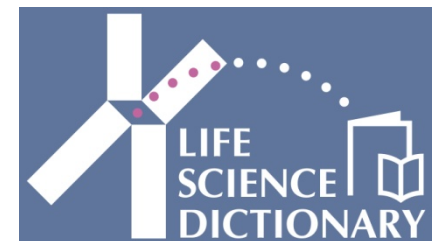


医薬品情報の構造化と 連想検索への応用

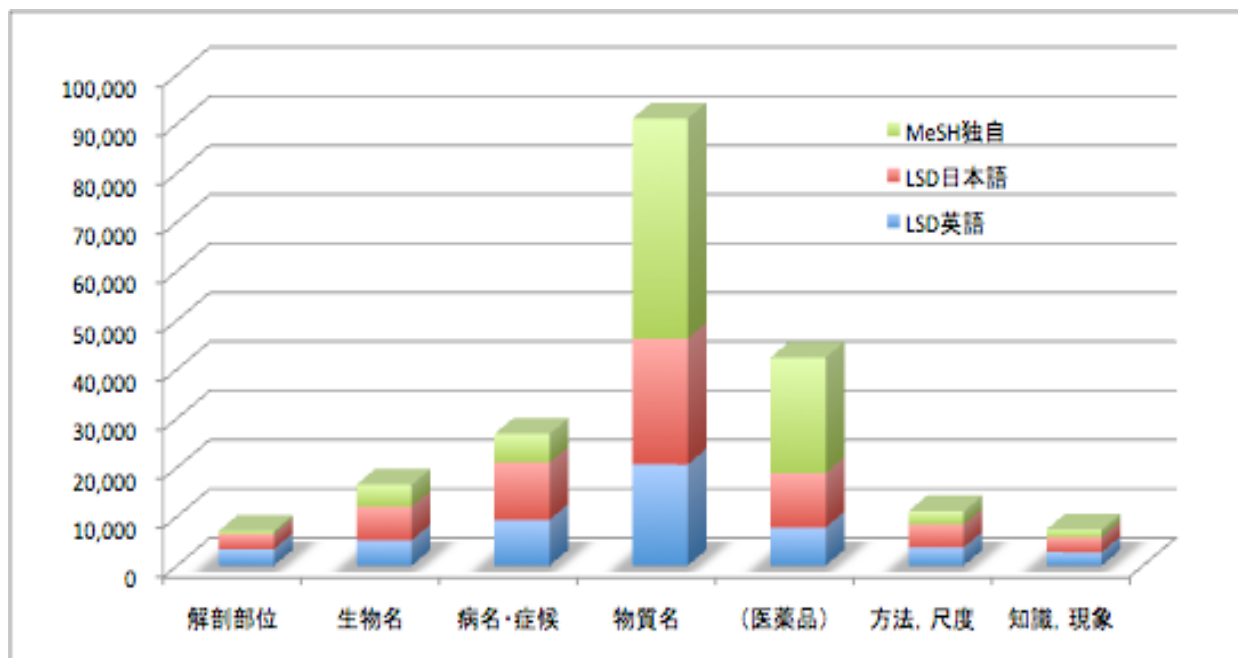
金子 周司

(京大・薬／ライフサイエンス辞書プロジェクト)



ライフサイエンス辞書(LSD)のシソーラス化

- ライフサイエンス辞書(1993～)
 - PubMed 文献抄録, 教科書での出現頻度に基づく 10 万語対訳辞書
 - 専門用語, 用言(動詞, 形容詞等), 用法, 用例, 音声などを収録
- MeSH に準拠した 2.8万語見出しシソーラスを制作(2006～)
 - LSD 名詞 12 万語を含む20万語が帰属



情報管理
53, 473-479 (2010)

対訳シソーラスの応用

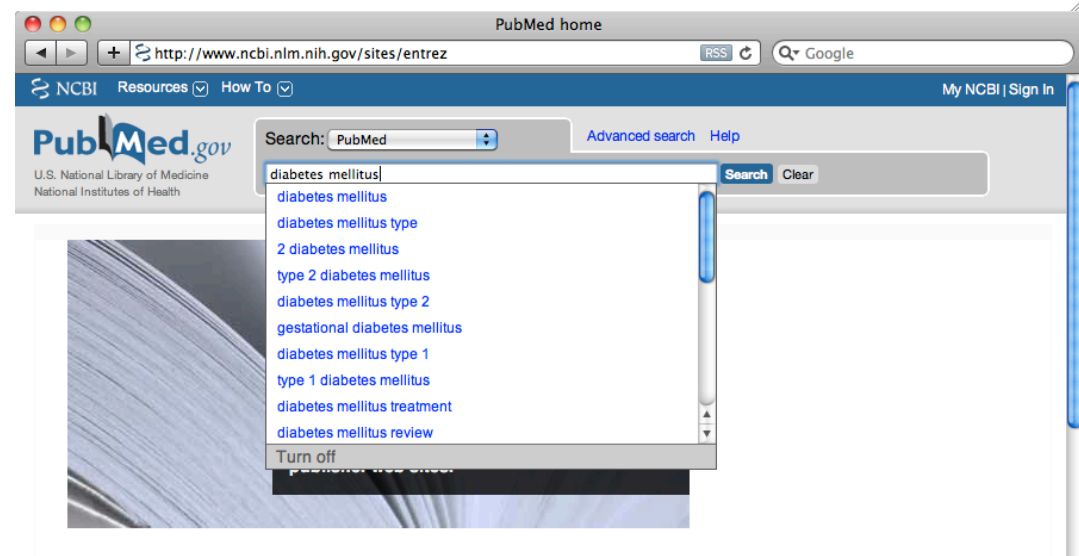
1. 対象の抽出

- 表記の解釈, 整理統一
 - 索引自動作成(学会抄録、医療文書)
 - 情報検索での表記のゆれを吸収
- 未収録語の抽出

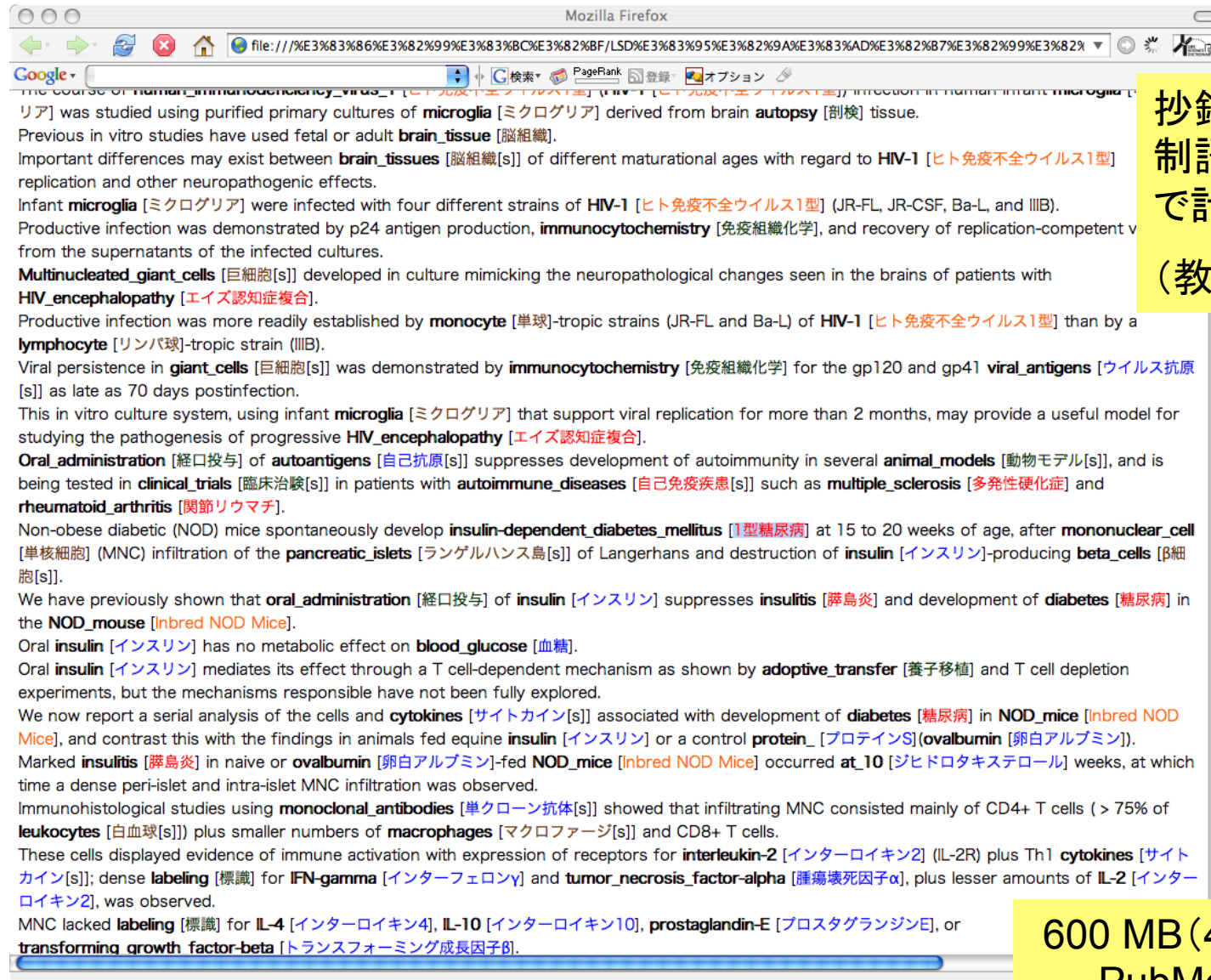


2. 関係の抽出

- 情報検索ヘルパー
 - **連想検索**
- データマイニング
 - AERS 情報解析
- テキストマイニング
 - 電子カルテ解析



簡易テキストマイニング



抄録中に共起する統
制語をPerlスクリプト
で計数

(教科書での解析例)

600 MB (45万論文, 1億語)
PubMed抄録に適用

概念共起解析

tf = 単語 t_1 と単語 t_2 の共起回数
 idf = \log (全文書数 N / 共起単語 t_2 が出現する文書数)
 $tf \cdot idf$ = $tf \times idf$

MeSH browser interface showing search results for 'Insulin' (統制語コード D007328, 出現文書数 17372). The interface includes a search bar, navigation controls, and a list of related concepts.

抄録内で共起する専門用語の共起頻度より tf-idf ソート

↓

- 相互作用および相互作用点
- 対比／並立概念
- 上位／下位概念

などが抽出された

単語	tf_{ij}	$tf \cdot idf$
グルコース	3838 / 16109	5481.35
インスリン受容体	431 / 1149	1109.79
グルカゴン	392 / 941	1043.37
血糖	289 / 963	766.32
トリグリセリド	323 / 3123	691.44
脂肪	371 / 6872	667.12
チロシン	360 / 12037	559.70
リン酸転移酵素	409 / 27292	490.48
Cペプチド	148 / 352	457.13
アディポネクチン	139 / 796	380.07
非受容体1型チロシンホスファターゼ	126 / 677	353.39
グリコーゲン合成酵素	115 / 421	346.26
PPAR γ	152 / 2569	338.27
グリコヘモグロビンA	117 / 570	336.89
ワートマニン	132 / 1257	334.74
ソマトスタチン	121 / 800	330.59
ホスファチジルイノシトール	158 / 4023	320.85
トログリタゾン	97 / 488	285.84
血糖降下薬	90 / 301	284.10
成長ホルモン	106 / 1052	277.00

対訳シソーラスの公開

WebLSD検索結果: ワーファリン

http://lsd.pharm.kyoto-u.ac.jp/cgi-bin/lspdproj/ejlookup04

ONLINE LIFE SCIENCE DICTIONARY
ライフサイエンス辞書オンラインサービス

プロジェクト WebLSD O/D英語教材 EtoJ Vocab EtoJ WebSpell

ミラーサイト [京都1 | 京都2 | 東京 (Weblio)] フォントサイズ [小 | 中 | 大 | 特大] ENGLISH

新規対訳の受け付け | 検索結果の読み方 |

英和・和英 シソーラス 英語共起表現 オプション表示/非表示

英和・和英検索: search clear

▶ 和英検索結果

▶ ワーファリン **わーふぁりん** **シソーラス** Scholar, Entrez, Google, Wikipedia
(n) ((商品名)) **warfarin** 共起表現

【関連語】 ワルファリン, ワルファリンカリウム, クマジン, ワルファリンナトリウム

219236710th transaction of WebLSD@Pharm node1 since Nov 11, 1996

WebLSDに未収録の専門用語 (用法) は "新規対訳の受け付け" から投稿できます。

Page Top

Copyright 2011 ライフサイエンス辞書プロジェクト



LSDシソーラス: ワルファリン Warfarin

http://lsd.pharm.kyoto-u.ac.jp/webbsd/c/tree/D014859

ONLINE LIFE SCIENCE DICTIONARY
ライフサイエンス辞書オンラインサービス

プロジェクト WebLSD O/D英語教材 EtoJ Vocab EtoJ WebSpell

ミラーサイト [京都1 | 京都2 | 東京 (Weblio)] フォントサイズ [小 | 中 | 大 | 特大] ENGLISH

新規対訳の受け付け | 検索結果の読み方 |

英和・和英 シソーラス 英語共起表現 オプション表示/非表示

シソーラス検索: search clear

▶ LSDシソーラス: ワルファリン Warfarin

ワルファリン (Warfarin) を Google Scholar, Entrez, Google, Wikipedia で検索

同義語 (異表記) :

- 4-Hydroxy-3-(3-oxo-1-phenylbutyl)-2H-1-benzopyran-2-one
- Aldo Brand of Warfarin Sodium
- Aldocumar
- Antigen Brand of Warfarin Sodium
- Apo-Warfarin
- Apotex Brand of Warfarin Sodium
- Bailly Brand of Warfarin Sodium
- Boots Brand of Warfarin Sodium
- Bristol-Myers Squibb Brand of Warfarin Sodium
- Coumadin
- Coumadine
- Estedi Brand of Warfarin Sodium
- Gen-Warfarin
- Genpharm Brand of Warfarin Sodium
- Goldshield Brand of Warfarin Sodium
- Marevan
- Tedicumar
- Waran
- Warfant
- Warfarin Potassium
- Warfarin Sodium
- クマジン
- ワーリン
- アレファリン
- クーマディン
- コーマディン
- ワーファリン
- ワルファリンカリウム
- ワルファリンナトリウム

概念ツリー:

- 複素環化合物 Heterocyclic Compound
 - 二環複素化合物 2-Ring Heterocyclic Compound
 - ベンゾピラン Benzopyran
 - クマリン Coumarin
 - 4-ヒドロキシクマリン 4-Hydroxycoumarin
 - ワルファリン Warfarin
- 複素環化合物 Heterocyclic Compound
 - ピラン Pyran
 - ベンゾピラン Benzopyran
 - クマリン Coumarin
 - 4-ヒドロキシクマリン 4-Hydroxycoumarin
 - ワルファリン Warfarin

英和・和英辞書からリンク
同義語や概念ツリーを提示

WebLSD シソーラス
無料公開

共起概念による連想検索

連想検索（共起語との組み合わせでウェブを検索）：

以下のリストで

日本語をクリック → **ワルファリン** との組み合わせでGoogleを検索
 英語をクリック → **Warfarin** との組み合わせでEntrezを検索

ビタミンK (Vitamin K); ヘパリン (Heparin); 血液凝固因子 (Blood Coagulation Factor); リファンピシン (Rifampin); アスピリン (Aspirin); アステミゾール (Astemizole); アンチトロンピン (Antithrombin); リスク回避行動 (Risk Reduction Behavior); 判断 (Judgment); パルプロ酸 (Valproic Acid); 自殺企図 (Attempted Suicide); フルコナゾール (Fluconazole); 麦角アルカロイド (Ergot Alkaloid); キニジン (Quinidine); 喫煙 (Smoking); アニシンジオン (anisindione); 抗体 (Antibody); 抑うつ (Depression); 活性化第X因子 (Factor Xa); 意図 (Intention); 意志決定 (Decision Making); 懸濁液 (Suspension); 静脈脂肪乳剤 (Intravenous Fat Emulsion); マトリックスGlaタンパク質 (matrix Gla protein); タンパク質同化薬 (Anabolic Agent); 神経筋遮断薬 (Neuromuscular Blocking Agent); ビタミンK2 (Vitamin K2); ペリオスチン (perioestin); ヨヒンビン (Yohimbine); アモキシシリン (Amoxicillin); ダンシル化合物 (Dansyl Compound); 3-ヒドロキシ-3-メチルグルタリルCoA (3-hydroxy-3-methylglutaryl-CoA); 査読 (Peer Review); リトナビル (Ritonavir); 第VIII因子 (Factor VIII); オレイン酸 (Oleic Acid); 自覚 (Awareness); アンジオテンシン変換酵素阻害薬 (Angiotensin-Converting Enzyme Inhibitor); 無機質 (Mineral); ビタミン (Vitamin); 副腎皮質ステロイド (Corticosteroid); 協調行動 (Cooperative Behavior); 矛盾 (Conflict); 不安 (Anxiety); 性自認 (Gender Identity); トリチウム (Tritium); 酸化還元酵素 (Oxidoreductase); 鉄 (Iron)

薬理作用と構造情報：

駆除薬 (Pesticide); 血液系作用薬 (Hematologic Agent); 抗凝固薬 (Anticoagulants); 殺鼠剤 (Rodenticide)

Page Top

Copyright 2011 ライフサイエンス辞書プロジェクト

- ・5700種(79%)の医薬品に連想検索
- ・主要1326種については92%以上で共起上位30語を提示

共起上位30語

日本語→Google で共起検索

英語→Entrez で共起検索

ワルファリン and リファンピシン - Google 検索

http://www.google.co.jp/search?q=ワルファリン%20AND%20リファンピ

Google

ワルファリン AND リファンピ

約 15,500 件 (0.16 秒... Google.com in English 検索オプション)

すべて

画像

動画

ニュース

ショッピング

もっと見る

京都府京都市 場所を変更

ウェブ全体から検索

日本語のページを検索

翻訳して検索

もっとツールを見る

調剤薬局日記/ワーファリンの相互作用

www.jade.dti.ne.jp/~ma-hata/a_02_wafarin.htm - キャッシュ

ワルファリンは、非常に飲み合わせの問題が多いことで知られ、その相互作用の機序も様々です。ワルファリンは血栓の生成を抑える大切な働きをして... リファンピシン、肝代謝酵素誘導、ワルファリンの代謝が亢進し、作用減弱。シメジニン、アルコール...

ワルファリンとリファンピシン併用患者のカルテ調査による相互作用の検討

ci.nii.ac.jp/naid/110001166582 - キャッシュ

和田 著 - 2002

ワルファリンとリファンピシン併用患者のカルテ調査による相互作用の検討 Interaction of Warfarin and Rifampicin on Medical Report. 和田 恭一 Wada Kyoichi; 国立循環器病センター薬剤部 Department of Pharmacy, National Cardiovascular Center ...

ワルファリンとリファンピシン併用患者のカルテ調査による相互作用の...

sc.chat-shuffle.net/paper/uid:110001166582 - キャッシュ

2010年5月13日 - ワルファリンとリファンピシン併用患者のカルテ調査による相互作用の検討. Sponsored Links. Rifampicin's ability to induce hepatic-microsomal enzymes is well known. As a result, rifampicin has often been implicated in drug ...

J-GLOBAL - ワルファリンとリファンピシン併用患者のカ... 【文献】

jglobal.jst.go.jp/public/20090422/200902144701200113

TOP > ワルファリンとリファンピシン併用患者のカ... J-GLOBALとは、研究者、文献、特許などの情報をつなぐことで、異分野の知や意外な発見などを支援する新しいサービス

カテゴリーによる共起語の偏り

糖尿病性腎症 $N=431,764$ $df_i=852$

0.6 GB Corpus

	$tf_{i,j}$	$tf \cdot idf$
蛋白尿	45 / 1095	116.81
糖尿病	71 / 12186	110.01
慢性腎不全	43 / 1610	104.42
腎疾患	42 / 1801	99.95
糸球体濾過率	35 / 1563	85.45
2型糖尿病	38 / 4055	77.04
高血圧	39 / 5623	73.53
1型糖尿病	29 / 2512	64.82
アンジオテンシン変換酵素阻害薬	24 / 1075	62.49
高血糖	23 / 1780	54.85
グルコース	29 / 16109	41.42
肥大	16 / 2593	35.54
腎移植	11 / 437	32.94
クレアチニン	14 / 2323	31.77
トランスフォーミング増殖因子 β	19 / 9336	31.64
アンジオテンシン	11 / 584	31.56
アルブミン	13 / 1881	30.69
リスク	32 / 47802	30.59
線維症	15 / 4967	29.09
低アルドステロン症	7 / 43	28.01

疾患→機序や治療薬など
バランス良い

アンジオテンシン1型受容体遮断 $N=431,764$ $df_i=135$

0.6 GB Corpus

	$tf_{i,j}$	$tf \cdot idf$
ロサルタン	35 / 488	103.14
アンジオテンシン変換酵素阻害薬	16 / 1075	41.66
アンジオテンシン1型受容体	9 / 414	27.16
アンジオテンシンII	9 / 1493	22.15
カンデサルタン	5 / 175	16.96
アンジオテンシン	5 / 584	14.34
酵素阻害薬	3 / 368	9.21
ブラジキニン	3 / 1005	7.90
バルサルタン	2 / 105	7.23
アンジオテンシンI	2 / 116	7.14
ネプリライシン	2 / 190	6.71
キニン	2 / 206	6.64
降圧薬	2 / 599	5.72
血管拡張薬	2 / 998	5.27
カンデサルタンシレキセチル	1 / 11	4.59
テルミサルタン	1 / 12	4.56
癌遺伝子産物c-Fos	2 / 2761	4.39
アルドステロン拮抗薬	1 / 34	4.10
一酸化窒素	2 / 5121	3.85
クロルタリドン	1 / 65	3.82

薬物分類→内包される薬
など下位概念が多い

ドネペジル $N=431,764$ $df_i=78$

0.6 GB Corpus

	$tf_{i,j}$	$tf \cdot idf$
ガランタミン	14 / 73	52.81
コリンエステラーゼ阻害薬	10 / 214	33.05
リバステグミン	7 / 26	29.54
タクリン	7 / 57	27.16
エドロホニウム	4 / 46	15.89
認知	4 / 1024	10.50
フィノスチグミンサリチル酸塩	2 / 4	10.07
コリンエステラーゼ	2 / 110	7.19
フィノスチグミン	2 / 129	7.05
アセチルコリン	3 / 3620	6.23
ビタミンE	2 / 944	5.32
アレコリン	1 / 24	4.26
ピリドスチグミン	1 / 38	4.06
オキソレモリン	1 / 51	3.93
ネオスチグミン	1 / 81	3.73
薬理作用	1 / 123	3.55
課題処理能力評価	1 / 211	3.31
ムスカリン受容体刺激薬	1 / 241	3.25
ピロカルピン	1 / 271	3.20
プロピジウム	1 / 278	3.19

単独薬物→併用薬や類似
薬が多い

Levenstein 距離とツリー距離を用いる補正

- Levenstein 距離: 2つの文字列の異なり度を表す尺度
 - Losartan と Varsartan は2文字消去して3文字追加=距離「5」
文字長8と9で合計17なので $5/17$ (=0.29)を類似度係数とする方法
 - しかし、interleukin-1 と interleukin-5 は名前は似ているが、機能は異なる文字列のみで考える限界
- ツリー距離: シソーラス階層での距離と深さを表す尺度
 - Type 1 Diabetes mellitus と Type 2 Diabetes mellitus は兄弟関係親である Diabetes mellitus はレベル4の深さ
この場合、ツリー近接度を $4*2 / (5+5) = 0.8$ とする
 - ツリーに収録された場合は効果的である
 - しかし、構造は違っても(=ツリーで離れていても)、薬効類似の薬物を排除できない
(例) ジルチアゼム と ベラパミル
- 医薬品分類による類似度
 - 薬効分類による係数が必要か



まとめ

- MeSH準拠によるLSDシソーラス化がひとまず完了したので、医学研究情報ポータルへの応用を試みた
- 英語抄録中に共起する統制語をカウントすることによって、2.8万の統制語に最大30個の共起語を選んだ(336万組)
- Webで公開している10万語対訳辞書に、シソーラスツリーと共起語による連想検索を実装した
- 医薬品名称を商品名までカバーするシソーラスに情報検索ヘルパーとしての連想検索が可能になった
- 類似薬や併用薬については、関連度による係数を設けて減算する必要がある
- シソーラスの維持も、今後の大きな課題である