

解剖学オントロジーを用いた情報抽出のための 機能要件の検討

Application Functions of Information Extraction with the Use of the Anatomical Ontology

湯本正典¹⁾ 今井健²⁾ 桜井亮太¹⁾ 林亜紀¹⁾ 佐藤恵¹⁾ 梶野正幸¹⁾ 河添悦昌²⁾ 大江和彦¹⁾²⁾

Masanori Yumoto, Takeshi Imai, Ryota Sakurai, Aki Hayashi, Megumi Sato,
Masayuki Kajino, Yoshimasa Kawazoe and Kazuhiko Ohe

- 1) 東京大学医学部附属病院 企画情報運営部
Department of Planning, Information and Management
The University of Tokyo Hospital
2) 東京大学大学院医学系研究科
Graduate School of Medicine, The University of Tokyo

Abstract: We are constructing a systematic knowledge foundation in medical electric information systems. To make the foundation act effectively, we are building a medical ontology on the strict rules, which we introduced, based on the study of ontology engineering. Computers will come to treat electric documents more correctly with the reference of it. We discuss necessary application functions to extract instructive medical information in the use of the anatomical region of our ontology. Assuming some questions related with anatomic, we put basic information, which is needed to answer them, in order. As the result, a set of fundamental anatomical concepts contains two typical categories: entity and relationship.

1. はじめに

今日、情報技術の進歩に伴い、各種業務におけるデータの記録方法や情報の伝達手段が高度に電子化されている。医療の分野でも、電子カルテシステム(以下、電子カルテと呼ぶ。)やレセプトシステムなどの導入が盛んになっている。電子カルテには、日々、診療の記録が大量に蓄積され、患者さんの再診や入院、手術や検査の際に、個人の記録として有効に活用されている。その一方で、これらのデータを医療関係者の間で共有し、新しい医学の知見を得る臨床研究を行うためのシステムの整備が期待されている。たとえば、多数の患者診療情報を集積し、数値やテキストの統計処理を行うことにより、類似症例に関する情報の分析、医学的な相関ルールの抽出などが行える。このような研究が進展することにより、さらなる医療の質の向上が期待できるが、そこに至るまでにはいくつかの課題がある。まず、医療情報を共有するためには、医療データに含まれる多彩な専門用語の標準化およびコード化が必要である。しかし、それだけでは十分ではない。用語の概念レベルの定義、そして、各概念間の意味的な関係性を含む、医療に関

する知識や技術の論理的な体系化が必要である。このような、体系化した知識が実装されている医療知識基盤を用いることにより、はじめて、情報の共有や知識の創発が効率的に、しかもより正確に行えるようになる。

このような背景のもと、筆者らはこれまで大阪大学溝口研究室との共同研究である厚生労働省「医療情報システムのための医療知識基盤データベース研究開発事業」にて、臨床医学オントロジーの構築を進めてきた。ここでいう臨床医学オントロジーとは、「臨床医学概念の定義を概念間の意味関係を用いて形式的に定義した体系」のことである。これまで、人体解剖構造領域(以下、この領域を解剖学オントロジーと呼ぶ)と、疾患領域を中心に記述を進め内容が充実してきている。これと並行して重要なのは、実際のアプリケーションにおける利用方法を考え、そのアプリケーションの機能要件を検討することである。そのための作業手順を以下のように考えた。まず、1) 解剖学オントロジーを利用して答える質問を列挙する。そして、2) その質問に答えるために必要な解剖学の情報を整理する。さらに、3) その情報が、現在の解剖学オントロジーに記述されているかを検証する。4) 記述されている場合は、その情報を抽出するための推論の想定し、整理する。5) 記述されていない

問1:与えられた解剖学的なエンティティのタイプは何か?

問2:与えられた解剖学的なエンティティの形状は何か?

問3:与えられた解剖学的なエンティティの周辺にあるエンティティは何か?

Case2):解剖学的なエンティティ同士の間の関係に関する質問

問4:解剖学的なエンティティAとBは接続しているか、もししていれば何を伝達する機能を持つか?

問5:ある特定の関節が、どのような動き方をするか?

問6:解剖学的なエンティティAとBは、空間的にどのような包含関係にあるか?

問7:解剖学的なエンティティAとBは、空間的にどのような位置関係にあるか?

Case3):より臨床的な質問、臨床医学における教育的な問い

問8:胃がんの時に摘出すべきリンパ節とそれが含まれる領域、すなわち切除領域はどこか?

問9:胃潰瘍が進行した時の深度によって、どのような治療が必要か?

問10:ある特定の臓器(ex. 肝臓)は再生するか否か?

問11:ある特定部位の神経が障害された場合、どこの領域にしびれや麻痺等の症状が発現するか?

問12:ある特定部位の血管が狭窄したら、どの領域に組織壊死などの影響が出るか?

問13:ある特定領域にしびれや麻痺が発現している場合、どの部位の神経が障害されている可能性があるか?(問11の逆問題。)

問14:ある特定領域が組織壊死している場合、どの部位の血管が狭窄している可能性があるか?(問12の逆問題。)

これらの知見を基にさらに詳細に検討した結果、質問に答えるためには、オントロジーに記述されるべきレベルの記述粒度において、少なくとも次の章に書かれているような情報が必要であることが明らかになった。

4. 知識提供のための情報の整理

前章のような質問に答えるために必要な情報を整理した。その結果を示す。以下順に、1) 特定の解剖学エンティティに関する情報(これを、解剖学エンティティ情報と呼ぶことにする。)、2) 複数の解剖学エンティティ間の関係に関する情報(これを解剖学リレーションシップ情報と呼ぶことにする)、3) 解剖学以外の情報に分けて述べる。

1) 特定の解剖学エンティティに関する情報(解剖学エンティティ情報)

(ア) 解剖学エンティティは、解剖学に含まれている構成部品と構成領域の2種類に分けられる。

(イ) 解剖学エンティティは、エンティティ名と、属性からなる。オブジェクト指向設計では、クラスに相当する。

(ウ) エンティティ名は、構成部品名と構成領域名のいずれかである。

① 構成部品名は、解剖学における構成部品の名前である。

➤ たとえば、尺骨、橈骨などである。

② 構成領域名は、種類の違う部品がいくつか集まって意味を成す領域の名前である。

➤ たとえば、上肢、上腕、前腕などである。

(エ) 構成部品の属性とは、以下の項目である。

① 物理的なスケールを表す属性

➤ 計量によりわかる属性である。

➤ ただし、各個人のインスタンスの計量でなく、解剖学における代表的な計量をとるとする。

➤ たとえば、代表的な、大きさ、長さ、重さ、深さなどである。

② 形状や色に関する属性

➤ 大域的な外観の観察でわかる属性である。

➤ ただし、各個人のインスタンスの観測で得られるのではなく、解剖学における代表的な観察の結果とする。

➤ たとえば、典型的な形や色などである。

③ 動きや機能を表す属性

➤ 時間の流れや、外界との関連で決まる属性である。

➤ たとえば、動きや機能、再生の可能性の有無などである。

④ 1つ上の概念の名前

➤ 一つリンクをたどった上位概念の名前を記述する。すなわち、is-a 関係で結ばれている上位の概念を記述する。

⑤ 1つ下の概念の名前

➤ 一つリンクをたどった下位概念の名前を列挙する。すなわち、is-a 関係で結ばれている下位の概念の名前を列挙する。

⑥ その概念の全体物の名前

➤ 一つリンクをたどった上の全体物の名前を記述する。すなわち、part-of 関係で結ばれている全体物の名前を記述する。

⑦ その概念の構成部品の名前

- 一つリンクをたどった下の構成部品の名前を列挙する。すなわち、**part-of** 関係で結ばれている構成部品の名前を列挙する。

⑧ 位置的に周囲にある構成部品の名前

- 位置的に隣接していると判断できる構成部品の名前を列挙する。

⑨ 所属する構成領域の名前

- その構成部品が属している構成領域の名前を列挙する。

⑩ 学問的概念として関連が近い構成領域の名前

- ある分野を指定することによって、関連してくる構成領域を記述する。
- たとえば、神経内科において、損傷した神経の名前と関連のある、神経の支配領域を記述する。

(オ) 構成領域の属性

① 位置的に周囲にある構成領域の名前

- 位置的に隣接していると判断できる構成領域の名前を列挙する。

② 学問的概念として関連が近い構成領域の名前

- ある分野を指定することによって、関連してくる領域を記述する。

③ 所属している構成部品の名前

- その構成領域に所属している構成部品の名前を列挙する。

ここまでの情報により、推論を適切に行えば、少なくとも、**Case1** に分類された質問に答えることができる。さらに、**Case3** に分類された問題10にも答えられる。

2) 複数の解剖学エンティティ間の関係に関する情報(解剖学リレーションシップ情報)

(ア) 解剖学リレーションシップとは、解剖学エンティティ同士の関係を記述する概念である。

(イ) 解剖学リレーションシップは、リレーションシップ名と、属性からなる。

(ウ) リレーションシップ名には以下の項目が考えられる。

① 接続名

- 接続ポートの名前を記述する。

② 空間領域の包含の概念的関係の名前

- 空間領域AとBが、含んでいるか、含まれているかの関係の名前を記述する。

- たとえば、AはすべてBに含まれている、Aは一部Bに含まれているなどである。

③ 相対的位置関係やトポロジーの名前

- 互いの相対的な位置概念の計量や、互いのトポロジ的な位置概念の名前を記述する。

- たとえば、隣接している、離れている、などの名前である。

④ 概念関係の名前

- このオントロジーに記載されている概念間の関係の名前を記述する。

- たとえば、**is-a** 関係や **part-of** 関係、ロールホルダーとクラス制約の関係など。

(エ) 解剖学リレーションシップの属性には、以下のような項目が考えられる。

① 接続の属性、例は以下のとおりである。

- 接続しているか、損傷しているか。
- 接続の種類（脈管接続など）

② 空間領域の包含の概念的関係の属性

③ 相対的位置関係やトポロジーの属性、例は以下のとおりである。

- 離れている距離、たとえば、解剖学的な典型では10センチメートル離れているなど。
- 配置のトポロジー形状、たとえばらせん状など。

④ 概念関係の属性、例は以下のとおりである。

- 通常の **is-a** 関係の属性
- 包含関係の属性：**part-of** 関係、**part-of-whole** 関係、**part-of-region** 関係、**part-of-gc** 関係
- ロールホルダーとクラス制約の関係といった属性

ここまでの情報により、適切な推論を行えば、少なくとも、**Case2** に分類された質問には答えることができる。また、**Case3** に分類された、問11から14に答えられる。

最後に残った、問8と9であるが、これに答えるためには、

3) 解剖学以外の情報

(ア) 疾患に関する情報

が必要である。

さて、解剖学の情報は、このように分類されるが、これらの情報が、現在の解剖学オントロジーに記載されているかを次の章で検証する。

5. 考察

まず、第4章に書いた解剖学の情報が、現在のオントロジーにすべて記述されているか否かを検証した。その結果、いくつかの情報の記述がまだ不足していることがわかった。その不足の原因は2つある。一つは、記述のルールは決まっているが、まだ時間が足りないことにより記述されていないという「量的な不足」である。そして、もう一つは、まだ記述のルールが定まっていない「質的な不足」である。量的な不足となっている情報には、「再生の可否」等、また、質的な不足となっている主な情報には、支配領域のほか、構成領域の記述方法(前腕、上腕など)、空間領域同士の包含関係の記述方法、相対的な位置関係やトポロジーの記述方法、が含まれている。今後はこのような情報の記述を行い、本オントロジーをより利活用しやすいものとする。

次に、推論としては、1) 特定の解剖学エンティティに関する情報を導く推論、2) 複数の解剖学エンティティ間の関係に関する情報を導く推論、3) 上記の2つを合わせて答えを導く推論、4) 上記の3つに加え、疾患に関する情報を合わせて答えを導くための推論が必要であると判明した。今後は、推論の設計と実装を進めていく。

第3に、このような、解剖学の知識を処理する機能の実装は、古くから人工知能の医療分野への適用という観点により、興味を持たれていたテーマである。たとえば、過去には Prolog による実装が試みられている[8,9,10,11]。しかし、このような研究のうち、発表されているものは数少ない。これは、計算機に知識処理を実装するときに、知識を「データベースに保持するか」、もしくは、「ロジックのコードの中に埋め込むか」の切り分けが明確に行えず、機能の拡張や維持保守ができなかったことが一つの原因であったといわれている。本研究のオントロジーは、記述する内容と方法を厳密に定義しており、それにより、「データベースに保持する知識」がはっきりしている。したがって、上記の切り分けが明確にできるようになったと考えている。本研究は、古くからの知識処理のアルゴリズムを実装するための、異なった視点からの研究であるともいえる。

以下は、今後の課題について述べる。まず、この解剖学知識の提供サービスの利活用のシーンを考えて、ユースケースをさらに詳細に洗い出す。

現在、第3章の間8や9のような、解剖学の教育的な知識を考えているが、それでも、疾患の知識と合わせることが必要であることはすでに記した。したがって、最低限、オントロジーの疾患領域の情報と、合わせて知識を抽出する推論を考案することが推論設計の最大の課題である。

最後に、このような機能要件から、アプリケーションの提供する機能を設計して仕様を決めることが次の課題でもある。そして、その、API仕様を実装してオープン化し、みなさまに利用していただくことが、本WGに課せられてミッションである。現在のところ、法造のAPIである「法造コア」を情報抽出のAPIとして最大限活用する方針でいる。

6. 謝辞

本研究の一部は、日本学術振興会の最先端研究開発支援プログラムおよび厚生労働省医療知識基盤研究開発事業により、助成を受けたものである。オントロジー構築において共同研究を行っている、溝口理一郎先生、古崎晃司先生はじめ、大阪大学産業科学研究所溝口研究室のみなさまに心より感謝する。

参考文献

- [1] 大江和彦, 今井 健. 臨床医学知識処理を目指した医療オントロジー開発. 人工知能学会誌 Vol.25, No.4, pp.493-500, (2010)
- [2] 古崎晃司, 国府裕子, 周俊, 今井 健, 大江和彦, 溝口理一郎. 臨床医学オントロジーの構築とその基本思想. 第19回セマンティックウェブとオントロジー研究会 SIG-SWO-A802-09, (2008)
- [3] SNOMED-CT. <http://www.ihtsdo.org/snomed-ct/>.
- [4] College of American Pathologists. SNOMED Clinical Terms Guid. Version 5.0, 2002-2006.
- [5] Schulz, S., Suntisrivaraporn, B. and Baader, F.: SNOMED-CT's problem list: ontologists' and logicians' therapy suggestions. Stud. Health Technol. Inform., Vol.129, Pt1, pp.802-806, (2007)
- [6] OpenGALEN. <http://www.opengalen.org/>.
- [7] FMA:The Foundational Model of Anatomy Ontology. <http://sig.biostr.washington.edu/projects/fm/>.
- [8] Ohe, K., Kaihara, S. : Intelligent Atlas: A method to Support Physicians' Spatial Reasoning. B.Barber et al.(eds), Elsevier Science Publishers B.V.(North-Holland), MEDINFO 1989, 175-179, 1989.
- [9] Ohe,K., Kaihara,S.: Representation of Neuroanatomical Knowledge by PROLOG. Expert Systems and Decision Support in Medicine, "Reichertz,P.L. and Lindberg,D.A.B (eds), 249-255, 1988.
- [10] 大江和彦, 開原成允:診断支援のための電子解剖学アトラス. 第9回医療情報学連合大会論文集: 799-802, 1990
- [11] 大江和彦: 脳神経の走行経路を PROLOG で記述した知的な解剖アトラス. PROLOG 産業応用シンポジウム論文集:25-30, 1991.