model system that reflects the process of terminal differentiation of connective tissue-type mast cells", *FEBS Lett.*, **582**, 1444-50, 2008
6. Okuno, Y., "In silico drug discovery based on the integration of bioinformatics and

chemoinformatics”, *YAKUGAKU ZASSHI*, 128 (11), 1645-51, 2008

7. ケミカルゲノミクス情報を用いた新規リガンド探索手法

藪内 弘昭, 奥野 恭史, *SAR News*, 14, 2-6, 2008

8. Inoue, T., Adachi, H., Murakami, S., Takano, K., Matsumura, H., Mori, Y., Fukunishi, Y., Nakamura, H., Kinoshita, T., Nakanishi, I., Okuno, Y., Minakata, S., Shimojo, S., Sakata, T. “New progress in crystallization technology of membrane protein and introduction of pharmaceutical innovation value chain” *YAKUGAKU ZASSHI*, 128 (4), 497-505, 2008

9. ケミカルゲノミクスに基づくインシリコ創薬
新島 聡, 奥野 恭史, *日薬理誌*, 133, 173, 2009

2. 学会発表

1. 関西バイオネットワーク「創薬バリューチェーンの構築に向けて」発表交流会

「非結晶性標的タンパク質に対する化合物探索」2008年12月8日

2. 応用トキシコロジーリカレント講座「ケモゲノミクスとトキシコゲノミクスの融合」

2008年9月12日

3. 日本たばこ産業株式会社 医薬総合研究所
社内講演会「ケミカルゲノミクス情報の活用によるインシリコ創薬」2008年9月2日

4. キッセイ薬品工業株式会社 中央研究所
社内講演会「ケミカルゲノミクス情報の活用によるインシリコ創薬」2008年8月20日

5. 生化学工業株式会社 中央研究所
社内講演会「ケミカルゲノミクス情報の活用によるインシリコ創薬」2008年8月19日

6. アスピオファーマ株式会社 生物医学研究所

セミナー「ケミカルゲノミクス情報を用いた生物活性に富んだケミカル空間の合理的探索」

2008年6月5日

G. 知的財産権の出願・登録状況（予定も含む）

無し

2. 実用新案登録

無し

3. その他

無し

研究成果の刊行に関する一覧表

書籍

著者氏名	書籍全体の編集者名	書籍名	出版社名	出版地	出版年
金子周司ほか	総編集 伊藤正男, 井村裕夫, 高久史麿	医学大辞典 第2版	医学書院	日本	2008
金子周司 野恭史ほか	藤井信孝・辻本豪三・ 奥野恭史 編集	インシリコ創薬 科学-ゲノム情報 から創薬へ-	㈱京都廣川書店	日本	2008

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
金子周司, 鷗川義弘, 大武博, 河本健, 竹内浩昭, 竹腰正隆, 天野博夫, 藤田信之	医学用語ソーラスに 基づく効率的医療情報 検索システムの開発	医療情報学	28 (suppl)	639-642	2008
van der Horst, E., Okuno, Y., Bender, A. and Ijzerman, A.P.	Substructure mining of GPCR ligands reveals activity-class specific functional groups in an unbiased manner	J. Chem. Inf. Model.	49	348-60	2009
Tsuchiya, S., Tachida, Y., Segi-Nishida, E., Okuno, Y., Tamba, S., Tsujimoto, G., Tanaka, S. and Sugimoto, Y.	Characterization of gene expression profiles for different types of mast cells pooled from mouse stomach subregions by an RNA amplification method	BMC Genomics		10 : 35	2009
Ruike, Y., Ichimura, A., Tsuchiya, S., Shimizu, K., Kunimoto, R., Okuno, Y., and Tsujimoto, G.	Global correlation analysis for micro-RNA and mRNA expression profiles in human cell lines	J. Hum. Genet.	53	515-23	2008
Kawanishi, H., Matsui, Y., Ito, M., Watanabe, J., Takahashi, T., Nishizawa, K., Nishiyama, H., Kamoto, T., Mikami, Y., Tanaka, Y., Jung, G., Akiyama, H., Nobumasa, H., Guilford, P., Reeve, A., Okuno, Y., Tsujimoto, G., Nakamura, E. and Ogawa, O.	Secreted CXCL1 is a potential mediator and marker of the tumor invasion of bladder cancer	Clin. Cancer Res.	14	2579-87	2008

Takano, H., Nakazawa, S., <u>Okuno, Y.</u> , Shirata, N., Tsuchiya, S., Kainoh, T., Takamatsu, S., Furuta, K., Taketomi, Y., Naito, Y., Takematsu, H., Kozutsumi, Y., Tsujimoto, G., Murakami, M., Kudo, I., Ichikawa, A., Nakayama, K., Sugimoto, Y. and Tanaka, S.	Establishment of the culture model system that reflects the process of terminal differentiation of connective tissue-type mast cells	FEBS Lett.	582	1444-50	2008
<u>Okuno, Y.</u>	In silico drug discovery based on the integration of bioinformatics and chemoinformatics	YAKUGAKU ZASSHI	128 (11)	1645-51	2008
藪内 弘昭, <u>奥野 恭史</u>	ケミカルゲノミクス情 報を用いた新規リガン ド探索手法	SAR News	14	2-6	2008
Inoue, T., Adachi, H., Murakami, S., Takano, K., Matsumura, H., Mori, Y., Fukunishi, Y., Nakamura, H., Kinoshita, T., Nakanishi, I., <u>Okuno, Y.</u> , Minakata, S., Shimojo, S., Sakata, T.	New progress in crystallization technology of membrane protein and introduction of pharmaceutical innovation value chain	YAKUGAKU ZASSHI	128 (4)	497-505	2008
新島 聡, <u>奥野 恭史</u>	ケミカルゲノミクスに 基づくインシリコ創薬	日薬理誌	133	173	2009

医学用語シソーラスに基づく効率的医療情報検索システムの開発

金子 周司¹⁾ 鷗川 義弘²⁾ 大武 博³⁾ 河本 健⁴⁾ 竹内 浩昭⁵⁾ 竹腰 正隆⁶⁾
天野 博夫¹⁾ 藤田 信之⁷⁾

京都大学大学院薬学研究科¹⁾ 宮城教育大学²⁾ 京都府立医科大学³⁾
広島大学医歯薬総合研究科⁴⁾ 静岡大学理学部⁵⁾ 東海大学医学部⁶⁾
製品評価技術基盤機構⁷⁾

Development of medical portal system based on the thesaurus and collocation analysis of biomedical terms

Kaneko Shuji¹⁾ Ugawa Yoshihiro²⁾ Ohtake Hiroshi³⁾ Kawamoto Takeshi⁴⁾
Takeuchi Hiroaki⁵⁾ Takekoshi Masataka⁶⁾ Amano Hiroo¹⁾
Fujita Nobuyuki⁷⁾

Kyoto Univ. Grad. Sch. Pharm. Sci.¹⁾ Miyagi Univ. Edu.²⁾ Kyoto Pref. Univ. Med.³⁾
Hiroshima Univ. Grad. Sch. Biomed. Sci.⁴⁾ Shizuoka Univ. Fac. Sci.⁵⁾ Tokai Univ. Sch. Med.⁶⁾
NITE⁷⁾

The internet has become a commonly used and popular form of media by which health care information can be searched and retrieved not only by general populace but also by medical students and professionals. Accordingly, an increase in the number of Japanese documents easily available through the Internet has caused the relative avoidance of original scientific papers and databases written in English. To develop a translational portal site that enables the use of Japanese for medical information retrieval, we have constructed an English-Japanese thesaurus containing 160,000 terms collected from biomedical literatures. We have assigned individual terms to the MeSH descriptors (25,000 English-Japanese pairs), and analyzed the top-30 collocations of the terms in medical literature (total 1.4 million pairs). The collocation data are available for associative search, equipped with a tree-style thesaurus in the online Life Science Dictionary (LSD). The free LSD portal site enables the use of Japanese terms for information searches linked to Entrez-PubMed and Google sites. In addition, we have developed a new dictionary for Mac OS X 10.5 that can be used in the Safari browser as a one-touch dictionary. The development of new portal systems and dictionaries will be useful in medical education and research activities.

Keywords: Thesaurus Tree, Synonyms, Dictionary, Associative Search

1. はじめに

インターネットを介して得られる医療健康情報は、一般市民だけでなく医療従事者および医療系学生にも影響を及ぼしつつある。特に、日本語で記述された解説記事などの増加は、相対的に英語で記述された原著論文など、科学的に評価の高い一次資料が利用されなくなる状況を招いている。しかし、もし英語のリソースを日本語で検索できるポータルが提供されれば、利便性と有益な波及効果が期待できる。そこで本研究では、インターネットで公表される広範な医学関連研究成果を、日本人が検索あるいは理解しやすくすることを目的として、過去10年以上にわたって構築してきたライフサイエンス辞書^{1,2)}の資源を利用するシソーラスの制作と日本語ポータルの研究開発を計画した。

本研究のゴールは、第一に医学研究論文やデータベースを日本語で検索する場合に、表記のゆれや翻訳を実装するとともに、入力したキーワードと密接に関連する別のキーワードを同時に提示することによって情報検索を容易にする連想検索サーバを開発することとした。また第二に、検索結果として表示される英語ページにおいて、利用者が求める箇所でオンデマンドに専門用語の対訳および解説を表示して、利用者の理解を助ける情報ポータルの試作を行った。英語で書かれた医学情報のあらゆるWebページを日本語で検

索して内容を理解できるサーバを無料で公開することで、医療および医学研究の成果を広く社会に提供する実用的なインターフェースとして幅広い利用や応用が見込める。

2. 方法

2.1 シノニム辞書の制作

医療情報の理解に必要と考えられる専門用語の異表記を統一するための統制語としては、後で情報検索に利用することを考えた結果、まずはアメリカ National Library of Medicine が構築している Medical Subject Headings (MeSH) に準拠することにした。そこで MeSH 2008 (2007年11月版) より解剖部位 (Tree A01-A17)、生物名 (Tree B01-B08)、病名および症候名 (Tree C01-C23 および精神疾患 F03)、生体分子および医薬品名 (Tree D01-D27 および Supplemental Concepts)、方法および尺度 (Tree E01-E07)、学問領域や現象 (Tree G01-G14) に帰属する専門用語から、上記のカテゴリに帰属できる Descriptor 21,684語と LSD に同一の見出し語が収録されていた Supplemental Concepts 2,668語を合わせた計 24,352語を統制語として採用した³⁾。これを元に、LSD と MeSH のすり合わせ作業を行い、シノニム辞書を制作した。また、一部の統制語においては、MeSH に準拠せず、出現頻度

の高い語句を採用した。

2.2 共起する統制語による関連概念データの制作

PubMed抄録を収集した文献コーパスに対して、シノニム辞書を適用して統制語によるタグ付けを行うPerlスクリプトを開発した。統制語タグが同一抄録中で共起する頻度を解析し、各用語について出現頻度、共起する他の統制語およびその共起頻度を得た。得られたデータが専門的に見て妥当な関連性を表すかどうかを、複数名の研究者による目視によって検討した。この評価に基づいて、検索キーワードの取捨選択を行い、最適化を試みた。

2.3 関連概念を提示する情報検索エンジンの開発

シノニムと共起解析データをオンライン版ライフサイエンス辞書WebLSDに実装することによって、日本語および英語のいずれによっても表記のゆれを吸収して統制語による情報検索を可能にするポータルシステムをPerl cgiにて開発した。

2.4 日本語訳を表示する辞書ツールの開発

ウェブブラウザで表示されるゲノム情報などの英語ページにおいて、可能な限り簡単な操作で専門用語を辞書引きできるツールを開発するため、Mac OS X 10.5においてシステム標準で利用できる辞書.appでの試作と検証を試みた。この辞書.appではブラウザであるSafariからショートカットで複合語レベルでの辞書検索が実現できる。また、辞書を制作するためのアプリケーションやテキスト使用がAppleにおいて公開されている。

3. 結果および考察

3.1 シノニム辞書の制作

2008年6月時点で、表1に示すカテゴリのMeSH DescriptorおよびSupplemental Concepts (SC)に帰属する統制語2.5万語の96%を日本語化し、英語表記と日本語表記を併記できるようにした。その上で、延べ約16万語の英語および日本語で記述されるLSD収録語およびMeSH用語を統制語に集約することで、対訳シノニムとシノニム辞書を制作した。このデータから、16万語の同義語のうち、LSD収録の英語と日本語、および新たに加えたMeSH英語が、それぞれほぼ3分の1ずつの割合を占めることがわかる。生体分子などの物質名、特に海外での医薬品商品名や化学一般名などの異表記を非常に数多く含む物質カテゴリにおいては、MeSHに由来する名称が半数に及び、これら新しく加えた用語によって欧米の文書に対する網羅性が高まったことが期待できる。

表1 ライフサイエンス辞書のシノニム化(概要)

MeSH Tree	カテゴリ	統制語数(a)	シノニム数(b)	平均英訳数 (b/a)	LSD英語	LSD日本語	MeSH物名
A	解剖部位	1,521	7,022	4.6	3,205	475	3,680
B	生物	3,419	18,439	5.4	3,157	374	6,670
C-D-E	病名・症候名	4,323	28,821	6.7	3,219	205	11,892
F	物質名	11,200	91,080	8.1	30,571	326	29,878
G	(化学)医薬品	2,863	42,307	14.8	2,380	184	11,072
H	方法・尺度	2,185	11,871	5.4	3,853	395	4,876
I	組織・構造	1,378	7,233	5.3	2,784	374	3,204
J	動物	24,283	134,628	5.6	44,732	326	18,023

百分率はシノニム数(b)に対する割合を表す。

しかしながら一方で、LSDに収録されながらMeSH

と照合できないため統制語に帰属されない用語が英語で1万語以上も存在することが明らかになった。特に、病名・症候名や解剖部位名においては、国内で用いられている標準病名マスターや国際的な有害事象報告のための統制語であるMedDRAにも収録されながらMeSHに帰属できない用語が数多く残された。また、医薬品としては国内医薬品において収録されていない用語が多数存在した。今後は、これらの語句を帰属させるためにツリーを拡張していく必要が示された。

3.2 共起する統制語による関連概念データの制作

PubMedより代表的な学術誌に掲載された10年分の論文抄録(600 Mバイト)をコーパスとして収集し、シノニム辞書によってテキスト中に最長一致で統制語のXMLタグを施した。このタグ付けテキストの内容をブラウザで確認しながら、曖昧性の排除と統制語の最適化を行った(図1)。この過程において、テキストでの一致のみによって統制語への変換を行う場合、曖昧性を排除するために多義性のある略語や商品名等、一部のシノニムをタグ付け辞書から除外する必要が生じた(約200語)。また、「ヒト human」、「病気 disease」、「酸 acid」等のように、非常に大きな概念は関連するキーワードとして不必要あるいは不適切と考えられたため、それら(約360語)もタグ付けから除外した。



図1 英語論文抄録にタグ付けを施したXMLデータをブラウザで表示した例

テキスト中出现する専門用語をすべて統制語に自動変換するPerlスクリプトを用いてXMLデータを作成した。日本語の統制語見出しを、物質や医薬品名は青色で、病名は赤色で、方法や尺度は緑色で表すことによって抄録で述べている内容に関連するキーワードの関係や、統制語の妥当性を判定できるようにした。この抄録の場合、「ベンゾジアゼピン」と「股関節骨折」の関係を述べている論文であることが一見してわかる。しかし「recipient」を「移植」という統制語に翻訳した箇所は誤りであるため、このように対応関係は解析辞書から除外する措置をとることによって最適化を行った。

次に、同一抄録中で共起する統制語のペアを収集することによって計100万対以上の統制語の共起頻度を求め、出現した2万語の統制語ごとに上位30対までの共起概念データを得た。ここで解析に用いるコーパスによって得られる共起概念のリストは大きく異なった。例えば、1.3 Gバイトの臨床研究抄録を用いた解析では、ある薬物と共起する概念はほとんど医薬品

名で占められ、標的となる生体分子や作用メカニズムを示唆するキーワードが得られない等、必ずしもコーパスが大きいからと言ってデータが適切にならない場合も見いだされた。

本研究は医療系学部あるいは大学院に所属する学生による教育現場での利用を想定していたため、上述した代表的学術誌にこの10年間に発表された先端的研究成果を記述する広範な学術分野の論文抄録に限ることとした。試行錯誤の末、主観的に見てほぼ全ての用語カテゴリーにわたってバランス良い共起結果を得ることができたと考えている。しかし今後、コーパス母集団を変えることでさらに専門家の知識を反映するような最適化を試みる必要があると思われる。



図2 WebLSDに実装したシソーラスと共起概念による医療情報ポータルの使用例

(左上)「どうしよう」と入力したときに表示される和英辞書で「糖尿病」シソーラスをクリック (右上)用語ツリーから下位概念である「2型糖尿病」をクリック (左下)2型糖尿病の共起概念リストから「遺伝子多型 Genetic polymorphism」をクリック (右下)PubMedに「type 2 diabetes mellitus」と「genetic polymorphism」の2つのMeSHが渡されて、ヒットする文献リストが表示される

3.3 関連概念を提示する連想検索エンジンの開発

このようにして得た共起概念データをシソーラスのツリー表やシノニム表示と組み合わせることによって、検索語として入力した日本語あるいは英語を自動的に統制語に直して表示するだけでなく、ツリーによって上位や下位の概念を探索できるようにしたり、関連性の高い共起概念を表示することで既存ポータルに適切なキーワード対を検索語として渡したりするためデータをXML形式で制作した。

これらデータをまずウェブブラウザで検索可能にするため、公開しているWebLSDのサブセットとして、英和・和英対訳辞書と一体で使うことができるような

cgiを制作し、2008年6月より公開した。

このWebLSDに実装したシソーラスを用いることによって、任意に入力する日本語の検索語が英語に訳されるだけでなく、MeSHに準拠した統制語について、シノニム、ツリー、共起概念が表示される(図2)。ツリーでは表示している統制語が赤字で表示され、その上位と下位に位置する概念をクリックで自由に移動することができる。共起概念は日本語と英語で最大30種類がそれぞれの統制語ごとに表示され、日本語をクリックした際には選んだ用語と統制語との組み合わせでGoogleへ検索キーワードが渡される。また、英語のリンクからはEntrez-PubMedにキーワード対が渡されるようにしてある。基本的にはURLを明示できる検索エンジンやデータベースに対して、このインターフェースを介してデータを渡すようにカスタマイズすることは容易にできるため、汎用性や応用性にも優れている。

3.4 日本語訳を表示する辞書ツールの開発

Mac OS X 10.5には標準で辞書ツールである辞書.appが付属している。この辞書.appは日本語にも対応しており、キーワード入力に応じて結果を表示するincremental searchを可能にした特徴を持っている。また、辞書.appはMac OS X標準ウェブブラウザであるSafariからショートカットキー(Command + Control + D)によって呼び起こすことができる。さらにこの時、カーソルが置かれている単語の前後を最長一致で判定し、最もその場所にふさわしい複合語を選び出して表示する他の辞書には見られない機能を有している。Apple社は辞書.appに対応する辞書を制作するための技術資料を公開しているため、今回、この辞書.appを用いる辞書を制作した。

その結果、WebLSDで実装したシソーラスとほぼ同様の機能を有するスタンドアロン辞書を制作することができた(図3)。辞書.appは検索語の途中でも先読みでキーワードを表示するため、前方一致するキーワードリストを見ながら、適切な用語のシソーラスを見ることができる。シソーラス内での操作はほぼWebLSDと同様であり、ツリーの上下移動や共起概念からの外部リンクを装備することができる。また、このようにして制作した辞書はSafariに表示されたhtmlページのカーソル位置からショートカットキーで呼び出すことができるため、英和の用語検索が容易に行える。今後、さらに辞書.appおよびSafariの連携が簡単かつ高機能になることを期待したい。



図3 Mac OS X辞書への実装

Safariで電子ジャーナルを検索した際にショートカットキーで辞書.appを起動した画面

4. おわりに

以上のように、本研究では当初の計画で予想した以上に有用なポータルを開発することができた。今後さら

にデータの最適化を計り、公開ポータルとしての利便性を向上させる予定である。また、提示する共起概念の視覚的な表示技術についても検討していきたい。しかし、教育における本シソーラスの利用経験はまだ浅いため、将来的にこれらを医療情報教育に活用し、客観的な評価を進めたいと考えている。

5. 謝辞

本研究は(財)電気通信普及財団研究調査助成(平成18年度)、厚生労働省科学研究費(平成18-20年度)および(独)日本学術振興会科学研究費研究成果公開促進費(平成17-19年度,177002)の研究助成を受けて行われた。辞書.appについては中村浩之氏から制作するきっかけとなる示唆をいただいた。ここに記して感謝の意を表したい。

参考文献

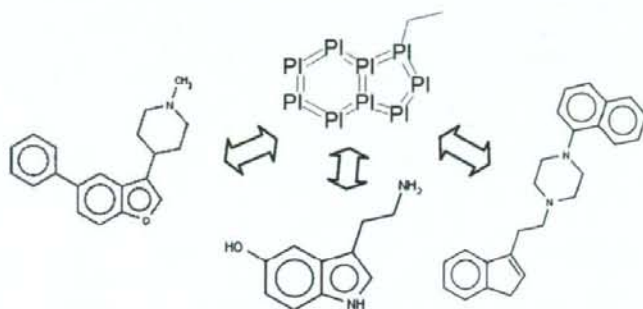
- [1] 金子周司, 鶴川義弘, 大武博, 河本健, 竹内浩昭, 竹腰正隆, 藤田信之. ライフサイエンス辞書2の制作と公開. コンピュータサイエンス, Vol. 2, No. 2, 135-142, 1995.
- [2] 金子周司, 藤田信之. 文献情報の解析に基づく対訳シソーラスの評価. 医療情報学, Vol. 25, No. 6, 475-483, 2005.
- [3] 金子周司. ライフサイエンス辞書とは. 情報管理, Vol. 49, No. 1, 24-35, 2006.
- [4] ライフサイエンス辞書. <http://lsd-project.jp/>.

Substructure Mining of GPCR Ligands Reveals Activity-Class Specific Functional Groups in an Unbiased Manner

Eelke van der Horst, Yasushi Okuno, Andreas Bender, and Adriaan P. IJzerman

J. Chem. Inf. Model., 2009, 49 (2), 348-360 • DOI: 10.1021/ci8003896 • Publication Date (Web): 03 February 2009

Downloaded from <http://pubs.acs.org> on March 2, 2009



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)



ACS Publications
High quality. High impact.

Journal of Chemical Information and Modeling is published by the American Chemical Society, 1155 Sixteenth Street N.W., Washington, DC 20036