

28PW-am 医薬品名の同義語辞書と自動分類タガーの開発

074

○金子 周司¹, 大武 博², 藤田 信之³

¹京大院薬, ²京都府立医大, ³製品評価技術基盤機構



① テキストで記述される医薬品情報は、医薬品名称だけを見て商品名、塩を含む一般名、有効成分名があり、さらに効能や適応症、使用上の注意、薬物動態、薬効薬理など、さまざまな用語で記述されている。しかし現在の記述には統制語がもちいられていないため、テキスト中の医薬品名を正しく抽出あるいは検索するにあたっては、同義語の処理と概念の包含関係を整理し、英語での表記にも耐える網羅的かつ頑強な辞書を意用する必要がある。

② 本研究において、我々は様々な医療テキストから正しく医薬品名を抽出する辞書とツールを完成させることを目標にして、医薬品名辞書とそれを利用する自動分類タガーを開発した。医薬品や化学物質に対して付与される名称については、日本医薬品名一般名称(JAN), JAPIC医療用医薬品集, MeSH, ATC, KEGGおよびライフサイエンス辞書等から商品名まで含めて抽出した。

③ これらを整理することによって重複のない39,105種類の医薬品名を3,482種類の統制語としての有効成分名に関連づける同義語(シノニム)辞書を構築した。また、統制語に薬効分類情報(Pharmacological Action)を付与することで、テキスト中の医薬品名を薬効ごとに自動分類するタガーをPerlスクリプトで制作した。

④ 実際にJAPIC医療用医薬品集テキストやPubMedから選んだ臨床研究抄録に対して、今回のタガーによってタグ付けを行ったところ、多義性のある略語以外は生理活性のある化合物の名称を拾うことができた。

⑤ これらにより薬効分類と疾患の共起関係などを解析することが可能となり、用語の統一あるいは表記のゆれの吸収が可能になるのみならず、医療情報のテキストマイニングに役立つと考えられる。

③ 本研究で完成させた医薬品名の同義語テーブル

英語および日本語をカバーする39,105種類の医薬品名を3,482種類の統制語(Descriptor)に集約した(上)。さらに、薬効分類(Pharmacological Action)を付与し、その同義語も収集した(右)。

冊子体もありますので、ご覧ください。

- 用語法 Terminology での多様性の解決
 - 「タミフル」を表す term
 - オセルタミビル シノニム
 - リン酸オセルタミビル シノニム
 - Oseltamivir phosphate オルトロジー
- 語彙の階層性や関連性を、どう構図に理解させるか
 - オセルタミビルは抗ウイルス薬である
 - A型およびB型インフルエンザウイルス感染症に用いられる
 - 作用機序はニューラミニダーゼ阻害である

① 医薬品情報の現実

Level	Descriptor
D00000	医薬品名
D00001	医薬品名
D00002	医薬品名
D00003	医薬品名
D00004	医薬品名
D00005	医薬品名
D00006	医薬品名
D00007	医薬品名
D00008	医薬品名
D00009	医薬品名
D00010	医薬品名
D00011	医薬品名
D00012	医薬品名
D00013	医薬品名
D00014	医薬品名
D00015	医薬品名
D00016	医薬品名
D00017	医薬品名
D00018	医薬品名
D00019	医薬品名
D00020	医薬品名
D00021	医薬品名
D00022	医薬品名
D00023	医薬品名
D00024	医薬品名
D00025	医薬品名
D00026	医薬品名
D00027	医薬品名
D00028	医薬品名
D00029	医薬品名
D00030	医薬品名
D00031	医薬品名
D00032	医薬品名
D00033	医薬品名
D00034	医薬品名
D00035	医薬品名
D00036	医薬品名
D00037	医薬品名
D00038	医薬品名
D00039	医薬品名
D00040	医薬品名
D00041	医薬品名
D00042	医薬品名
D00043	医薬品名
D00044	医薬品名
D00045	医薬品名
D00046	医薬品名
D00047	医薬品名
D00048	医薬品名
D00049	医薬品名
D00050	医薬品名
D00051	医薬品名
D00052	医薬品名
D00053	医薬品名
D00054	医薬品名
D00055	医薬品名
D00056	医薬品名
D00057	医薬品名
D00058	医薬品名
D00059	医薬品名
D00060	医薬品名
D00061	医薬品名
D00062	医薬品名
D00063	医薬品名
D00064	医薬品名
D00065	医薬品名
D00066	医薬品名
D00067	医薬品名
D00068	医薬品名
D00069	医薬品名
D00070	医薬品名
D00071	医薬品名
D00072	医薬品名
D00073	医薬品名
D00074	医薬品名
D00075	医薬品名
D00076	医薬品名
D00077	医薬品名
D00078	医薬品名
D00079	医薬品名
D00080	医薬品名
D00081	医薬品名
D00082	医薬品名
D00083	医薬品名
D00084	医薬品名
D00085	医薬品名
D00086	医薬品名
D00087	医薬品名
D00088	医薬品名
D00089	医薬品名
D00090	医薬品名
D00091	医薬品名
D00092	医薬品名
D00093	医薬品名
D00094	医薬品名
D00095	医薬品名
D00096	医薬品名
D00097	医薬品名
D00098	医薬品名
D00099	医薬品名
D00100	医薬品名

② ライフサイエンス辞書のシソーラス化とオートロジーへの展開

MeSH Descriptor (左) との対応付けにより3万語の専門用語ツールを完成させる。さらに共起する関連用語を用いてオートロジーを構築する。

- A. 解剖部位名
- B. 生物名
- C. 病名, 症候名
- D. 医薬品名, 生体内物質
- E. 方法, 技術

本発表では、医薬品名のシノニム作成とその応用について紹介する。

出典	内容	語数
MeSH	Substance table より 商品名および一般名(英語)	20,405
JAPIC医療用医薬品集	商品名および一般名(日本語)	18,109
JAN	一般名(日本語)	2,575
ATC/DDD	一般名(英語)	4,197
KEGG DRUG	一般名(英語)	5,286

これらから重複を排除し、MeSH Substance (6,545語)に割り当てた。

(結果の詳細) JAPICに収録された18,109種類の商品名のうち、13,619種類(75%)が割り当て可能であった。一般名もほぼ同じ割合であった。残された医薬品には、漢方製剤、生薬、配合剤などが挙げられる。

④ 医薬品名シノニムを利用した教科書テキストへの自動タグ付け

Perlで制作した解析スクリプト(左)を用いることによって、Goodman & Gilman 薬理学教科書テキスト(100万単語以上)に対して、パソコン(Apple Mac Pro)で70秒以内にすべての医薬品名にタグを付けられた(右)。

パソコンで実際のデータをお見せします。

⑤ 考察と展望

無料のオンライン辞書WebLSDへの実装により、医薬品情報ポータル構築が可能になった。

日本語テキストタガーと薬効分類による自動分類スクリプトを開発中である。

実際の医療テキストを材料とするマイニング実験の共同研究募集中!