

2万語かな漢字変換辞書 スタンドアロン辞書 (ベクターで無料配布)

MS-IME, こえり, ATOK

一般辞書によるかな漢字変換

→

ライフサイズシステム辞書によるかな漢字変換

→

Jamming (EPWING互換)

2008年1月22日 統合データベース

耳で覚えるライフサイエンス英語(無料配布)

必須2000語を iPod で

2008年1月22日 統合データベース

対訳の定義

頻度調査 → 大学院生アルバイトによる調査 → 校閲

- Web ベースの共同作業スペース
- 品詞, 属性(意味情報), 1行解説, MeSH term IDなどを付加
- 参考図書・辞書およそ100点の調査
- Google調査結果も併用し, 複数出典での確認で採用

Term	Definition	Source	Status
biology	生物学	辞書	採用
chemistry	化学	辞書	採用
physics	物理学	辞書	採用
mathematics	数学	辞書	採用

英語と日本語の対訳関係の複雑さ: 単語の置き換え

tumor promotor は 発癌プロモータ?

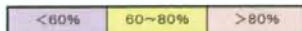
英語	PubMed ¹⁾	Textbook ²⁾	日本語	基礎系	臨床系
cancer!	28,302 ¹⁾ (43%) ¹⁾	5,158 ²⁾ (50%) ²⁾	癌	13,630 ¹⁾ (72%) ¹⁾	1,154 ²⁾ (52%) ²⁾
carcinoma!	5,846 ¹⁾ (9%) ¹⁾	1,368 ²⁾ (13%) ²⁾	癌腫	29 ¹⁾ (0%) ¹⁾	5 ²⁾ (0%) ²⁾
sarcoma!	1,235 ¹⁾ (2%) ¹⁾	340 ²⁾ (3%) ²⁾	肉腫	289 ¹⁾ (2%) ¹⁾	57 ²⁾ (3%) ²⁾
tumor!	31,207 ¹⁾ (47%) ¹⁾	3,242 ²⁾ (31%) ²⁾	腫瘍	4,192 ¹⁾ (22%) ¹⁾	989 ²⁾ (44%) ²⁾
neoplasm!	504 ¹⁾ (<1%) ¹⁾	207 ²⁾ (2%) ²⁾	新生物	4 ¹⁾ (0%) ¹⁾	6 ²⁾ (0%) ²⁾
oncogenesis!	373 ¹⁾ (<1%) ¹⁾	6 ²⁾ (0%) ²⁾	発癌	763 ¹⁾ (4%) ¹⁾	39 ²⁾ (2%) ²⁾
計	66,357 ¹⁾ (100%) ¹⁾	10,321 ²⁾ (100%) ²⁾		18,916 ¹⁾ (100%) ¹⁾	2,250 ²⁾ (100%) ²⁾

2008年1月22日 統合データベース

「protein」の表記における選択

多数の表記がある場合は、隣接語によって表記が使い分けられる

連接語	計	タンパク質	タンパク	蛋白質	蛋白	プロテイン
(総出現数)	28,201 ¹⁾	19,815 ²⁾	3,960 ³⁾	2,688 ⁴⁾	1,066 ⁵⁾	672 ⁶⁾
～合成	316 ¹⁾	229 ²⁾	52 ³⁾	24 ⁴⁾	11 ⁵⁾	0 ⁶⁾
結合～	1,234 ¹⁾	879 ²⁾	187 ³⁾	139 ⁴⁾	29 ⁵⁾	0 ⁶⁾
リン酸化～	154 ¹⁾	83 ²⁾	22 ³⁾	31 ⁴⁾	18 ⁵⁾	0 ⁶⁾
プリオン～	73 ¹⁾	8 ²⁾	9 ³⁾	2 ⁴⁾	54 ⁵⁾	0 ⁶⁾
～キナーゼ	483 ¹⁾	49 ²⁾	7 ³⁾	3 ⁴⁾	2 ⁵⁾	422 ⁶⁾
リボ～	207 ¹⁾	74 ²⁾	97 ³⁾	5 ⁴⁾	15 ⁵⁾	16 ⁶⁾
～尿	2 ¹⁾	0 ²⁾	2 ³⁾	0 ⁴⁾	0 ⁵⁾	0 ⁶⁾
(総数・臨床)	355 ¹⁾	0 ²⁾	0 ³⁾	8 ⁴⁾	340 ⁵⁾	7 ⁶⁾
～尿(臨床)	71 ¹⁾	0 ²⁾	0 ³⁾	0 ⁴⁾	71 ⁵⁾	0 ⁶⁾



2008年1月22日 統合データベース

用語の拡充に伴う問題

- 単語レベルでのカバー率向上 → 隣接語解析を用いた複合語への拡張
- 同義語の並立 → 関連語の整理(同じ訳語をもつもの)

The screenshot shows a search engine interface with the query 'diabetic neuropathy'. The results list several related terms and their corresponding Japanese translations. A table on the right side of the page lists the following information:

用語番号	用語名	訳語	関連語
0001001	糖尿病性神経障害	diabetic neuropathy	
0001002	糖尿病性末梢神経障害	diabetic peripheral neuropathy	
0001003	糖尿病性自律神経障害	diabetic autonomic neuropathy	
0001004	糖尿病性筋梗塞	diabetic muscle infarction	
0001005	糖尿病性腎臓病	diabetic nephropathy	
0001006	糖尿病性神経炎	diabetic neuritis	
0001007	糖尿病性網膜症	diabetic retinopathy	
0001008	糖尿病性皮膚病	diabetic dermopathy	
0001009	糖尿病性骨髄炎	diabetic osteomyelitis	
0001010	糖尿病性骨髄炎	diabetic osteomyelitis	
0001011	糖尿病性骨髄炎	diabetic osteomyelitis	
0001012	糖尿病性骨髄炎	diabetic osteomyelitis	
0001013	糖尿病性骨髄炎	diabetic osteomyelitis	
0001014	糖尿病性骨髄炎	diabetic osteomyelitis	
0001015	糖尿病性骨髄炎	diabetic osteomyelitis	
0001016	糖尿病性骨髄炎	diabetic osteomyelitis	
0001017	糖尿病性骨髄炎	diabetic osteomyelitis	
0001018	糖尿病性骨髄炎	diabetic osteomyelitis	
0001019	糖尿病性骨髄炎	diabetic osteomyelitis	
0001020	糖尿病性骨髄炎	diabetic osteomyelitis	
0001021	糖尿病性骨髄炎	diabetic osteomyelitis	
0001022	糖尿病性骨髄炎	diabetic osteomyelitis	
0001023	糖尿病性骨髄炎	diabetic osteomyelitis	
0001024	糖尿病性骨髄炎	diabetic osteomyelitis	
0001025	糖尿病性骨髄炎	diabetic osteomyelitis	
0001026	糖尿病性骨髄炎	diabetic osteomyelitis	
0001027	糖尿病性骨髄炎	diabetic osteomyelitis	
0001028	糖尿病性骨髄炎	diabetic osteomyelitis	
0001029	糖尿病性骨髄炎	diabetic osteomyelitis	
0001030	糖尿病性骨髄炎	diabetic osteomyelitis	
0001031	糖尿病性骨髄炎	diabetic osteomyelitis	
0001032	糖尿病性骨髄炎	diabetic osteomyelitis	
0001033	糖尿病性骨髄炎	diabetic osteomyelitis	
0001034	糖尿病性骨髄炎	diabetic osteomyelitis	
0001035	糖尿病性骨髄炎	diabetic osteomyelitis	
0001036	糖尿病性骨髄炎	diabetic osteomyelitis	
0001037	糖尿病性骨髄炎	diabetic osteomyelitis	
0001038	糖尿病性骨髄炎	diabetic osteomyelitis	
0001039	糖尿病性骨髄炎	diabetic osteomyelitis	
0001040	糖尿病性骨髄炎	diabetic osteomyelitis	
0001041	糖尿病性骨髄炎	diabetic osteomyelitis	
0001042	糖尿病性骨髄炎	diabetic osteomyelitis	
0001043	糖尿病性骨髄炎	diabetic osteomyelitis	
0001044	糖尿病性骨髄炎	diabetic osteomyelitis	
0001045	糖尿病性骨髄炎	diabetic osteomyelitis	
0001046	糖尿病性骨髄炎	diabetic osteomyelitis	
0001047	糖尿病性骨髄炎	diabetic osteomyelitis	
0001048	糖尿病性骨髄炎	diabetic osteomyelitis	
0001049	糖尿病性骨髄炎	diabetic osteomyelitis	
0001050	糖尿病性骨髄炎	diabetic osteomyelitis	

用語法 Terminology 多様性の解決

- 大量の生物情報の蓄積に伴う現実的な問題
 - (例) 1型糖尿病を表す term
 - diabetes mellitus type 1
 - juvenile-onset diabetes mellitus
 - insulin-dependent diabetes mellitus
 - autoimmune diabetes
 - IDDM, type 1 DM
 - 1型糖尿病, 1型糖尿病, 糖尿病1型
 - インスリン依存性糖尿病, インシュリン依存性糖尿病
 - などなど

LSD と MeSH の統合によるシソーラス構築

■ MeSH Descriptor (24,767語)

□ Tree (48,442項目)

- 1 Descriptor が複数Treeに属する

■ MeSH Supplemental Concepts (物質名 173,049語)

□ MeSH Concept (283,837語)
定義 Scope Note をもつ

- MeSH Term (541,756語)

↓↓ 一致する語句をIDでリレート

- LSD英語 (78,282語)

↓↓

- LSD日本語 (89,610語)

! MeSH terms を自然表記に自動訂正するスクリプトを開発

語順訂正

! Diabetes mellitus, Type 1

→ Type 1 Diabetes mellitus

複数単数をラテン語OK)

! Neurofibromatoses

→ Neurofibromatosis

! 人的作業による表記修正 (< 1%)

! ID管理で年次更新にも耐えうる

! 特定カテゴリーを集中的に作業

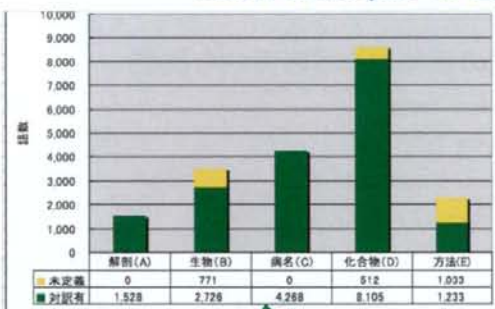
! 病名, 解剖部位名 (2006)

! 物質名, 生物学名 (2007)

! 方法, その他 (2008)

2008年1月22日 統合データベース

MeSH Descriptor の LSD 翻訳率

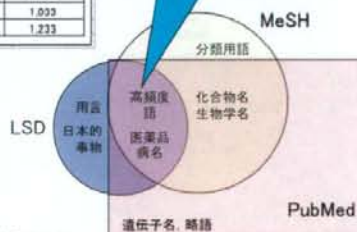


MeSH Descriptor (A-E, 20,178項目) の 89% をカバー (2008年1月)

2008年度中に完成予定 (>18万語)

LSD収録78,270語のうち、33,192語 (42%) が MeSH term

5カテゴリーで48,442語のうち、MeSHツリー下に34,791語 (75%) 入る



2008年1月22日 統合データベース

(A)解剖学用語

- ！解剖学用語集は、1万語前後のものが多い
 - 日本解剖学会改訂13版(2007)が最新
 - 「右肺上葉」「左肺外側肺底区」など部位名多い
 - ラテン語が主体で、英語と日本語が付随する
- ！MeSHでは、3,075 terms を1,528 descriptors に収録
 - 個別部位名少なく、解剖学用語としては不十分
 - 「肺泡」はあるが、「肺泡上皮細胞」など詳細概念が少ない
 - ラテン語はあまり収録されてなく、英語が主体
- ！LSDは5,583英語表記、5,247日本語表記を収録
 - 英語は3,145語(57%)がMeSHと重なるが2,438語がツリーに収まらず
 - 最大限、ツリーに収録する努力をするが、詳細概念はNLMに収録要望
 - LSD, MeSH 共に組織学用語、非ヒト組織、実験材料が多い
 - (例)「間葉系細胞」、「アフリカツメガエル卵母細胞」、「乳癌細胞株」
 - 略語はLSDのみにある
 - (例)「CNS」、「ES cell」、「BBB」、「CA1 region」

2008年1月22日 統合データベース

MeSH	日本語
D000001	解剖学
D000002	生理学
D000003	生化学
D000004	薬理学
D000005	病理学
D000006	免疫学
D000007	微生物学
D000008	寄生虫学
D000009	植物学
D000010	動物学
D000011	遺伝学
D000012	分子生物学
D000013	細胞生物学
D000014	発生生物学
D000015	神経生物学
D000016	免疫細胞生物学
D000017	分子免疫学
D000018	免疫遺伝学
D000019	免疫病理学
D000020	免疫療法
D000021	免疫学検査
D000022	免疫学検査法
D000023	免疫学検査薬
D000024	免疫学検査薬
D000025	免疫学検査薬
D000026	免疫学検査薬
D000027	免疫学検査薬
D000028	免疫学検査薬
D000029	免疫学検査薬
D000030	免疫学検査薬
D000031	免疫学検査薬
D000032	免疫学検査薬
D000033	免疫学検査薬
D000034	免疫学検査薬
D000035	免疫学検査薬
D000036	免疫学検査薬
D000037	免疫学検査薬
D000038	免疫学検査薬
D000039	免疫学検査薬
D000040	免疫学検査薬
D000041	免疫学検査薬
D000042	免疫学検査薬
D000043	免疫学検査薬
D000044	免疫学検査薬
D000045	免疫学検査薬
D000046	免疫学検査薬
D000047	免疫学検査薬
D000048	免疫学検査薬
D000049	免疫学検査薬
D000050	免疫学検査薬
D000051	免疫学検査薬
D000052	免疫学検査薬
D000053	免疫学検査薬
D000054	免疫学検査薬
D000055	免疫学検査薬
D000056	免疫学検査薬
D000057	免疫学検査薬
D000058	免疫学検査薬
D000059	免疫学検査薬
D000060	免疫学検査薬
D000061	免疫学検査薬
D000062	免疫学検査薬
D000063	免疫学検査薬
D000064	免疫学検査薬
D000065	免疫学検査薬
D000066	免疫学検査薬
D000067	免疫学検査薬
D000068	免疫学検査薬
D000069	免疫学検査薬
D000070	免疫学検査薬
D000071	免疫学検査薬
D000072	免疫学検査薬
D000073	免疫学検査薬
D000074	免疫学検査薬
D000075	免疫学検査薬
D000076	免疫学検査薬
D000077	免疫学検査薬
D000078	免疫学検査薬
D000079	免疫学検査薬
D000080	免疫学検査薬
D000081	免疫学検査薬
D000082	免疫学検査薬
D000083	免疫学検査薬
D000084	免疫学検査薬
D000085	免疫学検査薬
D000086	免疫学検査薬
D000087	免疫学検査薬
D000088	免疫学検査薬
D000089	免疫学検査薬
D000090	免疫学検査薬
D000091	免疫学検査薬
D000092	免疫学検査薬
D000093	免疫学検査薬
D000094	免疫学検査薬
D000095	免疫学検査薬
D000096	免疫学検査薬
D000097	免疫学検査薬
D000098	免疫学検査薬
D000099	免疫学検査薬
D000100	免疫学検査薬

(A)解剖学用語

The screenshot displays the MeSH database interface. On the left, a search for 'Dendritic Cells' (Descriptor D000071) is shown, with a tree view of related terms and a list of 14 synonyms. The main window shows a list of terms with their MeSH IDs and counts. A table on the right lists terms and their MeSH IDs.

Descriptor	Synonym数
D000071	14

MeSH ID	Term	Count
D000071	Dendritic Cells	14
D000072	Antigen Presenting Cells	13
D000073	Macrophages	12
D000074	Monocytes	11
D000075	Neutrophils	10
D000076	Eosinophils	9
D000077	Basophils	8
D000078	Platelets	7
D000079	Granulocytes	6
D000080	Leukocytes	5
D000081	Immune Cells	4
D000082	White Blood Cells	3
D000083	Granulocyte	2
D000084	Leukocyte	1
D000085	Cell	1
D000086	Granulocyte	1
D000087	Leukocyte	1
D000088	Cell	1
D000089	Granulocyte	1
D000090	Leukocyte	1
D000091	Cell	1
D000092	Granulocyte	1
D000093	Leukocyte	1
D000094	Cell	1
D000095	Granulocyte	1
D000096	Leukocyte	1
D000097	Cell	1
D000098	Granulocyte	1
D000099	Leukocyte	1
D000100	Cell	1

2008年1月22日 統合データベース

(B) 生物学名

- 生物学名は、何十万もある
 - 分類自体も複数の説がある
 - 命名規約により階層構造はしっかりしている
(界, 門, 綱, 目, 科, 属, 種)
 - ラテン語が主体で、英語と日本語が付随する
- MeSHは 9,129 terms を 3,497 descriptors に
 - 疾患関連の代表的な生物とウイルスを収録
 - ラテン語学名と英語通称名が混在
 - (Acaridani目 > Miteコダニ類 > Acaridaeコナダニ科)
- LSDは5,272英語表記, 7,142日本語表記を収録
 - 略語や英日混在表記を多数収録している
 - 「E. coli」, 「CMV」, 「Epstein-Barrウイルス」
 - 英語は4,552語(86%)がMeSHと重なるが720語がツリーに収まらず
 - 特定の生物名であることが多いので、上位のMeSHに収める

2008年1月22日 統合データベース

(B) 生物学名

The screenshot displays the MeSH browser interface. On the left, there are several panels showing hierarchical lists of terms under various categories like 'Fungi', 'Protozoa', and 'Mites'. The main area shows a list of terms with their MeSH IDs, such as '0012441 Saccharomyces cerevisiae' and '001438B Human T-lymphotropic virus 1'. On the right, a detailed view of the selected term '001438B' is shown, listing its parent terms (e.g., '001438A Human T-lymphotropic virus', '001438 Human T-lymphotropic virus 1') and child terms (e.g., '001438B Human T-lymphotropic virus 1', '001438C Human T-lymphotropic virus 2'). The interface includes navigation buttons like 'Home', 'Search', and 'Print'.

1年1月22日 統合データベース

(C) 病名・症候名

1 病名は数万語程度

標準病名マスター(2万語), MedDRA(6.5万語)
ICD-10 など、内外に標準表記(分類)あり

1 パリエーションや異字体が多い

1 MeSHは 13,330 terms を4,268 descriptors に

代表的な疾患のみを網羅

1 個別部位の病名は不完全

1 身体各所での炎症, 腫瘍, 痛み

1 急性, 慢性などの記述子

1 LSDIは 13,563 英語表記, 16,350日本語表記を収録

1 略語, 異表記, 英日混在表記を多数収録している

1 「SCA1」、「頻拍/頻脈」、「Addison病」

1 英語は9,124語(67%)がMeSHと重なるが4,439語がツリーに収まらず

1 MeSHにない概念多く, 他のソーラスとの照合も考えるべきか?

0001-0001	頭痛 (Headache)
0001-0002	頭痛 (Headache)
0001-0003	頭痛 (Headache)
0001-0004	頭痛 (Headache)
0001-0005	頭痛 (Headache)
0001-0006	頭痛 (Headache)
0001-0007	頭痛 (Headache)
0001-0008	頭痛 (Headache)
0001-0009	頭痛 (Headache)
0001-0010	頭痛 (Headache)
0001-0011	頭痛 (Headache)
0001-0012	頭痛 (Headache)
0001-0013	頭痛 (Headache)
0001-0014	頭痛 (Headache)
0001-0015	頭痛 (Headache)
0001-0016	頭痛 (Headache)
0001-0017	頭痛 (Headache)
0001-0018	頭痛 (Headache)
0001-0019	頭痛 (Headache)
0001-0020	頭痛 (Headache)
0001-0021	頭痛 (Headache)
0001-0022	頭痛 (Headache)
0001-0023	頭痛 (Headache)
0001-0024	頭痛 (Headache)
0001-0025	頭痛 (Headache)
0001-0026	頭痛 (Headache)
0001-0027	頭痛 (Headache)
0001-0028	頭痛 (Headache)
0001-0029	頭痛 (Headache)
0001-0030	頭痛 (Headache)
0001-0031	頭痛 (Headache)
0001-0032	頭痛 (Headache)
0001-0033	頭痛 (Headache)
0001-0034	頭痛 (Headache)
0001-0035	頭痛 (Headache)
0001-0036	頭痛 (Headache)
0001-0037	頭痛 (Headache)
0001-0038	頭痛 (Headache)
0001-0039	頭痛 (Headache)
0001-0040	頭痛 (Headache)
0001-0041	頭痛 (Headache)
0001-0042	頭痛 (Headache)
0001-0043	頭痛 (Headache)
0001-0044	頭痛 (Headache)
0001-0045	頭痛 (Headache)
0001-0046	頭痛 (Headache)
0001-0047	頭痛 (Headache)
0001-0048	頭痛 (Headache)
0001-0049	頭痛 (Headache)
0001-0050	頭痛 (Headache)

2008年1月22日 統合データベース

(C) 病名・症候名

The screenshot shows a complex medical database interface. At the top, there's a search bar and navigation tabs. The main area is divided into several panes. The left pane shows a list of search results for 'Diabetes Mellitus, Type 1'. The middle pane shows a detailed view of a selected item, 'Acute Myocardial Infarction', with various sub-panels for details and related terms. The right pane shows a list of related terms or synonyms. The interface is densely packed with text and icons, typical of a professional medical database.

2008年1月22日 統合データベース

(D) 生体分子・化合物名

- ！ 生体分子は数万語程度 (遺伝子数 + α)
 - ！ 略語や異表記が非常に多い
- ！ 化合物名は無量大
 - ！ 医薬品には商品名もある (ジェネリック名の問題)
- ！ MeSHは 43,874 terms を 8,617 descriptors に
 - ！ ツリーは構造や物性分類に基づき、応用性ない
 - ！ 他に 484,714 terms を 173,049 Suppl. Concepts に
 - ！ 商品名や開発コードまで広く網羅
 - ！ 種類の違いによるタンパク質の相違にまで及ぶ
- ！ LSDIは 19,046 英語表記, 20,247 日本語表記を収録
 - ！ 国内医薬商品名を追加収録した
 - ！ 「アダラート」、「ベシキア」など約5000種類 (英語綴りも収集中)
 - ！ 英語は 16,328語 (86%) が MeSHと重なるが 2,718語が ツリーに収まらず
 - ！ 上位概念にアサイン作業中だが、あまり問題とはならない

2008年1月22日 統合データベース

MeSH	日本語
D000000	メチル
D000001	メチル基
D000002	メチル化
D000003	メチル化剤
D000004	メチル化作用
D000005	メチル化酵素
D000006	メチル化阻害剤
D000007	メチル化阻害作用
D000008	メチル化阻害剤
D000009	メチル化阻害作用
D000010	メチル化阻害剤
D000011	メチル化阻害作用
D000012	メチル化阻害剤
D000013	メチル化阻害作用
D000014	メチル化阻害剤
D000015	メチル化阻害作用
D000016	メチル化阻害剤
D000017	メチル化阻害作用
D000018	メチル化阻害剤
D000019	メチル化阻害作用
D000020	メチル化阻害剤
D000021	メチル化阻害作用
D000022	メチル化阻害剤
D000023	メチル化阻害作用
D000024	メチル化阻害剤
D000025	メチル化阻害作用
D000026	メチル化阻害剤
D000027	メチル化阻害作用
D000028	メチル化阻害剤
D000029	メチル化阻害作用
D000030	メチル化阻害剤
D000031	メチル化阻害作用
D000032	メチル化阻害剤
D000033	メチル化阻害作用
D000034	メチル化阻害剤
D000035	メチル化阻害作用
D000036	メチル化阻害剤
D000037	メチル化阻害作用
D000038	メチル化阻害剤
D000039	メチル化阻害作用
D000040	メチル化阻害剤
D000041	メチル化阻害作用
D000042	メチル化阻害剤
D000043	メチル化阻害作用
D000044	メチル化阻害剤
D000045	メチル化阻害作用
D000046	メチル化阻害剤
D000047	メチル化阻害作用
D000048	メチル化阻害剤
D000049	メチル化阻害作用
D000050	メチル化阻害剤
D000051	メチル化阻害作用
D000052	メチル化阻害剤
D000053	メチル化阻害作用
D000054	メチル化阻害剤
D000055	メチル化阻害作用
D000056	メチル化阻害剤
D000057	メチル化阻害作用
D000058	メチル化阻害剤
D000059	メチル化阻害作用
D000060	メチル化阻害剤
D000061	メチル化阻害作用
D000062	メチル化阻害剤
D000063	メチル化阻害作用
D000064	メチル化阻害剤
D000065	メチル化阻害作用
D000066	メチル化阻害剤
D000067	メチル化阻害作用
D000068	メチル化阻害剤
D000069	メチル化阻害作用
D000070	メチル化阻害剤
D000071	メチル化阻害作用
D000072	メチル化阻害剤
D000073	メチル化阻害作用
D000074	メチル化阻害剤
D000075	メチル化阻害作用
D000076	メチル化阻害剤
D000077	メチル化阻害作用
D000078	メチル化阻害剤
D000079	メチル化阻害作用
D000080	メチル化阻害剤
D000081	メチル化阻害作用
D000082	メチル化阻害剤
D000083	メチル化阻害作用
D000084	メチル化阻害剤
D000085	メチル化阻害作用
D000086	メチル化阻害剤
D000087	メチル化阻害作用
D000088	メチル化阻害剤
D000089	メチル化阻害作用
D000090	メチル化阻害剤
D000091	メチル化阻害作用
D000092	メチル化阻害剤
D000093	メチル化阻害作用
D000094	メチル化阻害剤
D000095	メチル化阻害作用
D000096	メチル化阻害剤
D000097	メチル化阻害作用
D000098	メチル化阻害剤
D000099	メチル化阻害作用
D000100	メチル化阻害剤

シソーラスの応用

1. 対象の抽出

- 表記の整理統一
- 索引の自動作成
- 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

- 情報検索での関連キーワード先読み表示
 - 検索ヘルパー
- 病気と遺伝子
 - 遺伝子名の網羅性が不足
- 医薬品と副作用
 - そのためには主作用や標的を識別する必要

2008年1月22日 統合データベース

対象抽出実験

教科書テキスト に Perl で XML タグをつける

- シノニムを束
ねて1つの統
制語で表す
- カテゴリー毎
に色分け表
示
- タグ妥当性
チューニング
が必要

A screenshot of a web browser displaying a document. The document contains Japanese text with English translations in brackets. The text discusses medical research related to antibodies and diabetes. The browser's address bar shows a file path: 'file:///Users/kamekita/ragpp.html'. The document content includes paragraphs about antibody studies, insulin, and genetic factors like HLA-DQ2 and HLA-DQ8. The text is partially obscured by the sidebar on the left.

2008年1月22日 統合データベース

分子薬理学データベース

！ 遺伝子、医薬品、生体パスウェイ等について情報が整備される中、薬理作用を網羅的に記述するデータベースによって相互リンクが可能

The screenshot displays the JAPIC database interface. On the left, there is a sidebar with navigation options and search filters. The main area shows a detailed network diagram with numerous nodes and connecting lines, representing the relationships between various drugs and their pharmacological actions. The diagram is dense and hierarchical, with many nodes labeled with chemical names and IDs.

2008年1月22日 統合データベース

医薬品の機能分類 (MeSH Pharmacological Action)

The screenshot shows the MeSH Pharmacological Action classification interface. It features a central table with columns for drug names, MeSH terms, and synonyms. The table is organized into two main sections: 'LSD Synonym Table' and 'MeSH Tree'. The 'LSD Synonym Table' lists various LSD-related terms and their synonyms. The 'MeSH Tree' section shows a hierarchical structure of MeSH terms, with arrows indicating the relationship between 'Descriptor' and 'MeSH Tree' at the top, and 'Synonyms' and 'Pharmacological Action' at the bottom.

2008年1月22日 統合データベース

医薬品名と生体分子の関係抽出

医薬品と抄録中で共起する概念

2008年1月22日 統合データベース

医薬品名と生体分子の関係定義(1)

比較的少数の明らかな標的がある場合

2,800種類の医薬品の70%について定義した(未公開)

2008年1月22日 統合データベース

医薬品名と生体分子の関係定義(2)

医薬品名データベース

関係記述が簡単でない場合

医薬品名データベース

定量性も必要

2008年1月22日 統合データベース

The screenshot shows a database interface with two windows. The top window displays search results for 'フルボキサミン' (Fluvoxamine) and a table of relationships. A yellow box highlights the text '関係記述が簡単でない場合' (Relationship description is not simple). The bottom window displays search results for 'オラフィン' (Auranfin) and a table of relationships. A yellow box points to the text '定量性も必要' (Quantitative information also required).

クリニカルインフォマティクス

! バイオインフォマティクス

- 実験データの蓄積
- 法則性の発見
- 知識のネットワークによる新しい知識発見
- ゲノム創薬, トキシコゲノミクスへの応用

! クリニカルインフォマティクス

- 臨床データや医療テキストの蓄積
- 人間の知識の体系化
- 人間に適切なアドバイスを与える助言システム
- 人間のミス在未然に防ぐ警告システム

2008年1月22日 統合データベース

実際の診療録レビュー

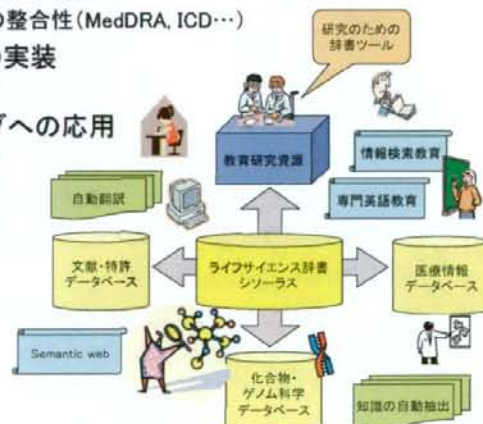
! 80代男性、経口摂取できなくなり入院。入院前から仙骨部にⅡ度の褥瘡あり。入院前に栄養状態のコントロール不良、入院後に栄養状態のコントロール不良、体交が適切に行われておらず、褥瘡がⅢ度に進行。デブリードマンを施行。

! 慢性硬膜下血腫除去術後で、ほぼ全介助の80代女性。発熱あり入院となり、誤嚥性肺炎の疑いにてセフメタゾンを投与。肺炎症状は軽快したが、偽膜性腸炎による軟便により、入院が延長した。

2008年1月22日 統合データベース

今後の展開予定

1. より頑強なシソーラスの構築
 - 他のシソーラスとの整合性 (MedDRA, ICD…)
2. 情報ポータルへの実装
 - 教育現場での評価
3. テキストマイニングへの応用
 - 臨床情報での評価



2008年1月22日 統合データベース

分子薬理学的知識を記述する 新たな三項関係データベースの開発

京都大学大学院 薬学研究科
伊藤 悦子, 金子 周司
skaneko@pharm.kyoto-u.ac.jp
2007年11月24日
医療情報学会(神戸)



Who are we ?

分子薬理学

電子辞書構築

京都大学大学院薬学研究科 生体機能科学分科

LIFE SCIENCE DICTIONARY PROJECT

2007年11月24日 医療情報学会

本研究の背景と目的

！ 遺伝子、医薬品、生体パスウェイ等について情報が整備される中、薬理作用を網羅的に記述するデータベースが存在しない。

→ 医薬品の分子作用点を記述する

The screenshot displays two browser windows. The left window shows the 'JAPIC 医薬品データベース' (JAPIC Drug Database) interface with various search and navigation options. The right window shows a detailed biological pathway diagram with numerous nodes and connecting arrows, representing molecular interactions. Below the diagrams, the text '2007年11月24日 医療情報学会' (November 24, 2007, Japanese Society for Medical Information) is visible.

方法

！ 医薬品名の収集

- ！ 英語: MeSH terms, ATC, KEGG Drug
- ！ 日本語: LSD, 日本医薬品一般名 (JAN), JAPIC医薬品集
- ！ 約18,000語を名寄せによって2,842語の統制語に整理
 - ！ MeSHと2,545語が一致, ATCと2,599語が一致

！ 生体分子名の収集

- ！ 英語・日本語対訳: LSDより15,595語を選択
- ！ MeSH termを用いて異表記を吸収
- ！ 遺伝子名はRefSeqより取り込み

！ 関係整理

- ！ PubMedコーパス (1.6 GB) での共起関係をカウント
- ！ 教科書の知識などから、人手によって分類、入力
- ！ 少数の記述子で関係を定義

2007年11月24日 医療情報学会

MeSH と LSD の統合による同義語テーブルの作成

2007年11月24日 医療情報学会

医薬品名と生体分子の関係抽出

2007年11月24日 医療情報学会

医薬品名と生体分子の関係定義(1)

比較的小数の明らかな標的がある場合

2007年11月24日 医療情報学会

医薬品名と生体分子の関係定義(2)

関係記述が簡単でない場合

2007年11月24日 医療情報学会