

薬理作用と作用機序に基づく化合物分類

MeSH descriptor_ID D018691
 descriptor_string Excitatory Amino Acid Antagonists
 編集したDescriptor **Excitatory Amino Acid Antagonist**
 ツリー日本語訳 **興奮性アミノ酸遮断薬**
 意味抽出辞書 興奮性アミノ酸遮断薬

MeSH tree

2	D27.505.519.625.190.300
	D27.505.696.577.190.300

MeSH Term	7
Antagonists Excitatory Amino Acid	0
EAA Antagonist	10
Excitatory Amino Acid Antagonist	13
Excitatory Amino Acid Antagonist	13
Excitatory Amino Acid Antagonist	13
Glutamate Antagonist	35
Glutamate Receptor Antagonist	120

LSD英語	6
excitatory amino acid antagonist	
excitatory amino acid antagonist	
glutamate antagonist	
glutamate antagonist	
glutamate receptor antagonist	
glutamate receptor antagonist	
excitatory amino acid antagonist	

LSD日本語	7
グルタミン酸遮断薬	0
グルタミン酸受容体遮断薬	0
グルタミン酸拮抗薬	1
グルタミン受容体遮断薬	0
グルタミン受容体拮抗薬	0
興奮性アミノ酸アンタゴニスト	0
興奮性アミノ酸遮断薬	0

薬物・化合物名	
Riluzole	リルゾール
Phenobarbital	フェノバルビタール
Phencyclidine	フェンシクリジン
Memantine	メマンチン
Kynurenic Acid	キヌレン酸
Ketamine	ケタミン
selfotel	
SDZ EAA 494	
NPC 12626	
LY 341495	
LY 235959	
licostinel	
lamotrigine	ラモトリジン
L 701324	
ketobemidone	ケトベミドン
ifenprodil	イフェンプロジル

MeSH descriptor_ID D018663
 descriptor_string Adrenergic Agents
 編集したDescriptor **Adrenergic Agent**
 ツリー日本語訳 **アドレナリン系薬物**
 意味抽出辞書 アドレナリン系薬物

MeSH tree

2	D27.505.519.625.050
	D27.505.696.577.050

MeSH Term	9
Adrenergic Agent	42
Adrenergic Drug	7
Adrenergic Neurohumor Depleter	0
Adrenergic Neuron Agent	0
Adrenergic Neuron Drug	0
Adrenergic Release Inhibitor	0
Adrenergic Synthesis Inhibitor	0
Adrenergics	1
Sympathetic Transmitter Releaser	0

LSD英語	12
adrenergic	
adrenergic	
adrenergic	
adrenergic agent	
adrenergic agent	
adrenergic agonist	
adrenergic agonist	
adrenergic agent	

LSD日本語	4
アドレナリン系薬物	0
アドレナリン作用薬	0
アドレナリン作用薬	0
アドレナリン作用薬	0
アドレナリン作用薬	

薬物・化合物名	
Yohimbine	ヨヒンビン
Xylazine	キシラジン
Xamoterol	キサモテロール
Viloxazine	ビロキサジン
Tyramine	チラミン
Trimipramine	トリミプラミン
tetrahydropalmatine	
DSP 4	

MeSH descriptor_ID D018712
 descriptor_string Analgesics, Non-Narcotic
 編集したDescriptor **Non-Narcotic Analgesic**
 ツリー日本語訳 **非麻薬性鎮痛薬**
 意味抽出辞書 非麻薬性鎮痛薬

MeSH tree

2	D27.505.696.663.850.014.040
	D27.505.954.427.040.100

MeSH Term	5
Antipyretics	42
Non-Narcotic Analgesic	1
Nonnarcotic Analgesic	3
Nonopioid Analgesic	14
Nonopioid Analgesic	14

LSD英語	7
antipyretic	
antipyretics	
antipyretics	
non-narcotic analgesic	
non-narcotic analgesic	
nonopioid analgesic	
nonopioid analgesic	
non-narcotic analgesic	

LSD日本語	5
解熱剤	1
解熱薬	31
非オピオイド性鎮痛薬	0
非オピオイド鎮痛薬	1
非麻薬性鎮痛薬	7
解熱薬	
非オピオイド性鎮痛薬	

薬物・化合物名	
Tolmetin	トルメチン
Tetrahydrocannabinol	テトラヒドロカンナビノー
Suprofen	スプロフェン
Sulindac	スリンダク
Sulfasalazine	スルファサラジン
Sodium Salicylate	サリチル酸ナトリウム
ziconotide	ジコノチド
tetrahydropalmatine	
samarium	
lappaconitine	
epibatidine	
E 2078	
dixyrazine	ジキシラジン
caffeine benzoate	

薬理作用と作用機序に基づく化合物分類

MeSH descriptor_ID D018680
 descriptor_string Cholinergic Antagonists
 編集したDescriptor **Cholinergic Antagonist**
 ツリー日本語訳 **抗コリン薬**
 意味抽出辞書 抗コリン薬

MeSH tree

2	D27.505.519.625.120.200
	D27.505.696.577.120.200

MeSH Term	8	LSD英語	24	LSD日本語	5	薬物・化合物名
Acetylcholine Antagonist	0	anticholinergic	抗コリン剤	0		Vecuronium Bromide
Anti-Cholinergic	2	anticholinergic	抗コリン薬	81		Tubocurarine ツボクラリン
Anticholinergic Agent	42	anticholinergic	抗コリン薬	81		Tropicamide トロピカミド
Anticholinergics	59	anticholinergic	抗コリン薬	81		Trimethaphan トリメタファン
Cholinergic Antagonist	36	anticholinergic agent	抗コリン薬	81		Trihexyphenidyl トリヘキシフェニジル
Cholinergic Antagonist	36	anticholinergic agent				Toxiferine
Cholinergic-Blocking Agent	0	anticholinergic agent				tiotropium チオトロピウム
Cholinolytic	0	anticholinergic	抗コリン			NS 21
		anticholinergic agent	抗コリン剤			desloratadine デスロラタジン
		anticholinergic drug	抗コリン作用薬			
		anticholinergics	抗コリン性			
		cholinergic antagonist				
		cholinolytic				
		cholinolytic drug				

MeSH descriptor_ID D019161
 descriptor_string Hydroxymethylglutaryl-CoA Reductase
 編集したDescriptor **Hydroxymethylglutaryl-CoA Reductase**
 ツリー日本語訳 **HMG-CoA還元酵素阻害薬**
 意味抽出辞書 HMG-CoA還元酵素阻害薬

MeSH tree

3	D27.505.519.186.071.202.370
	D27.505.519.389.370
	D27.505.954.230.202.370

MeSH Term	6	LSD英語	4	LSD日本語	5	薬物・化合物名
HMG-CoA Reductase Inhibitor	92	HMG-CoA reductase inhibitor	HMG-CoA還元酵素阻害剤	1		Simvastatin シンバスタチン
HMG-CoA Reductase Inhibitor	92	hydroxymethylglutaryl-CoA reductase	HMG-CoA還元酵素阻害薬	8		Pravastatin プラバスタチン
HMG-CoA Statin	0	statin	HMG-CoA還元酵素阻害薬	8		Meglutol メグルトール
Hydroxymethylglutaryl-CoA Inhibitor	0	statin	ヒドロキシメチルグルタリルコエンザイム	0		Lovastatin ロバスタチン
Hydroxymethylglutaryl-CoA	4		ヒドロキシメチルグルタリルコエンザイム	0		
Hydroxymethylglutaryl-Coenzyme A	0					rosuvastatin ロスバスタチン
						fluvastatin フルバスタチン
						compactin
		HMG-CoA reductase inhibitor	HMG-CoA還元酵素阻害剤			cerivastatin セリバスタチン
		hydroxymethylglutaryl-CoA reductase	スタチン			atorvastatin アトルバスタチン
		statin	スタチン系薬剤			
			ヒドロキシメチルグルタリルコエンザイムA還元酵			
			ヒドロキシメチルグルタリルコエンザイムA還元酵			

MeSH descriptor_ID D018926
 descriptor_string Anti-Allergic Agents
 編集したDescriptor **Anti-Allergic Agent**
 ツリー日本語訳 **抗アレルギー薬**
 意味抽出辞書 抗アレルギー薬

MeSH tree

1	D27.505.954.016

MeSH Term	7	LSD英語	9	LSD日本語	4	薬物・化合物名
Anti-Allergic	1	anti-allergic agent	抗アレルギー剤	0		Triprolidine トリプロリジン
Anti-Allergic Agent	0	anti-allergic agent	抗アレルギー薬	42		Tripelennamine トリベレナミン
Anti-Allergy Drug	0	antiallergic	抗アレルギー薬	42		Terfenadine テルフェナジン
Antiallergic Agent	1	antiallergic	抗アレルギー薬	42		Pyrilamine ピリラミン
Antiallergic Drug	2	antiallergic				Promethazine プロメタジン
Antiallergics	0	antiallergic drug				Pheniramine フェニラミン
Antiallergy Agent	0	antiallergic drug				tranilast トラニラスト
		anti-allergic agent	抗アレルギー			tixocortol pivalate
		antiallergic	抗アレルギー剤			suplatast tosilate トシル酸スプラタスト
		antiallergic drug	抗アレルギー性			oxatomide オキサトミド
		antiallergics				olopatadine オロパタジン
						N-acetyl-1-
						imometasone furoate
						loteprednol etabonate
						lodoxamide ethyl
						honokiol

薬理作用と作用機序に基づく化合物分類

MeSH descriptor_ID D020533
 descriptor_string Angiogenesis Inhibitors
 編集したDescriptor **Angiogenesis Inhibitor**
 ツリー日本語訳 **血管新生抑制薬**
 意味抽出辞書 血管新生抑制薬

MeSH tree

3	D27.505.696.377.077.099
	D27.505.696.377.450.100
	D27.505.954.248.025

MeSH Term	16
Angiogenesis Factor Inhibitor	0
Angiogenesis Inhibitor	147
Angiogenesis Inhibitor	147
Angiogenetic Antagonist	0
Angiogenetic Inhibitor	1
Angiogenetic Inhibitor	1
Angiogenic Antagonist	2
Angiogenic Antagonist	2
Angiogenic Inhibitor	26
Angiogenic Inhibitor	26
Angiostatic Agent	3
Anti-Angiogenetic Agent	0

LSD英語	8	LSD日本語	10
angiogenesis inhibitor		血管新生阻害剤	72
angiogenesis inhibitor		血管新生阻害物質	7
angiogenic inhibitor		血管新生阻害薬	0
angiogenic inhibitor		血管新生抑制因子	68
anti-angiogenic agent		血管新生抑制剤	5
anti-angiogenic agent		血管新生抑制剤	5
antiangiogenic agent		血管新生抑制薬	0
angiogenesis inhibitor		血管新生阻害剤	
angiogenic inhibitor		血管新生阻害物質	
antiangiogenic agent		血管新生阻害薬	
		血管新生抑制因子	
		血管新生抑制剤	
		抗血管新生薬	

薬物・化合物名	
Tumor Necrosis Factor	
Thalidomide	サリドマイド
Recombinant Interferon	
Interleukin-12	インターロイキン12
Interferon-beta	インターフェロンβ
Interferon-alpha	インターフェロンα
zhengguangmycin	
thymogen	
tetrathiomolybdate	
tetra(4-N-	
sunitinib	スニチニブ
SU 5416	
squalamine	スクアラミン
roquinimex	ロキニメックス
O-	
LECT1 protein, human	

MeSH descriptor_ID D043925
 descriptor_string Angiogenesis Inducing Agents
 編集したDescriptor **Angiogenesis Inducing Agent**
 ツリー日本語訳 **血管新生促進薬**
 意味抽出辞書 血管新生促進薬

MeSH tree

1	D27.505.696.377.077.077
---	-------------------------

MeSH Term	7
Angiogenesis Factor	30
Angiogenesis Inducer	4
Angiogenesis Inducing Agent	0
Angiogenesis Stimulating Agent	0
Angiogenesis Stimulator	6
Angiogenic Factor	319
Tumor Angiogenic Factor	1

LSD英語	6	LSD日本語	4
angiogenesis factor		血管形成因子	3
angiogenesis factor		血管新生因子	114
angiogenesis inducing agent		血管新生促進剤	0
angiogenesis inducing agent		血管新生促進薬	0
angiogenic factor			
angiogenic factor			
angiogenesis inducing agent		血管形成因子	
		血管新生因子	
		血管新生促進剤	

薬物・化合物名	
platelet-derived growth	
angiogenin	アンジオゲニン

MeSH descriptor_ID D044966
 descriptor_string Anti-Retroviral Agents
 編集したDescriptor **Anti-Retroviral Agent**
 ツリー日本語訳 **抗レトロウイルス薬**
 意味抽出辞書 抗レトロウイルス薬

MeSH tree

1	D27.505.954.122.388.077
---	-------------------------

MeSH Term	2
Anti-Retroviral Agent	4
Antiretroviral Agent	91

LSD英語	6	LSD日本語	3
anti-retroviral agent		抗レトロウイルス剤	0
antiretroviral		抗レトロウイルス薬	0
antiretroviral		抗レトロウイルス薬	0
antiretroviral agent			
antiretrovirals			
antiretrovirals			
anti-retroviral agent		抗レトロウイルス	
antiretroviral		抗レトロウイルス剤	
antiretroviral agent			
antiretrovirals			

薬物・化合物名	
Zidovudine	ジドブジン
Zalcitabine	ザルシタピン
Trichosanthin	
Stavudine	スタブジン
Saquinavir	サキナビル
Ritonavir	リトナビル

厚生労働科学研究費補助金（医療安全・医療技術評価総合研究事業）
分担研究報告書

医薬品安全性に関する文献情報

自動抽出システムの考案

研究協力者 天野博夫 京都大学大学院薬学研究科生体機能解析学分野

協力：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）

[研究要旨]

医薬品安全性監視制度の多重化へ向けた取り組みの一環として、1) 医薬品一般名（英名）による検索および適応・有害反応の日本語による表記と検索が可能な医薬品副作用データベース 2) PubMed アブストラクトからテキストマイニングの手法により医薬品の安全性に関する記載を HTML ファイルに選別・抽出して Web ブラウザ上で色別表示によりヒットした用語の認識性を改善する Perl スクリプトの両者を主要なコンポーネントとする医薬品安全性監視支援システムを構築した。本システムの開発は、副作用シグナルの早期検知あるいは副作用予測用の情報抽出に大きな寄与しうるものと期待される。

A. 研究目的

近年、医薬品の安全性は国内外を問わず、全ての医療関係者、患者およびその家族にとって重要な関心事項となっている。例えば、2005 年の米国における処方薬の副作用（有害反応）による死者数は交通事故によるそれを上回り、死因の第 6 位にランクされている。承認前臨床試験による医薬品安全性評価は、試験期間と被験者数の点で副作用の検出力に限界があることは広く認識されており、客観的かつ実効性のある市販後医薬品安全性監視制度の整備は、規制当局のみならず、製薬企業にとってもリスクマネジメント戦略としてきわめて重要な課題である。一方で、医療の現場と規制当局間の安全性情報伝達は、その主要部分を製薬企業に依存せざるを得ない点に構造的な危険性が存在するため、透明性の高い監視体制

を構築すると同時に、フェイルセーフとして医薬品安全性監視システムの多重化に早急に取り組む必要がある。本年度の研究においては、上記の医薬品安全性監視システム多重化へ向けた取り組みのひとつとして、製薬企業の介在が比較的少ない公開されている情報、即ち副作用データベースおよび文献情報をソースとして、副作用シグナルの早期検知あるいは副作用の予測に使用しうる情報を抽出するシステムを開発することを目的とした。

B. 研究方法

情報ソース

Drugs@FDA（米国 FDA）

AERS（米国 FDA）

PubMed（米国 NCBI）

Japan Pharmaceutical Reference (日本製薬工業会)
 EPARs for authorised medicinal products for human
 use (欧州連合 EMEA)
 MedDRA (日米欧医薬品規制調和国際会議)
 国内医薬品名 (日本医薬情報センター)

1. 医薬品名シソーラスの構築

Drugs@FDA、Japan Pharmaceutical Reference、
 EPARs for authorised medicinal products for human
 use 国内医薬品名の各データベースをソースとし
 て、Shell、AWK、Perl スクリプト群による自動処
 理を応用して商品名、別名、一般名 (いずれも英
 名) などの抽出、一般名表記の統一、重複の除去
 を行い、一般名をキーフィールドとする医薬品名
 シソーラスを構築した。

2. AERS 副作用データベースの改良・再構築と 一部日本語対応

FDA のサイトからダウンロードした AERS デー
 タベースに、医薬品一般名、適応和名、有害反応
 和名の各フィールドを付加した上で FileMaker Pro
 のデータベースとして再構築を行った。また、将
 来的な公開を前提として、Web 公開用フォームを
 設計した。

3. TSADR (Text-search System for Adverse Drug Reaction) システムの構築とトレーニング

テキストマイニングの手法を応用した、副作用
 シグナルの早期検知あるいは副作用予測システ
 ム構築の一環として、PubMed アブストラクトコ
 ーパスからの医薬品安全性に関連する情報の選
 別・抽出を補助するシステムを考案した。AERS
 データベースをソースに作成した語彙リストを
 辞書として PubMed アブストラクトから医薬品名
 と有害反応名が同時に記載されているセンテ
 スを html ファイルに抜き出し、Web ブラウザ上
 でヒットした用語の認識性を色別表示により改
 善する Perl スクリプト (TSADR.pl) を作成した。
 適応名の記載が同一のセンテンス内にあれば、こ

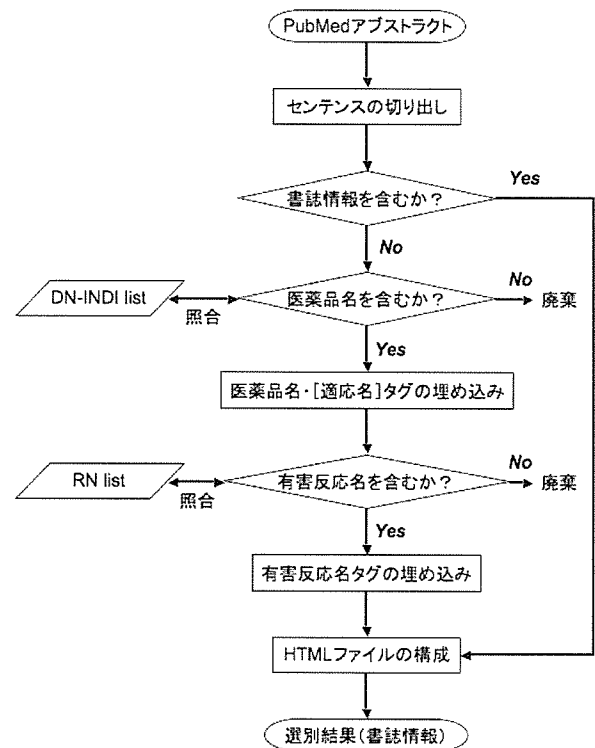


Fig. 1 TSADR 基本動作のフローチャート

れも表示させた。TSADR の基本的な動作の概要
 をフローチャート (Fig. 1) に示す。

PubMed アブストラクト は Limits 設定フォーム
 において only items with abstracts, English, Humans
 の 3 つ制限のみを設定し、キーワードを入力せず
 に取得した 500 件分のテキストを検索対象の 1 単
 位とした。TSADR により抽出されたセンテンス
 を含むアブストラクトを読み、医薬品の副作用の
 記載が正しく抽出・表示されているかを検証した。
 1 単位のテキストサンプルに対する一回のオペレ
 ーションの抽出率 (500 件中何件のアブストラク
 トが抽出されたか) および正解率 (抽出されたア
 ブストラクトの何%に医薬品の安全性に関する
 内容が記載されていたか) をシステムの評価基準
 とした。医薬品安全性情報の選別・抽出が正しく
 行われていない場合には、その原因を解消し、正
 解率を上げる方向の修正を、システムおよび実装
 した語彙リストに加えた (トレーニング)。

4. システムの実効性の検討

上記 AERS データベースおよびこれから派生する語彙リストの拡張、およびトレーニングを繰り返した後（カレントシステム）、TSADR による PubMed アブストラクト解析から得られた副作用シグナルに関して、AERS データベース上での報告例を検索して、システム全体としての実効性を検討した。

C. 研究結果

医薬品名シソーラス (dnamedic.txt) の構築およびアップデート用に作成したスクリプトは以下のものである。

exdiff.pl : Drugs@FDA の新旧 Product.txt データベースの差分抽出。

initiatedic.awk : カレント dnamedic.txt の初期化。

extdnames.awk : 差分ファイルから医薬品名を抽出。

ionsoff.awk : 一般名から対イオン表記を削除。

updatednfda.sh : 上記 4 スクリプト連動用。

checkdouble.awk : 一般名統一表記のチェック用。

addnestindex.awk : 下記 addgeneric.pl に実装時の配列情報を付加する。

AERS (Adverse Event Reporting System) データベースは米国の FDA が収集・構築しているテキストベースの医薬品副作用データベースで、世界各地からの医薬品安全性に関する報告が、自発報告、義務報告を問わず網羅されている。FDA のサイトから 2004 年以降のデータが四半期単位、ほぼ半年遅れでダウンロード出来る。本研究においては、TSADR のシステム構築用に 2004 年および 2005 年上半期の一年半分のレコードをベース語彙リストのソースとして使い、基本システム構築後に 2007 年 3 月現在入手可能な 2004 年から 2006 年第 3 四半期までの全データ（報告件数約 84 万件）に、以下のフィールド項目を付加した上で FileMaker Pro のデータベースとして再構築を行った。

a) 医薬品一般名（英名）フィールドの付加

AERS の医薬品名 (DRUGNAME) フィールドを、上記医薬品名シソーラスを実装した Perl スクリプト (addgeneric.pl) を用いて照合し、新たに一般名フィールドを付加した。現状において、総計約 316 万 8 千レコードの 92.8% の一般名が特定できている。この操作により、多数のジェネリック医薬品のレコード等も、一般名をキーワードとする統一的な検索が可能となった。

b) 適応和名フィールドの付加

c) 有害反応和名フィールドの付加

AERS の適応名 (INDI_PT) および有害反応名 (PT) フィールドは ICH 国際医薬用語集 MedDRA 収載用語 (PT=基本語階層) の使用による標準化が行われているため、日本語対照リスト MedDRA/J を関連フィールドとして表示させた。これにより、これらの項目の日本語表示と検索が可能となった。

本データベースは既に Web 公開が可能な状態にあるが (Fig. 2)、社会的な影響を考慮して、現在、適切な公開方法に関して慎重な検討を行っている。

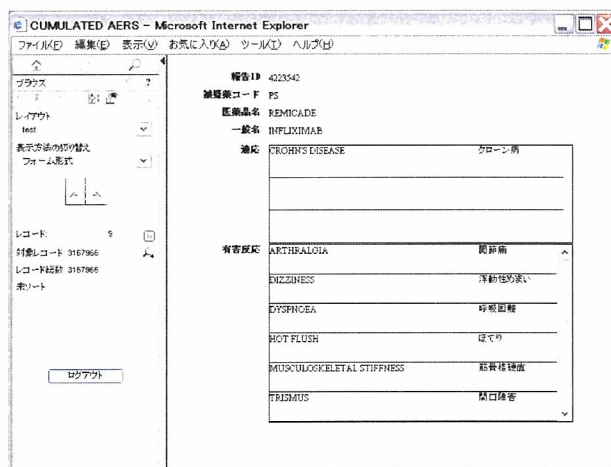


Fig. 2 AERS Web 公開フォームのサンプル

3. 医薬品名/適応対応リストの編纂

上記、AERS データベースから、医薬品名とその適応を対応させた語彙リスト (DN-INDI list) および有害反応名のリスト (RN list) を作成した。

"HAEMORRHAGE"や"ISCHAEMIA"等ラテン表記用語に関しては、文献情報における使用頻度がより高い英語表記のレコードを加えた。

初期システムにおいては、3813 の医薬品名 (drug name)、3824 の適応 (indication) を DN-INDI list に、9712 の有害反応名 (reaction name) を RN list に収録した。AERS データベースの拡張とトレーニングによる最適化に伴う増減を経て、カレントシステムにおいては医薬品名 6870、適応 5930、有害反応名 12875 がそれぞれ収録されている。

4. TSADR システムの構築

質的なばらつきが大きい大量のテキストソースを網羅的に検索して医薬品の安全性に特化した情報を選別・抽出する作業は通常のキーワード検索では事実上不可能であり、これを実現するために医薬品の副作用情報の収集・解析に最適化した DN-INDI list および RN list を実装した Perl スクリプトを新たに開発した。

初期システムを用いて、日本時間 06 年 5 月 29 日に取得した PubMed アブストラクト 500 件 (11,924 センテンス) より 74 件 (138 センテンス) が抽出された (抽出率 14.8%)。74 件中、有害反応名が正しく表示されていたものが 26 件 (正解率 35.1%)、32 件では医薬品名と直接には無関係な反応がヒットし、16 件は適応症が有害反応として誤って表示されていた。誤った選別のパターンとしては、"glucose"、"oxygen"、金属イオン等の生体成分が医薬品名として拾われて起こる事例や"alcohol"、"antibiotic"、"chemotherapy"等、医薬品分類名に関して誤った選別が起こる例が多く認められた。前者のパターンに関しては、隣接する単語との関連から医薬品名としての取捨を判断するフィルターをスクリプトに加えて対応し、後者のパターンは DN-INDI list の適応エントリー数を増やす手段で対応した。上記対応を施したシステムを、新たに (日本時間 06 年 6 月 14 日) に取得したシステムトレーニング用テキスト

(STTXT) に適用し、その結果を基に辞書ファイルを修正する作業を繰り返した。また、誤って選別されたアブストラクトに癌・腫瘍関係の雑誌のものが多かったことから、癌・腫瘍関係語彙用テキスト (CTBTXT: PubMed から Subsets の Limit に Cancer を設定して取得した) を用いたトレーニングも行った。これらのトレーニングによって最適化された語彙リストを用いて、両トレーニングテキスト自身を検索した最終的な成績 (正解件数/抽出件数) は STTXT が 35/62 (正解率 56.5%)、CTBTXT が 42/64 (正解率 65.6%) であった。必要な場合にはスクリプトの修正も行った。

TSADR の実用化に向けた、上記トレーニングの有効性を検討する目的で、新たに取得したテキスト (日本時間 06 年 8 月 1 日) をトレーニング前のシステム TSADR-original とトレーニング後のシステム TSADR-trained を用いて解析した。TSADR-original、TSADR-trained それぞれの成績 (正解件数/抽出件数) は 14/44 (正解率 31.8%) および 22/54 (正解率 40.7%) と算出され、抽出率、正解率ともにトレーニングの有効性が認められた。

一方、選別されるべきセンテンスの拾い漏れがどの程度起こっているかの予備的検討として、医薬品文献情報の有力サイトである英国の National electronic Library for Medicines にピックアップされた文献 (発行期日 06 年 8 月 1 日~18 日) のうち副作用情報に分類される 15 レコードを対象に TSADR による重要文献の抽出漏れを検討した。15 件中 3 件にはアブストラクトが無かった。残り 12 件のアブストラクトを PubMed より取得し TSADR-trained により解析すると、3 件が抽出されなかった。抽出漏れの原因は、いずれも DN-INDI list に医薬品名が収録されていなかったためで、1 件は第一層臨床試験における開発コード、2 件は治療方法の一般呼称で記載されていた。

Drug Names Reactions Indications Journal Information

Perphenazine-treated patients had a higher incidence of extrapyramidal symptom-related adverse events, mean increases (i.e., worsening) in extrapyramidal symptom rating scale scores, and a higher rate of elevated prolactin levels than aripiprazole (57.7% vs. 4.4%, $p < .001$).
 PMID: 17335319 = J Clin Psychiatry. 2007 Feb;68(2):213-23.

We also determined whether the strength of antipsychotic or combination trials was associated with age, the duration of the current depressive episode, medical burden, cognitive status, or the severity of depressive or psychotic symptoms.
 PMID: 17335316 = J Clin Psychiatry. 2007 Feb;68(2):194-200.

Bloodstream infections among patients treated with intravenous epoprostenol or intravenous treprostinil for pulmonary arterial hypertension—seven sites, United States, 2003-2006.
 In September 2006, CDC received a report from a PAH specialist of a suspected increase in the number of gram-negative bloodstream infections (BSIs) among PAH patients treated with IV treprostinil.
 The results do not suggest intrinsic contamination of IV treprostinil as a cause of the infections; the difference in rates might have been caused by differences in preparation and storage of the two agents, differences in catheter care practices, or differences in the anti-inflammatory activity of the agents.
 PMID: 17332729 = MMWR Morb Mortal Wkly Rep. 2007 Mar 2;56(8):170-2.

Fig. 3 TSADR による解析結果の一例（部分）

システムのトータルな実効性に関する検討の一例として、(日本時間 07 年 3 月 22 日) に取得した PubMed アブストラクトを TSADR のカレントシステムで解析して得られた副作用シグナル「Perphenazine 投与に関連した錐体外路症候群」(Fig. 3 参照) を AERS データベースで検索にかけた。「錐体外路症候群」は MedDRA 用語ではあるが、AERS で使用を認められている PT (主要語) の下の階層の用語であるため、この組み合わせではヒットがなかったが、有害反応のキーワードを「錐体外路」とすると、部分一致により「錐体外路障害」を含む 4 件の報告がヒットした。さらに、「錐体外路障害」に連想される「振戦」、「筋固縮」、「ジスキネジー」、「筋骨格硬直」等、運動機能システムの有害事象に広げると合計 34 件の報告がヒットした。

D. 考察

Fig. 4 は、ヒトに関する PubMed 英文アイテムの年間エントリー数の推移を示したものである。年を追ってエントリー数の増加が認められるが、2005 年のデータを参考にする、このコーパスから網羅性を重視して必要なアイテムを選別していくには、一日平均 800 ないし 1000 件を処理する必要があり、その実行には大きな組織、または自動化による補助が必須であると考えられる。

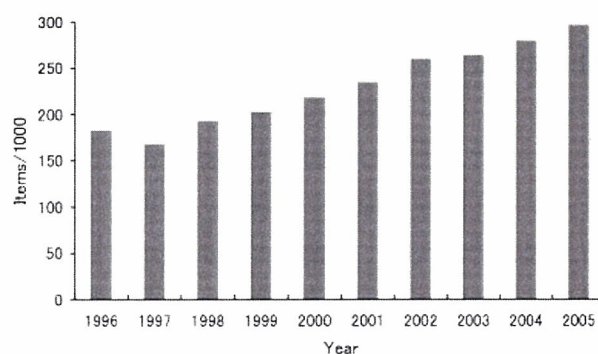


Fig. 3 PubMed 臨床医学系アイテムの年間エントリー数

本研究において医薬品安全性に関連する語彙のソース、および TSADR による副作用シグナルの実効性の検討に用いた AERS データベースは、米国 FDA が収集・管理している、製薬企業からの義務報告と医療従事者・患者およびその家族からの自発報告を総合した巨大な副作用データベースであり、四半期分 (掲載報告件数 8~9 万) ごとに半年遅れの生データが ASCII または SGML ファイルとして FDA のサイトから入手できる。

TSADR システムにおいては、医薬品の副作用が記載されているセンテンス中には医薬品名と有害反応・症状名が共起するという基本原則に従ってセンテンスの抽出が行われる。従って、いかに網羅的なデータ抽出が実行できるか (抽出率) は、実装された DN-INDI list および RN list の充実に依存する。一方、この原則に合致するセンテンスには、医薬品副作用以外にも、医薬品の適応をはじめ、生体成分でもある医薬品 (ホルモン等)

に関しては副作用のリスクファクターおよび臨床検査成績等の記載されたものもあり、原則的にはこれらも同時に抽出されてくる。従って、システム完成度の指標となる、いかに正しい抽出が行われたか（正解率）は医薬品副作用のシグナルをいかにしてこれらのノイズと分離するかに依存する。システムの構造上、語彙リストの医薬品名、有害反応名のレコード数を増やせば抽出率は上昇し、検索の網羅性に関しては有利に働く一方で、副作用以外の医薬品名と有害反応名の組み合わせを拾う可能性も高くなり、正解率が下がれば人間による最終的な選別操作の負担が大きくなる。これに対して、医薬品を投与する原因となる病態などの名称、すなわち適応名の語彙を増やすことは、誤った選別を抑制し、検索の精度を上昇させる。本研究において語彙リストの基本ソースとして用いた AERS データベースは、リレーショナルデータベースの構造を持つため、医薬品名と適応名を対応させて取得できる大きなメリットがある一方で、MedDRA に準拠した用語標準化によるある種の方言的な偏りが文献情報中のより多様な用語に対するヒット率を下げている可能性は否定できない。また、本研究において再構築した AERS データベースは、TSADR システムによるテキストマイニングの成績の検証ツールとしてもある程度の有用性は示せたが、MedDRA の階層構造を活用した柔軟な検索システムの開発が不可欠であることも明らかになった。

今後の研究においては、本検索システムの網羅性および選別性をさらに向上させるために、LSD シソーラスの実装を含めて、語彙リストの補強・改良を自動化する方法を検討する予定である。

E. 結論

本研究において開発された医薬品安全性監視支援システムは、副作用シグナルの早期検知あるいは副作用の予測用の情報抽出に大きな寄与をしようものと期待される。

F. 研究発表

1. 論文発表

1. 天野博夫, 金子周司, 医薬品安全性に関する文献情報自動抽出システムの考案, 医療情報学, 26 (Suppl.), 1193-1194 (2006)

2. 学会発表

1. 天野博夫, 金子周司. 医薬品安全性に関する文献情報自動抽出システムの考案. 第 26 回医療情報学連合大会 (札幌, 2006 年 11 月)

G. 知的財産権の出願・登録状況 (予定も含む)

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

厚生労働科学研究費補助金（医療安全・医療技術評価総合研究事業）
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・統合薬学フロンティア教育センター）

[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。

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」、「薬物の主作用点データベース（GPCRーリガンド相互作用データベース）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、テキストマイニング用の基幹システムとして、XMLデータベースの構築と、JAPIC 添付文書記載病名集のデータベース化を完了した。また、「薬物の主作用点データベース（GPCRーリガンド相互作用データベース）の開発」としては、医薬品の薬効や副作用の総合的な解析のインフォマティクス基盤として、薬物とタンパク質との相互作用データベース GLIDA を開発し、<http://pharminfo.pharm.kyoto-u.ac.jp/services/glida> より公開した。次年度以降は、日英の様々な医薬、医療文書のテキストマイニングエンジンの開発へと発展させると共に、包括的な薬物作用点データベースの開発にも着手する。

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開発素材として、(財)日本医薬情報センター（JAPIC）の添付文書記載病名集を用いた。JAPIC 添付文書記載病名集は、医薬品の薬効や副作用情報など本研究対象に必要な情報が記載されており、さらに XML 形式での電子データが供給されている。具体的には、XML データベース NeoCore を用いて、JAPIC 添付文書記載病名集からの用語の自動抽出、構造化を行い、JAPIC 添付文書記載病名集を包含するテキストマイニング用データベースの構築を行った。またテキストマイニングエンジンの開発としては、現在、全文検索アルゴリズムを実装中である。

2. 薬物の主作用点データベース (GPCR-リガンド相互作用データベース) の開発

医薬品による副作用の総合的な解析には、薬物の作用点となる標的タンパク質との相互作用様式を情報学的に処理する基盤技術の整備が必須となる。本研究では、薬物の主作用点データベースとして GPCR-リガンド相互作用データベース (GLIDA データベース) を構築し、公開している。市販医薬品の大半は、G タンパク質共役型受容体 (GPCR) を薬物作用点にしていることから、本データベースがプロトタイプになり得る。GLIDA は、GPCR のバイオ情報、リガンドのケミカル情報、および GPCR とリガンドの相互作用情報の 3 種類の情報より構成される。GPCR のエントリーはヒト、ラット、マウスに限定し、バイオ情報は GPCRDB から取得した。また、GPCR と結合するリガンドのエントリーとそのケミカルデータ (化学名、構造式、分子量、MDL Mol ファイルなど) は IUPHAR Receptor Database, PubMed, PubChem および MDL ISIS/Base 2.5 などの公共または商用のデータベースから取得した。GLIDA は LAMP (Linux, Apache, MySQL & PHP) プラットフォームで制作され、現在 Web 公開を行っている (<http://pharminfo.pharm.kyoto-u.ac.jp/services/glida>)

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

本年度は、テキストマイニング用データベースとして、JAPIC 添付文書記載病名集のデータベース化を完了した。

様々な医療情報、臨床データなどに柔軟に対応し、用語の自動抽出、データ構造化を可能にする基幹システムとして、XML データベース NeoCore の

実装を行った。XML データベースは、ツリー形式のデータ構造化を実現し、階層的な分類によるスマートな文書管理、検索の高速化にすぐれている。さらに、多種多様な医療情報に対し、データ構造を柔軟に編集できる頑強性においても突出した特長を有する。

また、データベースに登録する文書データとしては、JAPIC 添付文書記載病名集を用いた。JAPIC 添付文書記載病名集は XML 形式で供給されており、データベース構築が比較的容易であるという形式上の利点を有するほか、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おり、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であると考えられる。

2. 薬物の主作用点データベース (GPCR-リガンド相互作用データベース) の開発

本研究で開発した薬物と GPCR の相互作用データベースである GLIDA は、<http://pharminfo.pharm.kyoto-u.ac.jp/services/glida> より公開している。各エントリーの検索は、GPCR (又はリガンド) のキーワード検索およびクラス分類テーブルから行うことが可能である。ここで、GPCR 分類は、GPCRDB に定義されている進化系統樹由来の分類に従った。またリガンド分類は、KEGG で定義されている原子タイプの原子数/結合数に基づいた頻度プロファイルから距離行列を計算し、階層型クラスタリングを行った GLIDA 独自の分類木を作成した。検索された各 GPCR (又はリガンド) のページには、バイオ情報 (又はケミカル情報)、及びそれらに結合するリガンド (又は GPCR) のリストが同時に表示される。さらに、GLIDA の GPCR (又はリガンド) のページは GPCR-リガンド相互作用の解析機能を有している。すなわち、検索された GPCR (又はリガンド) と最も高い類似性を持つ 25 個の GPCR (又はリガンド) リストを表示するとともに、これら 25 個の

エントリと結合するリガンド（又は GPCR）との相互作用様式を 2 次元マップ表示する。このマップの 2 軸に並ぶ GPCR とリガンドの順番は、各々 GPCR とリガンドのクラスタリング結果を反映している。したがって、GPCR、リガンドの類似性情報と相互作用情報を同時に視覚化し、このパターンを分析して GPCR とリガンドの相互作用予測を実現し、薬物と作用基点の相互作用に関する情報を得ることができる。

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

本研究において、実装した XML データベースは、非常に汎用性に富んだシステムである。従って、本年度登録データとして用いた JAPIC 添付文書記載病名集はデータの一例であり、今後他の文書データを登録し、総合的なテキストマイニングを展開する。特に、今回は XML 形式のデータ入力を行ったため、比較的データベース構築が容易であった。次年度以降は、テキスト形式を XML 形式に自動変換することにより、多種多様な文書データに対するデータベース化を試みる。

また代表研究者が開発してきたライフサイエンス辞書との連携を図ることにより、日本語と英語の自動相互変換を行い、PubMed 情報の取り込みへと拡張する。これにより、日英の医薬、医療文書のテキストマイニングエンジンの開発へと発展させる。

2. 薬物の主作用点データベース (GPCR-リガンド相互作用データベース) の開発

GLIDA データベースは市販の医薬品の半分以上の標的分子となっている GPCR とそれに作用する薬物の相互作用に関する知識データベースであるとともに、その相互作用メカニズムの解明に関

する知識を提供し得るケミカルゲノミクスのためのデータベースである。医薬品の薬効や副作用の総合的な解析には、薬物の作用基点となる遺伝子との相互作用様式を情報学的に処理する基盤技術の整備は必須であり、本データベース構築によりその基盤は確立された。今後、GPCR タンパク質以外の他の標的タンパク質の情報を集積したデータベースへの拡張を行うほか、上述、テキストマイニングデータベースとの連携を図る。

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

テキストマイニング用の基幹システムとして、XML データベースの構築と、JAPIC 添付文書記載病名集のデータベース化を完了した。

2. 薬物の主作用点データベース (GPCR-リガンド相互作用データベース) の開発

医薬品の薬効や副作用の総合的な解析のインフォマティクス基盤として、薬物とタンパク質との相互作用データベース GLIDA を開発し、<http://pharminfo.pharm.kyoto-u.ac.jp/services/glida> より公開した。

F. 研究発表

1. 論文発表

1. Naito, Y., Takematsu, H., Koyama, S., Miyake, S., Yamamoto, H., Fujinawa, R., Sugai, M., Okuno, Y., Tsujimoto, G., Yamaji, T., Hashimoto, Y., Itohara, S., Kawasaki, T., Suzuki, A., Kozutsumi, Y., "Germinal center marker GL7 probes activation-dependent repression of N-glycolylneuraminic acid, a sialic acid species involved in the negative modulation of B cell

- activation.” **Mol. Cell Biol.**, 27(8), 3008-22, 2007.
2. Osada, S., Naganawa, A., Misonou, M., Tsuchiya, S., Tamba, S., Okuno, Y., Nishikawa, J., Satoh, K., Imagawa, I., Tsujimoto, G., Sugimoto, Y., and Nishihara, T. “Altered gene expression of transcriptional regulatory factors in tumor marker-positive cells during chemically induced hepatocarcinogenesis.” **Toxicology Letters**, 167, 106-113, 2006
 3. Tsuchiya, S., Okuno, Y., Tsujimoto, G. “MicroRNA: biogenetic and functional mechanisms and involvements in cell differentiation and cancer.”, **J. Pharmacol. Sci.**, 101(4):267-70, 2006
 4. Okuno, Y., Yang, J., Taneishi, K., Yabuuchi, H., Tsujimoto, G., “GLIDA: GPCR-Ligand database for Chemical Genomic Drug Discovery” **Nucleic Acids Research**, 34, D673-7, 2006
2. 学会発表
1. 日本薬物動態学会 第21回年会「ケミカルゲノミクスからの創薬インフォマティクス」2006年12月1日
 2. 第32回情報処理技術検討交換会「ケミカルゲノミクス情報のデータマイニング」2006年11月30日
 3. モレキュラーライブラリー研究会「ケミカルゲノミクス情報を用いた化合物ライブラリーの合理的設計」2006年11月16日
 4. 第34回構造活性相関シンポジウム奨励講演「ケミカルゲノミクス情報に基づく化合物探索」2006年11月14日
 5. 第269回CBI学会「ケミカルゲノム情報に基づくGPCR創薬」2006年11月1日
 6. 第21回21世紀の薬学を探る京都シンポジウム - 薬学教育フロンティア - 「インフォーマティクスと創薬」2006年10月14日
 7. 第2回バイオメディカル研究会「ケミカルゲノミクスのための創薬インフォマティクス」2006年9月12日
 8. 第4回先端医療セミナー「ケミカルゲノミクスからのIn silico創薬」2006年8月25日
 9. 第45回バイオグリッドビジネスサロン「創薬におけるバイオインフォマティクスの可能性」2006年8月18日
 10. 第16回近畿バイオインダストリー振興会議 技術シーズ公開会「創薬リード化合物自動合成装置の研究開発」2006年7月28日
 11. コンピュータ化学部会 平成18年度例会「ケミカルゲノム情報に基づくIn silico創薬」2006年6月13日
 12. バイオグリッド研究会 - ITプログラム成果報告と今後の展開 - 「化合物空間を利用した化合物検索」2006年5月27日
 13. (財)サントリー生物有機科学研究所 コロキウム「ケミカルゲノミクスのためのインフォマティクス」2006年2月21日
- G. 知的財産権の出願・登録状況（予定も含む）
1. 特許取得
 1. 特願2007-53322、「マイクロRNA標的遺伝子予測装置」、平成19年3月2日出願、出願人 東レ株式会社、発明者 奥野恭史、辻本豪三、国本亮、寺澤和哉、土屋創健、秋山英雄、妙本明
 2. 公開番号WO2007/004479 A1（特開2007-11752）、「データ処理装置、データ処理プログラム、それを格納したコンピュータ読み取り可能な記録媒体、およびデータ処理方法」、平成19年1月1日公開、出願人 京都大学、発明者 奥野恭史、辻本豪三、梁智允、種石慶

3. 特願 2006-147433、「ケミカルゲノム情報に基づく、タンパク質-化合物相互作用の予測と化合物ライブラリーの合理的設計」、平成 18 年 5 月 26 日出願、出願人 京都大学、発明者 奥野恭史、種石慶、辻本豪三

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	ライフサイエンス辞書プロジェクト監修	ライフサイエンス英語類語使い分け辞典	羊土社	日本	2006
奥野恭史ほか	日本バイオインフォマテイクス学会	バイオインフォマテイクス事典	共立出版(株)	日本	2006
奥野恭史ほか	石渡信一・桂勲・桐野豊・美宅成樹 編	生物物理学ハンドブック	朝倉書店	日本	2007/4/20 刊行予定

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	ライフサイエンス辞書とは	情報管理	49 (1)	24-35	2006
金子周司	無料ライフサイエンス辞書の活用と効能	ファルマシア	42 (5)	463-467	2006
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	25 (6)	475-483	2005
Naito, Y., Takematsu, H., Koyama, S., Miyake, S., Yamamoto, H., Fujinawa, R., Sugai, M., Okuno, Y., Tsujimoto, G., Yamaji, T., Hashimoto, Y., Itohara, S., Kawasaki, T., Suzuki, A., Kozutsumi, Y.	Germinal center marker GL7 probes activation-dependent repression of N-glycolylneuraminic acid, a sialic acid species involved in the negative modulation of B cell activation.	Mol. Cell Biol.		[Epub ahead of print]	2007
Osada, S., Naganawa, A., Misonou, M., Tsuchiya, S., Tamba, S., Okuno, Y., Nishikawa, J., Satoh, K., Imagawa, I., Tsujimoto, G., Sugimoto, Y., Nishihara, T.	Altered gene expression of transcriptional regulatory factors in tumor marker-positive cells during chemically induced hepatocarcinogenesis.	Toxicology Letters	167	106-113	2006

Tsuchiya, S., <u>Okuno, Y.</u> , Tsujimoto, G..	MicroRNA: biogenetic and functional mechanisms and involvements in cell differentiation and cancer.	J Pharmacol Sci	101(4)	267-70	2006
<u>Okuno, Y.</u> , Yang, J., Taneishi, K., Yabuuchi, H., Tsujimoto, G..	GLIDA: GPCR-Ligand database for Chemical Genomic Drug Discovery	Nucleic Acids Research	34	D673-7	2006

研究成果の刊行物・別刷

ライフサイエンス辞書とは

金子周司¹

著者抄録: ライフサイエンス辞書 (LSD) とは, 筆者が主宰するLSDプロジェクトが1993年以来作り上げてきた生命科学領域の電子辞書である。この辞書の最大の特徴は, 英語および日本語のいずれも大量の学術テキストを計量的に解析した独自のデータに基づいていることである。本稿では現在, 公開している辞書サービスの概要とLSDデータベースの構築法を紹介し, 併せて現在取り組んでいる対訳オントロジーの構築作業から見てきた英語と日本語の学術用語の相違点について述べている。

キーワード: 専門用語辞書, コーパス言語学, シソーラス, 音声データベース, オントロジー

What is the Life Science Dictionary?

KANEKO Shuji¹

Author Abstract: Life Science Dictionary (LSD) is a versatile database of life science terms developed by our volunteer researchers (LSD project) since 1993. The most prominent feature of the LSD is that both English and Japanese terms are collected based on the frequency in the corpora of scientific documents, abstracts and textbooks. Here, we describe the current status of LSD services and its original database, together with the differences between English and Japanese terms uncovered through our recent development of life science ontology.

Key words: scientific dictionary, corpus lexicology, thesaurus, voice database, ontology

原稿受理 (2006-02-22)

(情報管理 49 (1): 24-35)

1. はじめに

ライフサイエンス (生命科学) とは, 基礎学問としての生物学から応用学問としての医学までを基軸として, ささまざまな生命の理解に資する学術領域を包含する巨大かつ境界のない知識体系である。1970年代から今に至る分子生物学からゲノム科学への発展はライフサイエンスの急速な発展と拡大をもたらしたが, その知識のほとんどは論文, すなわちテキスト情報として記述され, 遺伝子情報等とともに電子的に蓄積された。そして, 研究者が見いだした新しい事物や概念は, 膨大な数の新しい専門用語を生み出してきた。

しかしながら歴史的に他の科学もそうであったように, 欧米から次々に「輸入」される新しい学問を記述する「ことば」は, わが国において十分に定義あるいは翻訳される間もなく, 研究者社会

において流通した。そういった学問の進展に対応すべく, 生命科学に関連する各種の学協会は用語集の整理や改訂を行ってきた。しかし, 爆発的な用語の増大はそれらの努力を超え, 現場で教育研究を担うわれわれからすると1990年代に制定されたさまざまな学術用語集は不十分な内容と映った。まして, 電子的に活用できる形態の学術用語集はまだなかった。

ライフサイエンス辞書プロジェクト (LSDプロジェクト) は, そういった状況下で生命科学の教育や研究現場で使いやすい辞書を自ら作ろうという目標を掲げて集まったボランティア研究グループである¹⁾。本稿ではまず現在LSDプロジェクトが行っている活動を紹介するとともに, 文献解析の過程において明らかになってきた英語と日本語の相違点から, 効率良い情報検索への展望について述べてみたい。

¹ 京都大学大学院薬学研究科 (〒606-8501京都市左京区吉田下阿達町) Tel. 075-753-4541
E-mail: skaneko@pharm.kyoto-u.ac.jp

¹ Kyoto University Graduate School of Pharmaceutical Sciences (Yoshida-shimo-adachi-cho, Sakyo-ku, Kyoto 606-8501)