

アノテーションT辞書仕様書 (2021年7月)

有限会社ティ辞書企画

〒349-0111 埼玉県蓮田市東 3-1-25

TEL:048-769-6662 FAX:048-765-2627

1. 「アノテーションT辞書」作成の目的

これまで「医薬 T 辞書」は、検索用辞書としてのご利用が主でしたが、最近文書解析のための用語抽出等に利用いただくケースが増えてきました。そこで、従来の医薬同義 T 辞書と展開型医薬 T 辞書を組み合わせ、ニーズに合わせた粒度で用語を抽出できるように考案した辞書が「アノテーション T 辞書」です。

2. 従来の辞書との違い (従来の T 辞書をご利用の皆様へ)

「アノテーション T 辞書」は文書中のより多くの語に意味付けを行うことができるよう、検索用辞書には含まれていない 2 語の略語や、医療機器名、レジメン名などより多くの用語を収録しています。また今後、部位名なども追加予定としています。また、新たに医薬品販売名コード (7 桁) を同義語の中を含めました。必要に応じてご利用ください。

従来の「医薬同義 T 辞書」は、コードが固定長でしたが、「アノテーション T 辞書」はコードが不定長です。ティ辞書企画では、コードを用語に代わるある程度意味をもつ文字列としてご利用いただくことで、より利用価値が高まると考えています。

3. 収録語

医療・薬学関連の語を収録し、次の通りコード頭文字のアルファベット 1 文字で分類しています。

コード 頭文字	分類	収録内容	例
L	臨床検査関連語	検査対象物質 試験・検査	炎症マーカー、抗核抗体、ミオグロビン インスリン負荷試験、マイクロサテライト不安定性検査、心血管造影、心電図
M	微生物名 ※主にヒトへの病原性が問題となるもの	細菌名 ウイルス名 寄生虫名	ジフテリア菌、リケッチア・チフィ ロタウイルス、ポリオウイルス 熱帯熱マラリア原虫、イエダニ

D	医薬品名	一般名 商品名 略語、レジメン名等	アズレン、アシクロビル アズノール、ゾビラックス AZL、ACV、SP療法
E	薬効名	薬効分類	血圧降下薬、免疫チェックポイント阻害剤、抗PD-1抗体製剤
S	疾患名	病名・症状 副作用用語 検査値異常	ドラベ症候群、筋損傷、鼻うっ血、フレイル、肥満 ワクチン副反応、ステロイド皮膚 腫瘍マーカー上昇、過敏性テスト陽性、 γ -GTP上昇

コード頭文字「G」は上記以外の語で、主に下記の用語を収録しています

コード	内容	例
G050010	実験動物	チンパンジー、ウサギ、ラット
G050020	剤形	カプセル、クリーム、ドライシロップ
G050030	試験方法	無作為化対照比較試験、多施設共同試験、オープン試験
G050040	投与関連	頓服、維持量、動脈注射
G050070	薬物動態	グルクロン酸抱合、組織内移行、バイオアベイラビリティ
G050080	診療科	眼科、ペインクリニック、渡航者外来
G051010	機関・組織	日本薬剤師研修センター、日本医師会、国立感染症研究所
G051030	医療従事者	ソーシャルケースワーカー、理学療法士、かかりつけ医
G200000	手術処置	胃切除術、電氣的除細動、中心静脈栄養法
G500000	医療機器	パルスオキシメータ、中心静脈カテーテル、磁気共鳴装置

※コード G は今後さらに充実する予定で、更新にとまないコードも変更されます。それ以外のコードは基本的には変更はありませんが、概念の統一や分解により変更される場合があります。(6. ご利用上の注意事項 2) バージョン管理もご参考ください。)

4. ファイルおよびファイルフォーマット

アノテーション T 辞書は 2 つのファイルで構成されています。

1) 同義語ファイル (ant_synyyyymmdd)

用語コードとその中にふくまれる同義語群を収録しています。

項目	内容
コード	アルファベット 1 桁 + 数値 (半角)。MAX19 桁
区分	区分 1 : 代表語 2 : 同義語 ※ 6 : 略語 E : 英語 Y : 販売名コード 7 桁
用語	全角 / 半角混合 (英数字記号は半角)
コード桁	コードの桁数 先頭が同じコードはコード桁が多いほど同じ概念の下層語となります。

※ 医薬品の場合、一般名に対して商品名は同義語として収録しています。

2) 用語関連ファイル (antyyyymmdd)

コードの包含関係を示したファイルです。あるコード (検索元コード) に対して、その中に包含される用語のコードを「対象コード」として示しています。

項目	内容
検索元コード	アルファベット 1 桁 + 数値。MAX19 桁
検索元代表語	検索元コードの代表語 (参考までに示しています)
対象コード	検索元コードに包含されるその他のコード
対象コード代表語	対象コードの代表語 (参考までに示しています)

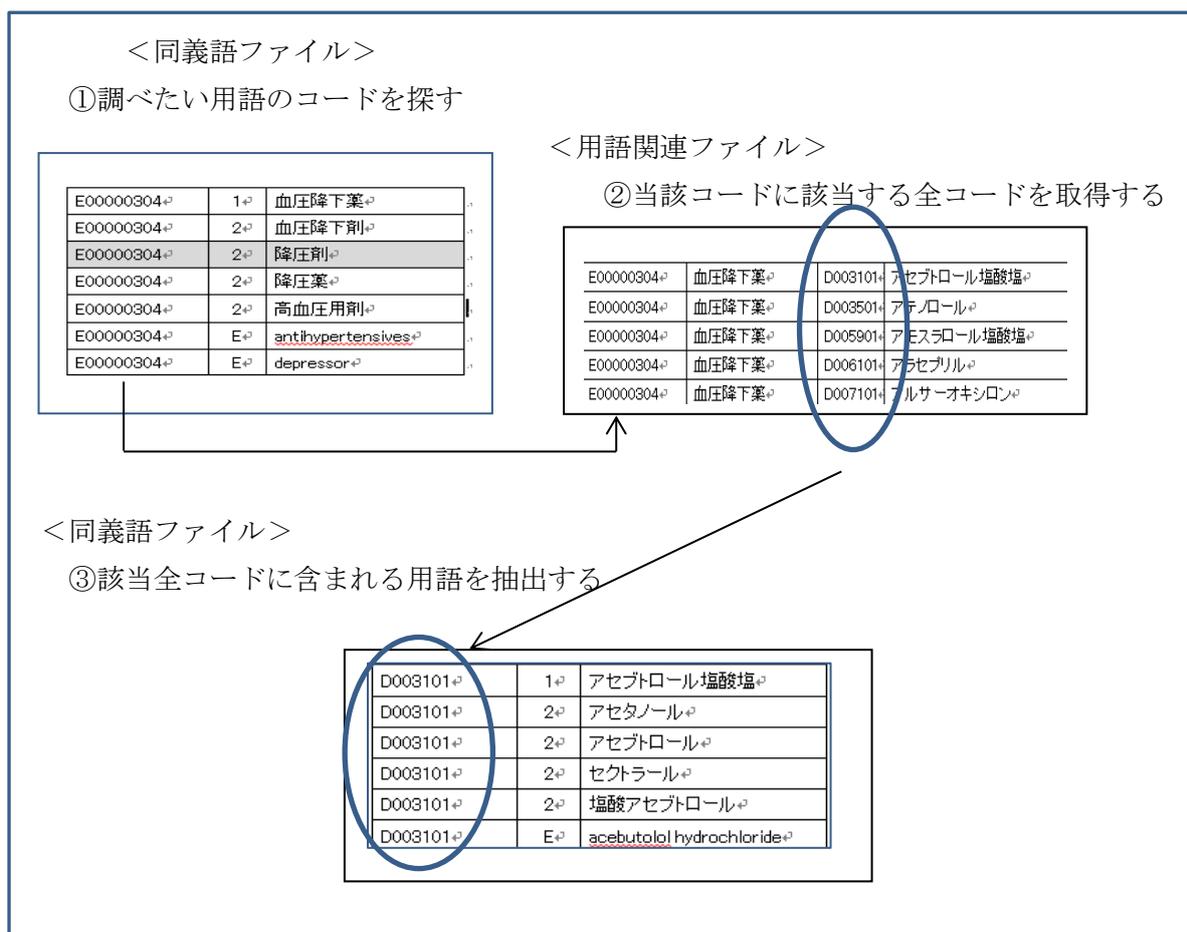
本来はコード間の関係のみを示すファイルですが、参考のため代表用語を表示しています。

5. ファイル間の関係・ご利用方法

基本的なご利用方法として、用語の検索、用語に対してのコード付与には「同義語ファイル」をご利用いただき、検出したい語を下層概念まで含めて抽出する場合には「用語関連ファイル」をご利用ください。

<ご利用例1>

降圧剤に含まれる語を全て抽出する場合



詳細：

①「同義語ファイル」で<降圧剤>を検索し、コードを見つけます

E00000304	1	血压降下薬
E00000304	2	血压降下剂
E00000304	2	降圧剤
E00000304	2	降圧薬
E00000304	2	高血压用剂

E00000304	E	antihypertensives
E00000304	E	depressor

コード「E00000304」が得られます

②次に「用語関連ファイル」で検索元コード「E00000304」を検索し、降圧剤に含まれるすべての対象コードを取得します。

検索元コード	検索元代表語	対象コード	対象コード代表語
E00000304	血圧降下薬	D003101	アセブトロール塩酸塩
E00000304	血圧降下薬	D003501	アテノロール
E00000304	血圧降下薬	D005901	アモスラロール塩酸塩
E00000304	血圧降下薬	D006101	アラセプリル
E00000304	血圧降下薬	D007101	アルサーオキシロン

以下省略（E00000304 は 100 件以上あります）

③「用語関連ファイル」より得られた対象コードを元に、改めて降圧剤に含まれる全用語を「同義語ファイル」から抽出します。

この際、区分を利用して、英語（区分 E）を除いたり、略語（区分 6）や販売名コード 7 桁（区分 Y）を除くこともできます。

コード	区分	用語
D003101	1	アセブトロール塩酸塩
D003101	2	アセタノール
D003101	2	アセブトロール
D003101	2	セクトラール
D003101	2	塩酸アセブトロール
D003101	E	acebutolol hydrochloride
D003501	1	アテノロール
D003501	2	アテロール
D003501	2	テノーミン
D003501	E	atenolol

一部のみ表示

<ご利用例 2>

文書にコードを付与することにより、より高度な文書解析を行う。

元文書

- ・ 血圧降下薬服用中にみられた γ -GTP の上昇
- ・ アセプトロールによる肝機能異常

① 「同義語ファイル」を用いてコード化を行います

- ・ **E00000304** 服用中にみられた **S2412131113**
- ・ **D003101** による **S2412**

② 「用語関連ファイル」を利用して次のような文書解析が可能です

降圧剤と症状（病名）が近くに記載してある文書の抽出

降圧剤として E00000304、D003101 が検出できます

症状は「S」で始まるコードが付与されている語で検出できます。

さらに、「肝機能異常」(S2412)には γ -GTP の上昇 (S2412131113) も含まれることがわかるため、この2つの例文は降圧剤の近くに肝機能異常について記載されている文書として抽出可能です。

<ご利用例 3 : 薬剤における「〇〇類」の利用>

薬剤の塩の違いや合剤も含めて抽出するために「〇〇類」の語を設定しています。

「用語関連ファイル」で「アムロジピンベシル酸塩」（単独）と「アムロジピン類」の違いは次のようになっています。

検索元コード	検索元代表語	対象コード	対象コード代表語
D200901	アムロジピンベシル酸塩	D200901	アムロジピンベシル酸塩
D200900	アムロジピン類	D200900	アムロジピン類
D200900	アムロジピン類	D200901	アムロジピンベシル酸塩
D200900	アムロジピン類	D200951	アムロジピンベシル酸塩/アトルバスタチンカルシウム水和物

D200900	アムロジピン類	D222152	カンデサルタン シレキセチル/アムロジピンベシル酸塩
D200900	アムロジピン類	D227652	バルサルタン/アムロジピンベシル酸塩
D200900	アムロジピン類	D234952	テルミサルタン/アムロジピンベシル酸塩
D200900	アムロジピン類	D234953	テルミサルタン/アムロジピンベシル酸塩/ヒドロクロロチアジド
D200900	アムロジピン類	D246151	イルベサルタン/アムロジピンベシル酸塩
D200900	アムロジピン類	D255951	アジルサルタン/アムロジピンベシル酸塩

D200900 アムロジピン類 の対象コードにはアムロジピンを含む合剤も設定してあるため、同義語ファイルとの連携で次のような合剤の商品名が抽出できます。

コード	区分	用語
D200901	2	アムロジン
D200901	2	ノルバスク
D200951	2	カデュエット
D200951	2	アマルエット
D222152	2	ユニシア
D222152	2	カムシア
D227652	2	エックスフォージ
D227652	2	アムバロ
D234952	2	ミカムロ
D234952	2	テラムロ
D234953	2	ミカトリオ
D246151	2	アイミクス
D246151	2	イルアミクス
D255951	2	ザクラス

同義語ファイルには販売名コード7桁の情報もありますので、販売名コードを付与されたデータの解析に利用できます。

コード	区分	用語
D200901	Y	2171022
D200951	Y	2190101
D200951	Y	2190102
D200951	Y	2190103
D200951	Y	2190104
D222152	Y	2149116
D227652	Y	2149114

D234952	Y	2149117
D234953	Y	2149122
D246151	Y	2149118
D255951	Y	2149121

6. コード桁（階層構造）

コードの体系には2系統ありますが、中間階層が無い場合も多く、階層としてご利用いただく場合は、該当の用語コードと前方一致するコード桁の多い語が下層語となります。

1) コード頭文字が L、M、S の場合

3桁コードが大分類であり、以後2桁ずつ多くなり下層語となります。最高17桁。

例

S18	感覚器障害
S1812	眼疾患
S181211	視覚障害
S18121111	視野異常
S1812111111	視野狭窄
S181211111112	視野欠損

2) コード頭文字が D、E、G の場合

7桁コードが大分類であり、以後2桁ずつ多くなり下層語となりますが、中間階層が無い場合が多くなっています。最高19桁。

例

E000001	神経系作用薬
E00000103	抗不安薬
E000001030105	セロトニン作動薬

7. ご利用上の注意事項

1) 用語の重複

代表語には重複はありませんが、同義語、特に略語には重複があります。（文字列は同じですが、概念が異なるためコードは別々となります。）

例

コード	区分	用語
S21131711	6	FMD
S61121413	6	FMD
L161221	6	FMD

線維筋異形成症(fibromuscular dysplasia)の略。

口蹄疫(foot-and-mouth disease)の略

血流依存性血管拡張反応(flow-mediated vasodilation)の略。

2) バージョン管理

「アノテーション T 辞書」は年1回更新予定です。更新により用語のコードが変更になる場合があります。特にコード頭文字「G」の用語は今後増やしていく予定であるため大きくコードが変更される可能性があります。また用語の関連性が変更される場合もありますので、何年版の辞書を利用された処理であるかのバージョン管理をお勧めいたします。

3) 同一成分で適応が異なる医薬品に関して

同一成分で投与経路や適応が異なる製剤が有る場合、成分名でそれを分別することができません。

このため、薬効分類からの検索では、意図しない結果が含まれることがあります。

例. アスピリンは解熱鎮痛薬と血液凝固阻止剤があります。このため下記のような系統の薬効分類に含まれます。例えば、**解熱鎮痛消炎薬**としてアスピリンを検出した場合、**血小板凝集阻害薬**として使われている場合も含まれてしまいます。

検索元コード	検索元代表語	対象コード	対象コード代表語
E000001	神経系作用薬	D001701	アスピリン
E00000105	解熱鎮痛消炎薬	D001701	アスピリン
E0000010501	非ステロイド性消炎鎮痛薬	D001701	アスピリン
E000001059101	サリチル酸系非ステロイド性消炎鎮痛薬	D001701	アスピリン
E000012	血液系用剤	D001701	アスピリン
E00001212	血液凝固阻止薬	D001701	アスピリン
E0000121201	経口抗凝血剤	D001701	アスピリン
E000012129102	血小板凝集阻害薬	D001701	アスピリン

また、同一成分で投与経路が異なる製剤も分別することができません。

外用抗ウイルス薬を検出した場合、全身投与のアシクロビルも含まれてしまいます。

