

SIP3医学概念・知識連結データベースとは

SIP3医学概念・知識連結データベース構築チーム
2024年9月20日

目的

本データは、既存の医療・医学関連のリソースをまとめた大規模な医学概念・知識連結データベースです。本データは、電子カルテ（EHR）や患者記録（PHR）など様々なテキストから構造化された医療知識を抽出する情報収集・分析の基盤として応用することを目指して開発されています。

本データの主な特徴は以下の3つです。

- 人間用ではなくAI用
 - 医療用語の言い換えや、外部のターミノロジーへの変換を目的とした、AI用のデータです。
 - 人間用の辞書のように、意味や正しい用語を調べるためのデータではありません。
- 紐付け（リンク）のデータ
 - 医療テキスト内に出現する語の中で、一定以上の頻度で出現する語はすべて掲載しています。
- 構築中を常態とする
 - 各エントリに付与された頻度・信頼度情報を利用してメンテナンスの優先度を把握できます。

本データは4つのサブデータから構成されます。

1. 病名データ: 症状や病名を収載。ICD-10や11, 標準病名などに紐付けが可能。
2. 医薬品データ: 医薬品名を収載。一般名, ATCコード, KEGGコードなどに紐付け可能。
3. 部位データ: 人体部分表現を収載。代表的な解剖学辞書の表現に紐付け可能。
4. 検査データ: 検査表現を収載。代表的な臨床検査項目の分類に紐付け可能。

仕様

属性情報

	ラベル	概要	タイプ
基本情報	ID	出現形に対応する一意に割り当てられたID.	文字列
	出現形	医療テキストから抽出した, 用語の出現形(ユニーク).	文字列
	出現形よみ	出現形のよみ	文字列
	正規形	<p>出現形を正規化した表記. 本データは, 既存リソースへのリンクが目的であり, 正しい正規形や標準的な表現を提案するものではないことに注意されたい. あくまで, ビューアー上での表示や, 概念の理解の補助を目的に正規形を設けている.</p> <p>サブデータごとに正規形のベースは異なる. それぞれの正規形のベースは以下の通りである.</p> <ul style="list-style-type: none">● 病名データ: 標準病名マスター病名● 医薬品データ: 一般名● 部位データ: ICD-11のエクステンションコードをベースとした表現● 検査データ: 医師国家試験出題基準の「主な検査項目の表記」をベースにした表現 <p>出現形に対して複数の正規形が該当する場合, 最大3つまでを半角セミコロン (;) で接続し列挙する.</p>	文字列
	TREE	最大4桁の記号による分類. ビューアー使用時のツリー構造の表示に用いる.	文字列
翻訳情報	出現形EN	出現形の英語訳.	文字列
	出現形ZH	出現形の中国語訳.	文字列
関係情報	REL:IS-A	上位下位関係.	文字列

	REL:PART-OF	部分全体関係.	文字列
	REL:CAUSAL	因果関係.	文字列
その他	用語難易度	用語の難易度	整数
	YEAR	エントリの収集年度	整数

リンク先

リンク先のリソースによって、ライセンスの取り扱いが異なります。本データでは、最小セットとしてライセンスを持たない利用者でも利用可能なリソースへのリンクを収録しています。これらのリソースへのリンク情報は、どの利用者でも利用できます。加えて、個別リソースのライセンスを持つ利用者には、そのリソースへのリンク情報を加えたデータを提供します。

- 最小セットとして収録するリソース
 - TO:SNOMED-CT-IPS
 - TO:ICD-10
 - TO:HPO
 - TO:薬効分類
 - TO:一般名処方マスター
 - TO:JLAC10コード

リソース名	概要	リンク	病	薬	検	部
TO:日本医学会医学用語辞典	日本医学会が出版している日英辞書.	見出し語	✓	✓	✓	✓
各種厚労省標準マスター関連用語集	MEDISが公開しているマスター.		✓		✓	
TO: CLIDAS	循環器データベースへ用語紐付け.		✓			

TO: J-CaseMap	自治医科大・東大が開発している辞書.		✓			
TO:Lilak	医学知識連結DB. 東京大学が開発したオントロジー.	疾病名または部位コード	✓			✓
TO:薬効分類	総務省が定めた日本標準商品分類番号にしたがって分類された, 医薬品の効能に関する分類.	分類番号		✓		
TO:JAPICID	医薬品に関するコードへの紐付け.			✓		
TO:KEGGエントリ	京都大学が管理している医薬品DBへのID.	エントリID		✓		
TO:PubChem	アメリカNCBIが管理する化学物質データベースのID.	SID		✓		
TO:ATC	WHOによる薬品の分類コード.	ATCコード		✓		
TO:JLAC10	日本臨床検査医学会が制定した臨床検査項目の分類コード(第10回改訂版)				✓	
TO:JLAC11	日本臨床検査医学会が制定した臨床検査項目の分類(第11回改訂版)					
TO:ICD-10	WHOが制定した国際的な病名分類(第10回改訂版).	ICDコード	✓			
TO:ICD-11	WHOが制定した国際的な病名分類第11回改訂版).	ICDコード	✓			
TO:HPO	Human Phenotype Ontology (HPO) 研究用のフェノタイプ用語集.	ID	✓			
TO:SNOMED-CT-IPS	SNOMED-CTの縮小版. SNOMED-CT/International Patient Summary Terminology	ID	✓			
TO:MedDRA/J	医薬品規制調和国際会議(ICH)が制定した医薬用語集の日本語版.	見出し語	✓			
TO:UMLS	アメリカNLMが公表している, 医療概念のメタソース. CUIを収載.	ID	✓	✓	✓	✓
TO:一般名処方マスタ	厚生労働省が策定している医薬品の一般名処方の標準的な記載.	一般名コード		✓		

信頼度

信頼度	概要
S	信頼度最大. 既存のリソースと照合済み.
A	信頼度高. 人間3名以上がチェック済.
B	信頼度中. 人間2名がチェック済.
C	信頼度低. 人間1名がチェック済.
D	信頼度低. 自動推定された結果.

見方

実際のデータをもとに、掲載されている情報の見方を説明します。

以下の図は、病名データをExcelで開いた場合の表示画面です。本データでは、一行一エントリとして、属性情報・リンク情報を各列に設定し、個別のエントリの値を入力しています。

例えば、赤枠で囲んだ行ID5の出現形「高血圧」のよみがなは同じ行のE列(出現形よみ)に記載されている「こうけつあつ」です。また、同じ行のF列(出現形よみ_flag)の値が「B」であることから、この出現形よみはすでに人間2名がチェック済みであることがわかります。このように、本データでは属性情報およびリンク情報に加えて、個別に各情報に信頼度のフラグを付与しています。

A	B	C	D	E	F	G	H	I	J	K	L	M
行ID	ID	出現形	出	出現形よみ	出現形よみ	正規形	正規形 f	正規形よ	正規	TREE	TREE fla	出
0	fe55650df3	副作用		ふくさよう	C		-1 C				-1 C	sid
1	17eced89ff	合併症		がっべいしよう	B		-1 B				-1 B	com
2	b41877e5d5	糖尿病		とうにょうびよう	S	糖尿病	S			E14	S	dia
3	49aec2159a	再発		さいはつ	C		-1 C				-1 C	rec
4	4a53019649	癌		がん	S	癌	S			C809	S	can
5	b9c56cc645	高血圧		こうけつあつ	B	高血圧症	B			I10	B	Hyp
6	2fb92977db	痛み		いたみ	B	疼痛	A			R529	A	pai
7	4324a7a86c	腫よう		しゅよう	C		-1 C				-1 C	tum
8	d6924ed9ca	心不全		しんふぜん	S	心不全	S			1509	S	hea
9	e8182c7776	感染症		かんせんしよう	B		-1 B				-1 B	inf
10	9139b0afb9	出血		しゅっけつ	S	出血	S			R58	S	hem
11	cf46a0e39e	腫瘍		しゅよう	S	腫瘍	S			D489	S	tum
12	55cc77a8bb	感染		かんせん	C		-1 C				-1 C	inf
13	c7332fe2bd	とう痛		とうつう	C	疼痛	C			R529	C	pai
14	f159aa5214	脳梗塞		のうこうそく	S	脳梗塞	S			1639	S	cer
15	fb093bf93f	炎症		えんしよう	B		-1 B				-1 B	inf
16	d76afb2069	障害		しょうがい	C		-1 C				-1 C	dis
17	3f6d5e353a	肺癌		はいがん	S	肺癌	S			C349	S	lun
18	41c9ce71cf	2型糖尿病		にがたとうにょうびよう	S	2型糖尿病	S			E11	S	typ
19	7170bd6bf3	疼痛		とうつう	S	疼痛	S			R529	S	pai
20	f14ce6c29f	骨折		こっせつ	S	骨折	S			T1420	S	fra
21	d7f5001c42	乳癌		にゅうがん	S	乳癌	S			C509	S	bre
22	452e73e41e	発熱		はつねつ	S	発熱	S			R509	S	fev
23	b4aad70e4b	I		あい	C	[ERR]	C			K051	C	
24	0a2cdd33f4	胃癌		いがん	S	胃癌	S			C169	S	gas
25	150c5a382f	病変		びょうへん	B		-1 C				-1 C	les
26	83e098c49e	肥満		ひまん	B	肥満症	A			E669	A	obe
27	80165378ca	肝硬変		かんこうへん	B	肝硬変症	B			K746	B	liv
28	fe41ea81f3	転移		てんい	B	転移性腫瘍	B			C799	B	met
29	0f1cb05516	肺炎		はいえん	S	肺炎	S			J189	S	pne
30	cccaaf177f	貧血		ひんけつ	S	貧血	S			D649	S	ane
31	35147056a2	肝細胞癌		かんさいぼうがん	S	肝細胞癌	S			C220	S	hep
32	7240d5c573	意識障害		いしきしょうがい	S	意識障害	S			R402	S	dis
33	4c723dfb89	腫りゅう		しゅりゅう	C	腫瘍	C			R229	C	mas
34	9fc8c8fb1d3	呼吸困難		こきゅうこんなん	S	呼吸困難	S			R060	S	dys
35	776931bb51	統合失調症		とうごうしつちようしよう	S	統合失調症	S			F209	S	sch
36	5eb7bb5ba4	浸潤		しんじゆん	C		-1 C				-1 C	inv
37	ee30f070a7	不整脈		ふせいみゃく	S	不整脈	S			1499	S	arr
38	e1595234a7	有害事象		ゆうがいじしよう	B		-1 B				-1 B	adv
39	ed67e16796	虚血性心疾患		きょけつせいしんしっかん	S	虚血性心疾患	S			1259	S	isc

以下の図は、メンテナンス用のWEBアプリケーション「KnitBoat」でデータを開いた際の画面です。KnitBoatでデータを閲覧するには、画面左側の分類ツリーを選択し、次に正規形を選択します。さらに出現形に表示されるエントリを選択することで、そのエントリの詳細を見ることができます。

例えば、以下の図は画面左側の分類で「I10 本態性(原発性<一次性>)高血圧(症)」を選択し、次に正規形で「高血圧症」を選択、最後に出現形で「高血圧症」を選択した場合に表示される画面です。画面右側の「詳細」には出現形「高血圧症」のエントリに付与された各種の情報と信頼度が表示されています。

病名

検索条件 OR AND

作業担当 作業中 コメントあり

確認担当 確認中 信頼度

テキストで絞り込み

章	コード	コード名
>	I	感染症及び寄生虫症 (A00 - B99)
>	II	新生物<腫瘍> (C00 - D48)
>	III	血液及び造血系の疾患並びに免疫機構の障害 (D50 - ...)
>	IV	内分泌, 栄養及び代謝疾患 (E00 - E90)
>	V	精神及び行動の障害 (F00 - F99)
>	VI	神経系の疾患 (G00 - G99)
>	VII	眼及び付属器の疾患 (H00 - H59)
>	VIII	耳及び乳様突起の疾患 (H60 - H95)
>	IX	循環器系の疾患 (I00 - I99)
>	100-102	急性リウマチ熱 (I00 - I02)
>	105-109	慢性リウマチ性心疾患 (I05 - I09)
>	110-115	高血圧性疾患 (I10 - I15)
>	110	本態性(原発性<一次性>)高血圧(症)
>	111	高血圧性心疾患
>	112	高血圧性腎疾患
>	113	高血圧性心腎疾患
>	115	二次性<続発性>高血圧(症)
>	120-125	虚血性心疾患 (I20 - I25)
>	126-128	肺性心疾患及び肺循環疾患 (I26 - I28)
>	130-152	その他の型の心疾患 (I30 - I52)

正規形

検索条件 OR AND

作業担当 作業中 コメントあり

確認担当 確認中 信頼度

テキストで絞り込み

正規形	コメ...	作業	確認	信頼度
高血圧症				C
本態性高血圧症				C
収縮期高血圧症				C
高血圧合併原糖尿病				C
悪性高血圧症				C
糖尿病				C
血圧異常				C
境界型高血圧症				C
高血圧性緊急症				C
肺動脈性肺高血圧症				C
白衣高血圧				C
腎性高血圧症				C
若年高血圧症				C
高血糖症				C
高脂血症				C
本態性血小板血症				C
肺高血圧症				C
低レニン性高血圧症				C
高レニン性高血圧症				C
低血圧症				C

出現形

検索条件 OR AND

作業担当 作業中 コメントあり

確認担当 確認中 信頼度

テキストで絞り込み

出現形	コメ...	作業	確認	信頼度
高血圧				C
高血圧症				C
軽症高血圧				C
早朝高血圧				C
仮面高血圧				C
高血圧緊急症				C
治療抵抗性高血圧				C
高齢高血圧				C
高齢者高血圧				C
血圧高値				C
HT				C
未治療高血圧				C
糖尿病合併高血圧症				C
軽症高血圧症				C
重症高血圧症				C
異常高血圧				C
夜間高血圧				C
肥満高血圧				C
外来高血圧				C
日本人高血圧				C

詳細 頻度レベル: 95-100% 信頼度: C

【高血圧症】

		作業済	確認済	信頼度
出現形	高血圧症	<input type="checkbox"/>	<input type="checkbox"/>	
出現形よみ	こうけつあつしょう	<input type="checkbox"/>	<input type="checkbox"/>	S
正規形	高血圧症	<input type="checkbox"/>	<input type="checkbox"/>	S

コメント

ここにメッセージ内容を入力