

MeSHタームとの合致度を調べた結果、対象語彙数は8954、その中で意味属性が関連付けの1位で一致したものは6805件(76%)、以下2位が654件、3位が249件で、5位以内で一致したものは7926件(89%)であった。試みに、意味的情報での一致度を、MeSHカテゴリー、セマンティックネットワーク、SNOMED-CT(MeSHとの単純一致)、SNOMED-CT(5階層以内にMeSHが一致する場合)において行ったこの結果を図3に示す。

6. 考察

本論文では、Medline文献のうち2000年から2004年までの5年分を対象としているが、試みに1995年以降の10年分を対象として文献の数をほぼ2倍に増やしても、検出できる語彙の数は、およそ8%しか増加していなかった。実は本研究ではTFIDFの計算を10年分の文献に対して行おうとしたのであるが、計算機リソースの関係で実行できなかった。しかし検出することのできた語彙数を見る限り、計算範囲を5年に絞っても辞書見出しのカバー率は5%程度しか低下しておらず、それほど大きな影響はないと言える。本研究における評価方法では、MeSHと一致した見出し語だけを対象としているため、それ以外の見出し語に対しては、文献に索引付けされたMeSHタームとの関連性が異なるのではないか、そのために成績も異なるのではないか、ということが懸念される。しかし重み行列の計算は、あくまで文献をもとにした見出し語とMeSHとの関連であり、見出し語がMeSHと一致しないものに対しても同様な成績であることが見込まれる。本研究での意味的関連は、実際には文献に付されたMeSHタームとの関連性に大きく依存する。MeSHタームからMeSHの意味カテゴリーやUMLSを介した他の統制用語集に付随する意味情報を取得して意味的関連としているからである。これらの意味情報は統制用語集に応じてほぼ一意的に定まるため、MeSH概念の一致度の成績と意味情報一致度の成績はほとんど変わらない。SNOMED-CT分類への一致度だけ成績が悪いのは、MeSHとSNOMED-CTとの対応が悪いことに起因している。Medline文献から意味的関連を取得するには、Medline文献を索引付けているMeSHの多くに付されているQualifierを利用することもできる。本研究で用いたのと同じ手法により、MeSHタームの代わりにQualifierとの関連性を調査すれば良い。あるいはMeSHタームとQualifierとの組み合わせを利用することも可能である。医学用語辞典の約3万語がMedline中に出現していないという結果について。原因として、最近は使われていない語彙である、複数形などの語尾変化により一致しない、辞書見出しにスペルミスが存在する、などが考えられる。それぞれ、より古い文献も検索範囲に含める、stemming処理を行う、辞書のスペルミスが存在するのならばそれらを修正する、などの対策が考えられる。最後に本研究での意味的関連の意味について。この関連はMedlineの文献における関連性である。NormalizeによりUMLSと一致した語彙を介して取得した意味属性の場合には、統制用語の分類を反映するのみであるため、例えば「胃癌」は「悪性腫瘍」となるが、文献的な関連性では「胃癌」に「化学療法」、「胃透視」、「内視鏡下手術」などが対応する点に注意を要する。また、文献における関連性であるため、これを診療録や読影報告書などの文書に適用することが可能であるか否かについては、別に考慮する必要がある。しかし医学領域での名詞、特に専門用語に関しては、それほど大きなずれは生じないと予想できる。

7. 結論

日本語医学用語に関連する意味的情報を取得するために、Medline文献5年分における英語見出しと文献索引との関連性を計算した結果、医学用語辞典の62%をカバーする49384語彙に対して、1位で76%、5位以内での精度89%で意味的関連を付すことができた。

参考文献

- [1] Yuzo Onogi et al.: Mapping Japanese Medical Terms to UMLS Metathesaurus. Medinfo 11, pp.406-410, 2004.
- [2] 大江和彦: 電子的診療情報交換のための実用的な病名概念マスターの在り方. 医療情報学 20(suppl.2) 37-38, 2000
- [3] 竹村匡正, 松井弘子, 声田信之: 用例に基づく医療用語知識の体系化について. 医療情報学 24(1), pp.139-145, 2004
- [4] 小野大樹, 高林克己, 鈴木隆弘, 横井英人, 井宮淳, 里村洋一: テキストマイニングによる退院サマリー自動分類の試み. 医療情報学 24(1), pp.35-44, 2004
- [5] 竹村匡正, 佐藤純三, 黒田知宏, 長瀬啓介, 高田彰, 田中亨治, 郭錦秋, 吉原博幸: MML化された退院時サマリーを知識ソースとした類似症例検索システムの開発. 第24回医療情報学連合大会論文集, pp.464-465, 2004
- [6] Srinivasan P, Hristovski D: Distilling conceptual connections from MeSH co-occurrences. Medinfo. 11(Pt 2):808-12, 2004
- [7] Zhu AL, Li J, Leong TY: Automated knowledge extraction for decision model construction: a data mining approach. AMIA ... Annual Symposium Proceedings/AMIA Symposium. :758-62, 2003
- [8] Burgun A, Bodenreider O: Methods for exploring the semantics of the relationships between co-occurring UMLS concepts. Medinfo. 10(Pt 1):171-5, 2001
- [9] Bodenreider O, Nelson SJ, Hole WT, Chang HF: Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. Proceedings / AMIA ... Annual Symposium. :815-9, 1998.

$$w_{i,j} = tf_{i,j} \times idf_i = \frac{freq_{i,j}}{\max_l freq_{l,j}} \times \log \frac{N}{n_i} \quad \text{TF}$$

-IDFによる計算

辞書見出し	MeSH JPN	MeSH分類(1)	MeSH分類(2)
ヒカントン	ヒカントン	Chemicals and Drugs	Heterocyclic Compounds
バルトネラ科	バルトネラセエ	Organisms	Bacteria
新生児強皮症	新生児強皮症	Diseases	Congenital, Hereditary, and Neonatal Di
こう(鏝)疫	こう疫	Diseases	Parasitic Diseases
だに中毒	だに中毒	Diseases	Disorders of Environmental Origin
モニチア症	モニチア症	Diseases	Disorders of Environmental Origin
スルフヘモグロビン血症	スルフヘモグロビン血症	Diseases	Hemic and Lymphatic Diseases
動物流行性運動失調症	ヒツジよろめき病	Diseases	Nutritional and Metabolic Diseases
過受胎	過受精	Biological Sciences	Reproductive and Urinary Physiology
強皮症	新生児強皮症	Diseases	Congenital, Hereditary, and Neonatal Di
キサントプテリン	ウロプテリン	Chemicals and Drugs	Heterocyclic Compounds
凍傷	浸水足	Diseases	Skin and Connective Tissue Diseases
悪夢	依存性人格障害	Psychiatry and Psychology	Mental Disorders
胸膜痛	Bornholm病	Diseases	Virus Diseases
ふくじゅそう(福寿草)〔属〕	ホストロファンチン	Chemicals and Drugs	Polycyclic Compounds
男女性胚(細胞)腫	性腺組織腫瘍	Diseases	Neoplasms
黄癰(せん)	黄せん	Diseases	Bacterial Infections and Mycoses
肋骨軟骨炎	Tietze症候群	Diseases	Musculoskeletal Diseases
舌痛(症)	灼熱舌	Diseases	Stomatognathic Diseases
梅毒性下疳	下疳	Diseases	Bacterial Infections and Mycoses
バイオクリーンルーム	ライフアイランド	Analytical, Diagnostic and Ther	Equipment and Supplies
X染色体	パール小体	Anatomy	Cells
ペータロピオラクトン	オキセタン	Chemicals and Drugs	Organic Chemicals
高アミノ酸尿〔症〕	アミノ酸尿症-腎性	Diseases	Urologic and Male Genital Diseases
腎溶質負荷	クロライド	Chemicals and Drugs	Heterocyclic Compounds
レトルタモナス〔属〕	肉質鞭毛虫類	Organisms	Animals
歯科治療学	局方注解	Information Science	Information Science
レットレル・ジーベ病	Letterer-Siwe病	Diseases	Neoplasms
巨大腎	Diectophyma上科	Organisms	Animals
ヘモバルトネラ症	アルスフェナミン	Chemicals and Drugs	Organic Chemicals
肛門周囲の	肛門周囲腺	Anatomy	Animal Structures
あぶみ骨可動化術	アブミ骨手術	Analytical, Diagnostic and Ther	Surgical Procedures, Operative
発泡剤	頸管びらん	Diseases	Female Genital Diseases and Pregnancy
ドレスラー症候群	交連切開後症候群	Diseases	Cardiovascular Diseases
腓腹筋	動脈内膜炎	Diseases	Cardiovascular Diseases
疑双口吸虫〔属〕	双口吸虫	Organisms	Animals
自然排液	気管支ドレナージ	Analytical, Diagnostic and Ther	Therapeutics
リケッチャ科	バルトネラセエ	Organisms	Bacteria
ルードウィッヒアンギナ	Ludwigアンギナ	Diseases	Bacterial Infections and Mycoses
麻痺性痴呆	オリーブ脊髄路	Anatomy	Nervous System
梅毒性髄膜炎	運動失調症-歩行性	Diseases	Bacterial Infections and Mycoses

見出し語と関連するMeSHカテゴリー(部分)

	MeSH	MeSH分類	SMN	SNOMED	SNOMED-5
1位で一致	76%	76%	76%	55%	70%
5位以内で一致	89%	89%	89%	63%	81%

ターム、および意味的関連付けの一致度

MeSH

セマンティックWebエンジンを用いたMedDRA/Jのオントロジ化とその 応用方法の研究

○鈴木 博道¹⁾ 清水 昇²⁾ 中須 弘平³⁾

(財)国際医学情報センター¹⁾

慶應義塾大学SFC研究所²⁾

(株)サイバーエッジ³⁾

MedDRA ontology using the "Semantic Web Engine", and its application study

○SUZUKI HIROMICHI¹⁾ SHIMIZU NOBORU²⁾ NAKASU KOHHEI³⁾

International Medical Information Center, Tokyo, Japan¹⁾

Keio Research Institute at SFC, Fujisawa, Japan²⁾

CyberEdge CO., LTD., Tokyo, Japan³⁾

Abstract: We have successfully ported MedDRA thesaurus into OWL ontology, and generated a new MedDRA ontology. The MedDRA ontology has almost same size of Galen ontology in UK.

Keywords: Semantic Web , RDF , OWL , MedDRA , Thesaurus , Medical Ontology ,
SemanticWeb Engine

1. 目的

セマンティックWebは、国際的なWeb技術の標準化団体であるW3C(World Wide Web Consortium)が中心となって標準化を進めている次世代Web技術である。セマンティックWebは、Web全体を一つのデータベース化する事とWeb上の情報を自動処理する事を狙いとしている。この狙いを実現する為、マシンリーダブルなメタデータ記述言語であるRDF(Resource Description Framework)とRDFを拡張した用語間の関係(W3Cでは用語間の関係をオントロジと呼んでいる)を記述する為の言語であるOWL(Web Ontology Language)とを定義している。セマンティックWebのオントロジ技術を用いる事により、概念の階層関係だけでなく、従来の関係データベースでは記述することが難しかった概念間のネットワーク関係や複合概念や条件付概念など様々な関係を知識として記述し維持管理する事が可能である。一方、「ICH国際医薬用語集日本語版」(以下MedDRA/Jと略す)は、5階層の階層構造を持つ用語シソーラスである。

本研究では、(株)サイバーエッジが販売している日本語処理を可能なセマンティックWeb基盤ソフトウェアである「Semantic Webエンジン」を用いて、MedDRA/JをセマンティックWebのオントロジにポーティングし(セマンティックWebのオントロジにポーティングされたMedDRA/Jを以下MedDRAオントロジと述べる)、MedDRAオントロジを生成する事を目的とする研究を実施した。

MedDRA/Jをオントロジ化する事により、MedDRA/Jの意味補強や他のオントロジや辞書やシソーラスとの連携が容易になるので、用語の同義語の追加や、各製薬会社などで利用している用語辞書や他の辞書シソーラスとMedDRA/Jシソーラスとの用語の違いを検出する機能やユーザ特有の変換ルールを補助オントロジとして追加する事でユーザ固有の用語をMedDRA/Jで規定された用語に自動変換する機能の実現が容易になる。また、MedDRA/Jをオントロジ化した事により、オントロジビューワ機能によりMedDRA/Jの意味を分かり易く表示したり、必要な情報を意味検索したり、GUIエディタ機能による変更修正したりする事が可能となった。

2. 方法

本研究では、医薬規制用語集バージョン5.1(MedDRA/J V.5.1J)を用い、次の変換規則を設定し、それに基づいてオントロジ化を行った。

①器官別器官別大分類(SOC)、高位グループ用語(HLGT)、高位用語(HLT)、基本語(PT)、下層語(LLT)及び特別検索カテゴリーの用語は総てクラスとして定義する。

②器官別大分類、高位グループ用語、高位用語、基本語及び下層語との間の階層関係をサブクラス関係として定義する。

③特別検索カテゴリーとそれに対応する基本語との間の関係を特別検索カテゴリーに対するサブクラスとして基本語を定義する。

④階層関係を示さない次の10個のフィールド属性をプロパティとして定義する。

・器官別大分類略称、・特別検索カテゴリー略称、・医薬品副作用用語集[WHOART]コード、・HARTSコード、・有害事象の用語[COSTART]、・ICD-9コード、・ICD9CMコード、・国際疾病分類[ICD10]コード、・カレント、・医薬品副作用用語集[JART]コード

⑤上記の④のフィールド属性値は、④で定義されるプロパティのプロパティ値として定義する。

⑥日本と英語とが定義されている用語の場合、分かり易くする為、日本語と対にして[]の中に英語を記述する。

⑦日本語の定義されていない用語、すなわち、英語のみの用語の場合、英語の用語だけを[]で括らずに記述する。

⑧日本語が未定義の用語は、未翻訳クラスを新設し、そのクラスのサブクラスとした。

⑨カタカナの振り仮名はオントロジビューワなどで表示した時、冗長になるので取込まない。

⑩下層語(PT)には日本語は同じだが英語の記述が異なるものがあるが、微妙に意味が異なると思われるので、別のものと見做して取込む。

生成されたMedDRAオントロジの「器官別大分類の眼障害」の部分を示すと次の様になる。

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [
<!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
<!ENTITY "http://example.org/meddra#" >]
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
```

```

xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
xmlns="http://example.org/meddra#"
xmlns:meddra="http://example.org/meddra#">
<owl:Class rdf:about="眼障害[Eye disorders]"/>
<owl:Class rdf:about="視覚障害[Vision disorders]">
<rdfs:subClassOf>
<owl:Class rdf:about="眼障害[Eye disorders]"/>
</rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="視覚路障害[Visual pathway disorders]">
<rdfs:subClassOf>
<owl:Class rdf:about="視覚障害[Vision disorders]"/>
</rdfs:subClassOf>
</owl:Class>
<owl:Class rdf:about="視覚路障害NOS[Visual pathway disorder NOS]">
<rdfs:subClassOf>
<owl:Class rdf:about="視覚路障害[Visual pathway disorders]"/>
</rdfs:subClassOf>
</owl:Class>
...

```

このオントロジ化されたMedDRA情報は、次の様にセマンティックWebエンジンのオントロジビューワで表示する事が可能である。

図1: オントロジビューワ

3. 結果

元データである医薬規制用語集バージョン5.1(MedDRA/J V.5.1J)とそれから生成されたMedDRAオントロジを比較して見ると、MedDRA/J V.5.1Jは用語と用語間をリンクするリンクポイントとデータとから構成されており、その用語の総数は75,137個、リンクポイントの総数は24,765個であり、総てを合計すると99,902個のデータから構成されている。一方MedDRAオントロジのステートメント数は、クラス定義とクラス関係定義とを行うステートメント数が109,180、フィールド属性をプロパティ属性に置き換えた定義ステートメント数が97,900となった。この結果は、MedDRA/J V.5.1Jの用語とリンクポイントとが、MedDRAオントロジのクラス定義及びクラス関係定義ステートメントにマッピングされ、MedDRA/J V.5.1Jのフィールド属性が、MedDRAオントロジのプロパティ定義ステートメントにマッピングされた事により、この分だけデータ量が増加している。

フィールド属性はMedDRA固有のテーブル構造を認識している事を前提としたデータ定義であるのに対して、MedDRAオントロジのプロパティ定義は、<主語><述語><目的語>の3つ要素から構成されるトリプルモデルに則した記述であり、特定のデータ構造を前提知識として必要としない汎用的なマシンリーダブルのデータである。

MedDRAオントロジは、オリジナルのデータと比較してデータ量が増加する問題はあるが、ソフトウェアで意味を解釈し易いマシンリーダブルのデータなので、知的検索、知識追加、意味変換、意味補強、推論及び他のオントロジとの連動が可能になるので有用な増加である。

4. 結論

本研究により関係データベースに格納する事を前提としたMedDRAシソーラスをセマンティックWebのOWL言語を用いて記述したMedDRAオントロジに変換可能である事が実証できた。

MedDRAオントロジは、MedDRAシソーラスに比べてデータ量は増加するが、特定のデータ構造を前提知識として必要としないデータなので、様々な既存のセマンティックWebツールやこれから開発されるであろうセマンティックWeb関係のツールを用いて容易に処理する事ができる。

セマンティックWebのツールを用いる事で、MedDRAの情報を分かりやすく表示したり、適切に管理したりする事が可能となる。例えば、MedDRAオントロジをセマンティックWebエンジンのオントロジビューワで表示させてみて気づいた事であるが、MedDRA/J V.5.1Jの日本語の用語には次の様な問題がある。

- ①下層語(LLT)に日本語の用語の定義されていないものが多い。
- ②日本語の用語は、同一だが英語の用語が異なるものがある。すなわち、複数の英語の用語が、ただ一つの日本語の用語に対応付けられているものが幾つか存在する。

5. 考察

本研究により開発されたツールを用いて生成されたMedDRAオントロジは、次の様な特徴がある。

①MedDRAオントロジの総トリプル数は約20万トリプルであり、海外の大規模医療オントロジと比較すると、英国のGalenオントロジが20万トリプル、米国のNCI(National Cancer Institute)オントロジが50万トリプルあるといわれているので、MedDRAオントロジは大規模オントロジの一つであると言える。

②MedDRAオントロジは階層概念定義とプロパティ属性定義しか持たず、条件付概念定義や複合概念定義を有するGalenオントロジやNCIオントロジに比べて単純なオントロジである。

MedDRAオントロジは、次の様な機能や新たなサービスを実現するのに利用する事ができる。

- ①知的検索、例えば「ICD9CMコードとして”371.23”を持つものの上位用語は何か?」と言った検索が可能となる。
- ②知識追加、オントロジはテーブル形式ではないので、記述用の語彙を追加すれば新たな知識(情報)を自由に追加できる。
- ③意味変換、或るクラス(用語)と別のあるクラス(用語)とが同一のものを意味するとか、もしくは、異なるものであるとか言う事を定義する意味変換用のオントロジを追加すれば、MedDRAとベンターの持つ他の用語集との変換が容易になる。

④推論、例えば、(i)アンピリシンの禁忌は伝染性単核球症であり、(ii)ABPCがアンピリシンの略称であり、(iii)SBT/ABPCがSBTとABPCとの合剤であることが判明している場合、SBT/ABPCも伝染性単核球症に対して禁忌である事が推論できる。

⑤他のオントロジとの連動、MedDRAオントロジと類似した医療オントロジに英国のGalenオントロジや米国で開発されたNCIオントロジがある、これ等のオントロジと連動する為のゲートウェイオントロジを追加すれば、これ等を統合したより広範囲をカバー可能で有用性の高い医療知識データベースを実現できる。

The screenshot shows a software interface for an ontology browser. On the left side, there is a hierarchical tree structure representing the ontology. The right side displays a detailed list of terms, likely from the MedDRA ontology, with columns for term names and other attributes. The interface includes standard window controls and a search bar at the top.

オントロジビューワ