

Title	医療知識基盤の構築に向けた疾患オントロジーの Linked Open Data化
Author(s)	古崎, 晃司; 山縣, 友紀; 国府, 裕子; 今井, 健; 大 江, 和彦; 溝口, 理一郎
Citation	人工知能学会論文誌, 29(4): 396-405
Issue Date	2014
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/12918
Rights	Copyright (C) 2014 人工知能学会. 古崎晃司, 山縣友 紀, 国府裕子, 今井健, 大江和彦, 溝口理一郎, 人工 知能学会論文誌, 29(4), 2014, 396-405. http://dx.doi.org/10.1527/tjsai.29.396
Description	

特集論文 「Linked Open Data とセマンティック Web」

医療知識基盤の構築に向けた疾患オントロジーの Linked Open Data 化

Publishing Linked Open Data from a Disease Ontology toward a Knowledge Infrastructure

古崎 晃司

Kouji Kozaki

大阪大学産業科学研究所

I.S.I.R., Osaka University

kozaki@ei.sanken.osaka-u.ac.jp, <http://www.ei.sanken.osaka-u.ac.jp/~kozaki/>

山縣 友紀

Yuki Yamagata

(同上)

yamagata@ei.sanken.osaka-u.ac.jp

国府 裕子

Hiroko Kou

情報通信研究機構

National Institute of Information and Communications Technology

h.kou@nict.go.jp

今井 健

Takeshi Imai

東京大学病院 企画情報運営部

Dept. of Planning, Information and Management, The University of Tokyo Hospital

ken@hcc.h.u-tokyo.ac.jp

大江 和彦

Kazuhiko Ohe

東京大学大学院医学系研究科

Department of Medical Informatics, Graduate School of Medicine, The University of Tokyo

kohe@hcc.h.u-tokyo.ac.jp

溝口 理一郎

Riichiro Mizoguchi

北陸先端科学技術大学院大学

Japan Advanced Institute of Science and Technology

mizo@jaist.ac.jp

keywords: ontology, disease ontology, causal chainviewpoint, linked data

Summary

Publishing open data as linked data is a significant trend in not only the Semantic Web community but also other domains such as life science, government, media, geographic research and publication. One feature of linked data is the instance-centric approach, which assumes that considerable linked instances can result in valuable knowledge. In the context of linked data, ontologies offer a common vocabulary and schema for RDF graphs. However, from an ontological engineering viewpoint, some ontologies offer systematized knowledge, developed under close cooperation between domain experts and ontology engineers. Such ontologies could be a valuable knowledge base for advanced information systems. Although ontologies in RDF formats using OWL or RDF(S) can be published as linked data, it is not always convenient to use other applications because of the complicated graph structures. Consequently, this paper discusses RDF data models for publishing ontologies as linked data. As a case study, we focus on a disease ontology in which diseases are defined as causal chains.

1. はじめに

近年、医療分野における情報化は世界的に進められており、電子カルテをはじめとする病院システムの情報化やそれに伴い蓄積された多量の情報の有効活用などが盛んに行われている。そのような医療情報システムの高度化を支える知識基盤として、医療分野を対象としたオントロジー構築が求められている [大江 10]。このような背景の下、筆者らは厚生労働省の支援を受け、オントロジー工学の成果を駆使した適切な概念化に基づく洗練された

臨床医学オントロジーを目指し、オントロジー構築を進めてきた [古崎 08, Mizoguchi 09, 11]。その構築過程は、オントロジー工学者と臨床医を含む医療領域の専門家の綿密な協力の下で進められており、その成果となる臨床医学オントロジーは、深い専門性とオントロジー的な整合性を両立したものであると言える。構築された臨床医学オントロジーには、13 診療科の代表的な約 6,000 の疾患定義を対象とした疾患オントロジー、人体の全血管、神経、筋肉および骨の接続関係の記述を含む、主要な解剖学構造を定義した解剖学オントロジーなどが含まれる。

これらの成果を有効に活用し、様々な医療情報システムの開発を促進するには、臨床医学オントロジーをオープンな知識基盤として公開することが望まれる。そこで本研究では、近年、様々な領域におけるデータの公開手法として注目されている Linked Data 技術を利用し、臨床医学オントロジーを Linked Open Data (LOD) として公開することを検討する。Linked Data は、形式上は RDF で書かれたデータであるため、OWL や RDF(S) などの RDF 形式で表現したオントロジーを Linked Data として公開することは容易にできる。しかし、OWL などのオントロジー記述言語は適切な推論処理を行うことを前提としたクラスレベルの記述を中心に設計されているのに対し、Linked Data ではインスタンス間のリンクを中心とした処理を中心に考えられている。そのため、OWL 形式のオントロジーをそのまま Linked Data として公開しただけでは、必ずしも、利便性に富んだ LOD になるとは限らない。そこで本研究では、疾患オントロジーを例として、臨床医学オントロジーの有効活用を促進するにあたって適切な LOD としての公開方法を検討する。

以下、2 章では Linked Data におけるオントロジーの役割について述べ、続く 3 章では本研究で LOD 化の対象とする疾患オントロジーの概要を述べる。4 章では疾患オントロジーの LOD 化の方針について考察し、疾患を構成する因果連鎖を対象とした LOD 化を行う。5 章で関連研究について述べ、6 章で本研究を総括すると共に、臨床医学オントロジーを LOD 化するにあたっての今後の課題と展望を述べる。

2. Linked Data におけるオントロジーの役割

オントロジー (Ontology) は存在論を意味する哲学の用語であるが、情報学の分野においては「概念化の明示」[Gruber 92] や「人間が対象世界をどのように見ているかという根源的な問題意識を持って物事をその成り立ちから解き明かし、それをコンピュータと人間が理解を共有できるように書き記したもの」[溝口 05] という意味で用いられる。セマンティック Web においてメタデータを表現するスキーマや語彙を提供すると共に、データのセマンティクスを与える役割を担う。Linked Data においては相互に連結 (リンク) されたインスタンスを公開することがオントロジーを厳密に定義することよりも重要視される傾向がやや見られるが、オントロジーの果たす役割は変わっていない。他のデータと多くのリンクを持つ膨大なデータセットは、たとえオントロジーに基づいていなくとも十分な価値がある場合もあるが、それらのデータの効率的かつ一貫性の持った活用には共通語彙やスキーマを提供するオントロジーの役割が重要となる。

実際に、Linked Open Data として公開されているデータセットの多くはオントロジーに基づいている。例えば、

DBpedia^{*1} の英語版は 377 万の事物に関するデータが含まれているが、そのうち 235 万が 500 クラスと 1900 プロパティの定義を含むオントロジー^{*2} に基づき分類されている。これは Linked Data におけるオントロジーの典型的な利用法であり、多くのオントロジーが同様の目的で公開されている。バイオインフォマティクス分野のオントロジーを公開するための著名なポータルサイトである NCBO BioPortal[Whetzel 11] には 355 のオントロジーが公開されると共に相互のリンク付けが行われており、生命科学分野における Linked Data の一翼を担っている^{*3}。

一方、インスタンス (データ) 記述のためのスキーマや語彙を提供することよりも、オントロジー自体を意味処理のための知識基盤として用いることを中心的な役割として公開されているオントロジーがある。このようなオントロジーは、領域独立な汎用的知識の提供を目的とした Cyc^{*4} や WordNet^{*5} と特定領域における専門知識の提供を目的とした SNOMED-CT^{*6} や Galen^{*7} などに大別できる。これらのオントロジーは先に述べた Linked Data におけるスキーマや語彙の提供を目的としたものに比べ、多くのクラス定義を持つ傾向にある。例えば、日本語における汎用知識の提供を目的として開発されている日本語 Wikipedia オントロジー (JWO)^{*8} は 160,000 のクラス定義を含む [玉川 11, Tamagawa 12]。同じく Wikipedia を元に構築されている DBpedia 英語版では 500 クラスしか定義されていないことを考えると、この違いは、主として想定しているオントロジーの利用方法の違いがその一因であると思われる。すなわちオントロジーそのものを構造化された知識として用いる際には、より詳細なクラス定義が必要とされる。

本研究における疾患オントロジーの LOD 化は、疾患定義そのものを構造化された医療知識基盤として利用することを想定したものである。

3. 疾患オントロジーの概要

近年の医療情報システムには膨大なデータが蓄積されており、それらのデータを活用したシステムの高度化を目指した知識処理のための基盤知識として、様々なオントロジーの構築が進められている。疾患オントロジーとして代表的なものには、OGMS [Scheuermann 09]、DOID [Osborne 09]、IDO [Cowell 10] などがあるが、これらのオントロジーは属性 (プロパティ) を中心とした疾患定義を行っている。それに対して筆者らは、疾患を異常状態の因果連鎖から構成されるものとして捉える River Flow Model of Disease というモデル [Mizoguchi 11] を提案し

*1 <http://dbpedia.org/>

*2 <http://mappings.dbpedia.org/server/ontology/>

*3 <http://bioportal.bioontology.org/>. 2013 年 8 月 17 日時点

*4 <http://www.cyc.com/>

*5 <https://wordnet.princeton.edu/>

*6 <http://www.ihtsdo.org/snomed-ct/>

*7 <http://www.opengalen.org/>

*8 <http://www.wikipediaontology.org/query/>

ている。本研究で LOD 化を行う疾患オントロジーはこのモデルに基づいており、「ある疾患の原因となる異常状態は何か？」や「ある疾患が、どのような異常状態の遷移を引き起こし、どのような症状が現れるか？」といった質問に答えることができる因果連鎖に関する詳細な情報が含まれている。本章では、この疾患オントロジーにおける疾患定義とその計算機表現の概要について述べる。なお、本研究では疾患オントロジーの LOD 化を中心課題としており、River Flow Model や疾患定義のオントロジーとしての表現の詳細については別稿での報告を予定している。

3.1 疾患の概念定義

疾患を理解する上では、その疾患が何を起因としており、どのような結果（病態や症状）を引き起こすかを適切に捉えることが重要とされる。そこで本研究では、疾患概念を「その原因と途中経過を含めた一連の状態変化の連鎖と、それにより引き起こされている結果状態との総体」として捉える [大江 10, Mizoguchi 11]。すなわち、疾患は人体に現れる状態変化による因果連鎖から構成される従属的実在物として定義される。ここで因果連鎖は川の流れるように自らを変化させるものとして捉えることができ、オントロジー的には持続物の下位概念として定義できる [Mizoguchi 11, 古崎 11]。疾患定義に用いられる因果連鎖（以下、疾患連鎖と呼ぶ）は人体に現れる（異常）状態とその間の因果関係から成り、その種類は、汎用疾患連鎖 人体で起こりうると思われる全連鎖。すべての疾患から共通して参照される。
 疾患定義連鎖 疾患定義に必要な連鎖。その疾患を発症した患者には共通して見られる。
 派生連鎖 ある疾患定義連鎖から上流/下流方向に汎用疾患連鎖を辿って得られる連鎖。上流側はその疾患の原因と考えられうる連鎖、下流側はその疾患を発症した患者に起こりうる連鎖を意味する。
 に分かれる。

疾患定義の中心は「疾患定義連鎖」であり、疾患の is-a 関係は、疾患定義連鎖の範囲の包含関係で表される*9。例えば、糖尿病の疾患定義連鎖が「インスリン作用不足 高血糖」であるとすると*10、「細胞の破壊 血中インスリン量の絶対的低下 インスリン作用不足 高血糖」を疾患定義連鎖とする 1 型糖尿病は糖尿病の下位概念となる（図 1）。すなわち、疾患の下位概念の定義は、疾患定義連鎖の範囲を派生連鎖に沿って上流・下流に広げたものとなる。

一方、派生連鎖は、定義には含まれないがその疾患で典型的に見られると思われる異常状態の連鎖を表す「疾

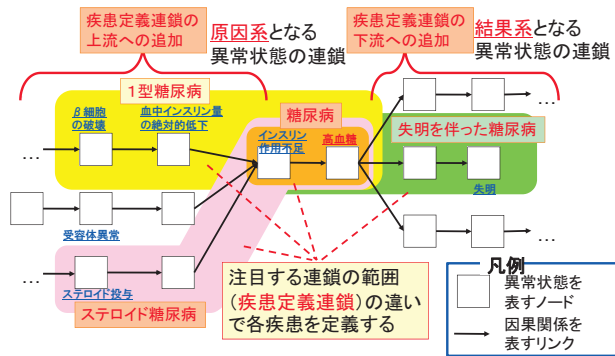


図 1 糖尿病における疾患連鎖の例

患派生連鎖」と、すべての汎用疾患連鎖を辿ること得られる「一般派生連鎖」に分かれる。本研究では、臨床医が自らの持つ専門知識と臨床の教科書の記述をもとに、各疾患の疾患定義連鎖および一般派生連鎖を定義することで、疾患オントロジーを構築した。

このように臨床医の手によって定義された各疾患を構成する因果連鎖（疾患定義連鎖と疾患派生連鎖）を一般化することで、通常は、定義することが現実的には不可能と思われる人体で一般的に起こりうると考えられている全ての因果連鎖（汎用疾患連鎖）を、ボトムアップ的に定義することができる。当然ながら、この方法で定義した汎用疾患連鎖は、現実世界で人体に起こりうる全ての因果連鎖を網羅したものではないが、本疾患オントロジーで定義した全疾患において臨床医が関心を持つ範囲で起こりうると思われる全連鎖に相当する。よって、この汎用疾患連鎖を用いることで、疾患定義の際に直接記述した因果連鎖のみならず、全診療科を横断した汎用疾患連鎖を用いた推論が可能となる。

なお、2013 年 3 月 11 日時点で本疾患オントロジーには、13 診療科に渡る 6,302 の疾患と、その疾患を構成する 21,669 の異常状態が定義されている*11。

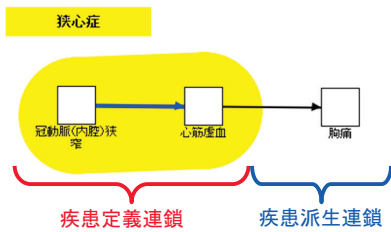
3.2 疾患オントロジーの計算機表現

図 2 に疾患オントロジーにおける疾患定義の「法造」による計算機表現の例を示す。疾患（「狭心症」）の右に示された「因果構造物」と書かれた長方形は、因果連鎖の一般構造を定義した概念への参照を示しており、疾患が因果連鎖から構成されることを表している。疾患連鎖に現れる異常状態は「構成要素」として列挙され、それらが疾患定義連鎖/疾患派生連鎖のいずれに属するかは「連鎖種別」属性で表される。異常状態間の因果関係は「原因/結果」スロットの相互参照で表現されており、必要に応じて発生確率属性が付与される。下位疾患の定義では、上位疾患から継承された構成要素の特殊化や、新たに疾患定義連鎖/疾患派生連鎖の範囲に入った異常状態の構成

*9 疾患定義連鎖の範囲は変わらないが、それを構成する異常状態がより詳細な状態に特殊化されることにより下位疾患が定義されることもある。
 *10 これは説明用に簡略化したものであり、本研究の疾患オントロジーにおける糖尿病の定義とは異なる。

*11 このうち 2,103 疾患（13,910 の異常状態）の定義が 2010-2012 年度のプロジェクトで大幅に見直されたものである。

狭心症の疾患連鎖



疾患オントロジーにおける狭心症の定義

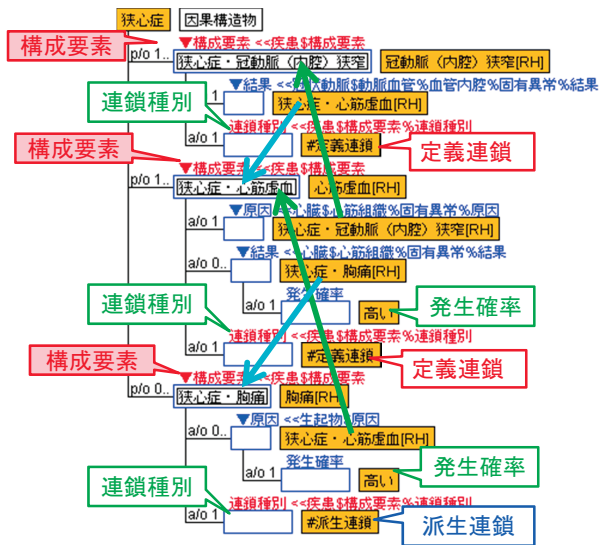


図2 疾患定義の計算機表現例

要素との追加が行われる。

なお、各構成要素は汎用連鎖として定義された異常状態を参照し、各異常状態の定義からその疾患定義に必要な定義のみを継承することで、疾患の種類に依らない汎用疾患連鎖との違いを明示している。すなわち、疾患定義における構成要素は、異常状態オントロジーの3階層モデルにおけるレベル3「特定のコンテキスト(疾患)を構成する異常状態」に相当し、レベル2「対象物内で汎用的に現れうる異常状態」を参照することで定義されている [山縣 12, Yamagata 12]。

4. 疾患オントロジーの Linked Open Data 化

4.1 LOD 化の基本方針

疾患オントロジーを LOD 化するにあたり、最も容易な方法としては、「法造」のオントロジーエクスポート機能を用いて、疾患オントロジーを法造形式から OWL 形式のオントロジーに変換すればよい。しかし、OWL 形式に変換した疾患オントロジーをそのまま LOD として公開したとすると、疾患の因果連鎖を取得するために複雑な SPARQL クエリを用いる必要がある。例えば、異常状態 A から結果として起こりうる異常状態を辿る際に

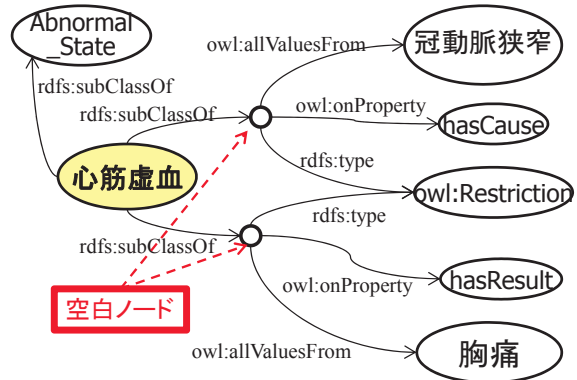


図3 OWL を用いた汎用疾患連鎖の表現例

は、図3のような空白ノードを含む RDF グラフの検索を繰り返す必要がある。さらに、図1のようなある疾患定義に含まれる因果連鎖を取得しようとする時、図3のような構造を持つグラフの検索に加え、上位疾患から継承された owl:Restriction を考慮した、更に複雑なクエリが必要となる。これは、結果として得られる図1のような因果連鎖を表すグラフ構造とは直観的に大きく異なり、LOD として公開した際の利便性を考慮すると、あまり好ましくないと思われる。

一般に、推論機構を用いて「ある因果連鎖のインスタンスが疾患オントロジー上で定義されたどの疾患に分類されるかを判定する」といった推論処理を行う際には、OWL 形式のオントロジーが用いられる。しかし、「疾患オントロジーを疾患定義が体系化された知識基盤として参照するのみで、特定の疾患のインスタンスを対象とした処理をしない」ことを前提とすると、より簡易な RDF 表現を用いることにより、クラス定義として記述された情報を効率的かつ直感的なクエリで取得できるような RDF データモデルを用いる利点が大きいと考えられる。

そこで本研究では、後者の立場から、LOD としての利用する際の利便性を考えた、疾患オントロジーの LOD 化を検討した。なお、初期的な検討を目的としたため、LOD 化の対象は疾患オントロジーのなかでも中心となる疾患連鎖に限定した。

4.2 疾患連鎖 LOD の RDF データモデル

前節で述べた方針のもと、疾患オントロジーから疾患連鎖に関する情報のみを抽出して RDF 化した疾患連鎖 LOD を構築した。疾患連鎖 LOD において、本来は「クラス(概念)レベル」の情報である疾患および異常状態の定義を、LOD として利用する際の利便性を考え、全てインスタンス(RDF リソース)に変換している。

異常状態は Abnormal_State タイプ(クラス)のインスタンスとして表現され、異常状態間の因果関係は hasCause および hasResult プロパティを双方向に記述するこ

とで表現される。発生確率が高い因果関係を表現するときには、これらのサブプロパティとして定義である `hasProbableCause` および `hasProbableResult` を用いる。これらのプロパティは、各異常状態の原因/結果となりうる異常状態を表し、それらを全ての異常状態から辿ることで汎用疾患連鎖が得られる。

一方、疾患は `Disease` タイプのインスタンスとして表現され、疾患定義連鎖および疾患派生連鎖を構成する異常状態は、`hasCoreState` および `hasDerivedState` プロパティを用いて表される。また、通常の OWL や RDF(S) では `rdfs:subClassOf` を用いて `is-a` 関係が表されるが、疾患連鎖 LOD では各異常状態および疾患がインスタンスとして表現されているため、異常状態および疾患の `is-a` 関係を表すために、それぞれ `subStateOf` および `subDiseaseOf` というプロパティを導入した。

図 4 に疾患 A とその下位疾患 B の疾患連鎖（上図）を RDF で表現した例（下図）を示す。各疾患を構成する疾患定義連鎖および疾患派生連鎖を取得するには、各疾患との間に `hasCoreState` および `hasDerivedState` プロパティを持つ異常状態を全て取得し、それらの異常状態間の `hasCause`/`hasResult` プロパティを辿ればよい。ここで、各異常状態は、3.2 節で述べた疾患オントロジーにおける扱いは異なり、「汎用疾患連鎖を表す異常状態」と「疾患の構成要素となる異常状態」は区別せずに、同一のインスタンスとして表されている。しかし、`hasCause`/`hasResult` プロパティでつながった異常状態が同じ疾患との間に `hasCoreState`/`hasDerivedState` プロパティを持つか否かで、その因果関係がその疾患の疾患定義連鎖/疾患派生連鎖に含まれるか否かの判定ができる。図 4 の疾患 A の場合、異常状態 2 の結果として異常状態 4 が記述されているが、異常状態 4 は疾患 A との間に `hasCoreState` も `hasDerivedState` も持たないので、疾患 A の定義連鎖にも疾患派生連鎖にも含まれないことが分かる。

なお、疾患 B のように上位疾患を持つ疾患の疾患連鎖を取得する場合には、`subDiseaseOf` を辿って得られる上位疾患に含まれる疾患連鎖を合わせて取得する必要がある。このような処理を避け、上位疾患の有無に関わらず同じ方法で疾患連鎖を取得できるようにするには、あらかじめ全上位疾患から継承された構成要素となる異常状態に対して `hasCoreState`/`hasDerivedState` プロパティを持たせておくという方法が考えられる。しかし、この方法では、下位の疾患になればなるほど上位と同様に記述を多く持たせる必要があり冗長となるため、本疾患連鎖 LOD ではこの方法を採用しなかった。また、3.2 節では紙面の都合で説明を省略しているが、疾患の構成要素となる異常状態が下位疾患でより詳細な異常状態に特殊化される場合も、異常状態間の `isSpecializedFrom` プロパティで適切に表現できる。

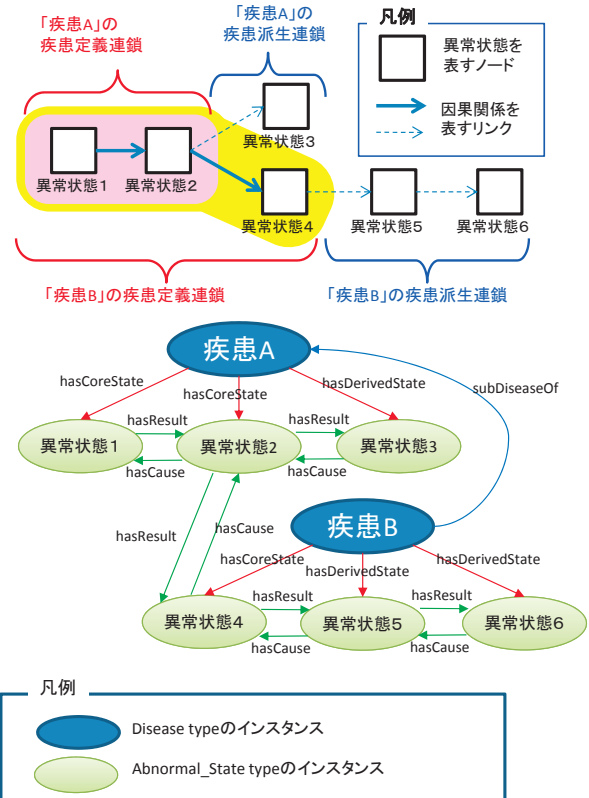


図 4 疾患連鎖（上図）の RDF 表現（下図）例

4.3 SPARQL クエリの例

図 5 に疾患連鎖 LOD から異常状態および疾患定義に関する情報を取得する SPARQL クエリの例を示す^{*12}。全ての異常状態の一覧を取得するには、(a1) に示すように `Abnormal_State` タイプの全インスタンスを取得すればよい。一方、指定した異常状態の原因/結果となる異常状態を取得するには、`hasCause`/`hasResult` プロパティをたどればよい。例えば、(a2) は指定した異常状態の原因となる異常状態を全て取得するクエリである。さらに、指定した異常状態を通る因果連鎖（一般派生連鎖）に含まれる全異常状態は、(a3) のクエリで取得することができる。

続いて、疾患定義に関するクエリ例について述べる。全ての疾患の一覧を取得するには、異常状態の場合と同じように、(d1) で示すクエリで `Disease` タイプの全インスタンスを取得すればよい。また、ある疾患の上位概念となる疾患（以下、上位疾患と呼ぶ）の一覧は、`subDiseaseOf` プロパティを再帰的に辿る (d2) のクエリで取得できる。一方、ある疾患定義に含まれる、疾患定義連鎖および疾患派生連鎖を得るには、上位疾患の定義を考慮しない場合は、それぞれ (d3) および (d4) に示すクエリを用いて、指定した疾患と `hasCoreState` および `hasDerivedState` プロパティを持つ異常状態を取得すればよい。さらに、(d2)、

*12 ここで示すクエリ例は、SPARQL 1.1 (<http://www.w3.org/TR/sparql11-query/>) を用いている。

```

(a1) 全ての異常状態を取得
select ?abn
  where { ?abn rdf:type dont:Abnormal_State }
(a2) 指定した異常状態を含む原因となる全異常状態を取得
select ?o
  where { <abn_id> dont:hasCause* ?o }
(a3) 指定した異常状態を通る因果連鎖(一般派生連鎖)に含まれる全異常状態を取得
select ?o
  where { {<abn_id> dont:hasCause* ?o}
         UNION {<abn_id> dont:hasPossibleCause* ?o}
         UNION {<abn_id> dont:hasResult* ?o}
         UNION {<abn_id> dont:hasPossibleResult* ?o} }
(d1) 全ての疾患を取得
select ?dis
  where { ?dis rdf:type dont:Disease }
(d2) 指定した疾患の上位概念となる疾患を全て取得
select ?o
  where { <dis_id> dont:subDiseaseOf* ?o }
(d3) 指定した疾患定義に含まれる疾患定義連鎖を取得
where { <dis_id> dont:hasCoreState ?o }
(d4) 指定した疾患定義に含まれる疾患派生連鎖を取得
select ?o
  where { <dis_id> dont:hasDerivedState ?o }
(d5) 指定した疾患定義に含まれる全因果連鎖を取得
select ?o
  where { <dis_id> dont:subDiseaseOf* ?dis .
         { ?dis dont:hasCoreState ?o }
         UNION { ?dis dont:hasDerivedState ?o } }
(d6) 指定した疾患定義に含まれる全因果連鎖を因果関係の一覧を取得
select DISTINCT ?abn1 ?p ?abn2
  where { <dis_id> dont:subDiseaseOf* ?dis .
         ?abn1 ?p ?abn2 .
         { ?dis dont:hasCoreState ?abn1 }
         UNION { ?dis dont:hasDerivedState ?abn1 }
         { ?dis dont:hasCoreState ?abn2 }
         UNION { ?dis dont:hasDerivedState ?abn2 } }

```

※abn_idには任意の異常状態, is_idには任意の疾患を指定する。

図5 SPARQL クエリの例

(d3), (d4) を組み合わせた (d5) のクエリを用いることで、上位疾患からの継承も考慮した、ある疾患定義に含まれる全因果連鎖を取得することができる。なお、ここまでの例では、因果連鎖に含まれる異常状態の一覧を取得するクエリを示したが、因果連鎖に含まれる「異常状態間の因果関係」を得るには、(d6) のようなクエリを用いれば良い。このクエリでは、指定した疾患定義を構成する因果連鎖に含まれる全異常状態間のプロパティを取得できる。

上述のように 4.2 節で示した RDF データモデルでは、疾患連鎖の情報が OWL 形式に比べて直観的に表現されており、必要な情報を比較的簡易なクエリで取得することができる。なお本データモデルでは、現状の疾患オントロジーで定義されている疾患連鎖に関する情報は、全て表現されている^{*13}。

*13 [山縣 12, Yamagata 12] で議論されている異常状態オントロジーの内容については、疾患オントロジー自体でも構築途中のため、現状の疾患連鎖 LOD には含まれていない。

4.4 RDF 化に伴う短所

前節で述べたように、疾患オントロジーにおける疾患連鎖の情報は、4.2 節で示した RDF データモデルですべて表現されている。しかし、OWL や RDF(S) といった標準化されたオントロジー記述言語で予め定義されている語彙を用いていないため、それらの記述言語に準拠したツールや推論機構を用いた意味処理が行えないという欠点がある。

例えば、概念間の is-a 関係は OWL や RDF(S) では `rdfs:subClassOf` 用いて表されるが、本モデルでは異常状態および疾患の is-a 関係を表すために、それぞれ `subStateOf` および `subDiseaseOf` という独自のプロパティを導入しているため、利用者はこの点を考慮した検索・推論処理を行う必要がある。前述のように、上位概念の一覧の取得 (図 5(d2)) や上位概念から継承されるプロパティの取得 (図 5(d5),(d6)) であれば、比較的簡単な SPARQL クエリで行うことが出来る。しかし、OWL におけるプロパティの対称性 (`owl:SymmetricProperty`)、推移律 (`owl:TransitiveProperty`) などの特性記述といった、より高度な推論を必要とする場合は、独自の RDF データモデルで表現したオントロジーでは処理が複雑になると思われる。

すなわち、OWL のオントロジー記述言語を用いないということは、その言語で標準的にサポートされており、そして多くのツールで標準的に実装されている推論処理を適用することができなことを意味する。この欠点は独自の推論機構を実装することである程度は軽減することができるが、一般に、OWL や RDF(S) に準拠した推論処理が多く必要とされる場合には、本論文で示したオントロジーを独自のデータモデルに基づいて RDF 化することは好ましくないとされる。この点については、本論文の手法を一般化する際に、より詳細な考察が必要な課題である。

なお、本論文で対象とした疾患オントロジーにおいては、OWL における推論処理の利用を前提としておらず、疾患連鎖に基づく疾患定義の情報を取得することが主として想定されている利用方法であるため、この範囲においては推論機構を利用できないという短所は問題にならない。ただし、より汎用性・利便性を高めるために、`rdfs:subClassOf` などの基本的な語彙の一部については、クエリが複雑にならない範囲で使用することも検討したいと考えている。

4.5 疾患連鎖 LOD の試作と公開

4.2 節で述べたデータモデルに基づき、疾患オントロジーから抽出した疾患連鎖を RDF 形式に変換し、LOD として公開した^{*14}。変換元の疾患オントロジーは代表的な 6 診療科を対象とした 2013 年 1 月 29 日時点のもの

*14 <http://lodc.med-ontology.jp/>, 疾患連鎖 LOD の公開サイト

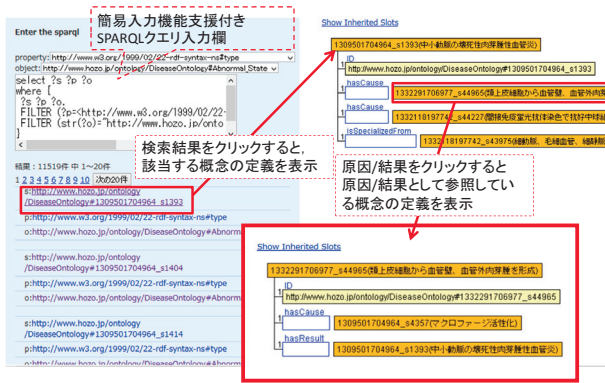


図6 疾患連鎖 LOD の SPAQL エンドポイント

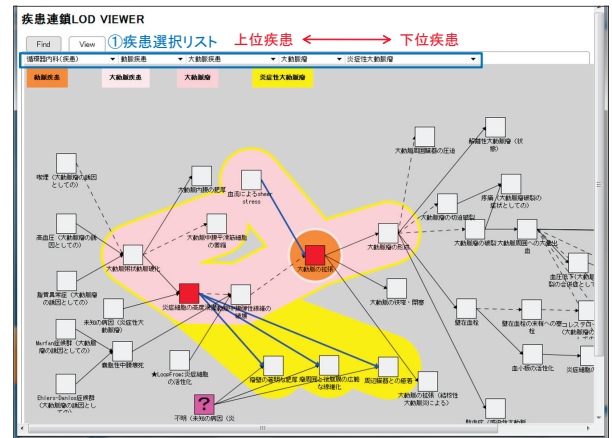


図7 疾患連鎖 LOD の可視化ツール

で、約 1,800 の疾患定義とそれを構成する約 12,000 の異常状態の定義を含む。なお、すべての疾患および異常状態を表す RDF リソースには、それぞれ日本語および英語のラベルを `rdfs:label` を用いて付与している。

合わせて疾患連鎖 LOD を外部システムから検索・利用するための API (SPARQL エンドポイント) と、疾患連鎖を図 1 のようなグラフ構造で可視化する疾患連鎖 LOD Viewer を開発した。SPARQL エンドポイント (図 6) には、初心者でもリストを選択するだけで簡単なクエリを試すことができる「簡易クエリ入力支援機能」と、検索結果から選択した RDF リソースを可視化する機能、および、検索結果に疾患が含まれる場合は「疾患連鎖 LOD Viewer」と連携して疾患連鎖を可視化する機能も実装した。

一方、疾患連鎖 LOD Viewer (図 7) は、

- 「疾患名」、「その疾患に含まれる異常状態名」から疾患定義を検索
- リスト形式で表示した「疾患の is-a 階層」(図 7-①) から表示する疾患を選択

の何れかの方法で選択した疾患の疾患連鎖を可視化することができる。本 Viewer は PC のみならず、タブレット端末、スマートフォンなど様々なデバイスから利用可能な Web サービスとして開発されている^{*15}。技術的には、疾患連鎖 LOD の SPARQL エンドポイントに対して選択した疾患の疾患連鎖の情報を取得するのに必要な SPARQL クエリを発行し、その結果を用いて可視化している。すなわち、本 Viewer で疾患連鎖が正しく表示されるということは、3.2 節で議論した方法に沿って正しく疾患連鎖の情報が取得できていることを示している。表示する疾患の種類やクライアントの環境にも依存するが、無線 LAN 接続下において疾患連鎖が表示されるまでに要する時間は、概ね 2 秒前後であり、現時点においては十分な速度であると思われる。なお上述のシステムの開発にあたり、RDF データベースには AllegroGraph4.10 を可視化部分には HTML5 を利用している。

*15 ただし、HTML5 に対応したブラウザが必要。

5. 関連研究

2 章で述べたように、多くのオントロジーが LOD として公開されているが、それらの多くがセマンティック Web の標準技術として広く使われているオントロジー記述言語である OWL, RDF(S) を用いてクラス定義を表現している。一方、本研究では、オントロジーを Linked Data として利用する際の利便性を考慮し、クラス定義をインスタンスとして表現するために、独自の RDF データモデルを導入している。

クラス定義のような概念レベルの情報を Linked Data として公開するにあたり、本研究と同様に独自の RDF データモデルを導入するアプローチを採用している例は、いくつかの研究で見られる。例えば、Assem らは WordNet [Assem 06a, 06b] の RDF 表現について考察し、Linked Data として公開している。同様の議論は小出らも行っている [Koide 06, 小出 11]。これらの研究では、ある単語に対する同義語集合である *synset* を表現することに主眼が置かれている。また、Nuzzolese らは、FrameNet と呼ばれる言語リソースを Linked Data として公開するためのパターンについて議論しており、FrameNet を RDF および OWL に変換する手法を提案している。これらの研究と比較した本研究の特徴は、LOD 化の対象としている疾患オントロジーにおける疾患概念のクラス定義が、単に属性を列挙したのではなく、因果連鎖という比較的複雑な概念構造により表されている点にある。そのため、クラス定義の概念構造を取得するには、より直感的に理解しやすいクエリを用いることが重要になると考えられる。

一方、OWL 形式を用いて公開したオントロジーを、利用しやすくする別のアプローチとして、必要な情報を取得するためのクエリの複雑さを、独自のクエリ言語や検索用インタフェースを導入することで吸収するという手法もある。OWL 用の検索言語として DL-query が提案されており、Protégé などのオントロジー構築環境でもサポー

トされている。また、BioPortal [Whetzel 11] や Spatial Decision Support^{*16} [Raskin 12] といったオントロジー公開用のポータルサイトでは、SPARQL に依らないオントロジー検索や、RDF グラフをそのままではなくオントロジーの表現に特化した形式で可視化する GUI が用いられている。オントロジーの検索・閲覧を行う際には、これらのアプローチも有用であると思われる。しかし、複数の Linked Data を横断的に利用したシステムを開発する際には、全て SPARQL で検索できる方がより効率的であり、その際、直感的なクエリで必要な情報を取得できることが、利便性の観点から重要と思われる。

特に疾患オントロジーが対象としている疾患定義は、対象となる部位といった解剖学的情報、各疾患に関連する生命科学分野の各種データ（疾患に関わる遺伝子、臨床データ、治療方法、薬剤、など）と結びつけることで、新たな知見を得られることが求められている。これらの生命科学データの Linked Data 化は、国外では米国の NCBO (The National Center for Biomedical Ontology)^{*17}、国内ではバイオサイエンスデータベースセンター (NDBC)^{*18} が推進しており、それらのデータベースと疾患オントロジーを統合し横断的に検索することで、様々な応用システムの開発できると期待される。

6. まとめと今後の課題

本論文では、臨床医学オントロジーを知識基盤として活用するために LOD として公開する際の手法について、疾患オントロジーを例として考察した。その結果、疾患オントロジーにおける概念（クラス）定義を RDF におけるインスタンスとして表現することによって、検索などの利用時に直観的かつ簡易なクエリで必要な情報を取得できるような LOD 化が可能であることを示した。

解剖学オントロジーにおける血管や神経の接続構造など、臨床医学オントロジー^{*19}における他のオントロジーについても、同様のアプローチが適用可能と思われる。しかしその際に、対象に応じて異なる RDF データモデルを検討することは、臨床医学オントロジー全体の LOD 化を考えた際に一貫性や整合性を乱す恐れがある。そのような問題を避けるために、対象とする概念の種類に依らず適用可能な、オントロジーの概念定義を RDF インスタンスに変換する汎用ルールを導入の検討が必要であると考えられる。

その一方で、OWL や RDF(S) など既存のオントロジー記述言語をそのまま RDF としてもものを LOD として公開し、必要な情報を取得するためのクエリの複雑さは、検索に用いる RDF グラフをパターン化することで吸収するというアプローチも考えられる。上述の BioPortal [Whetzel

11] など、SPARQL クエリを用いずキーワード等の条件を入力することでオントロジーの概念定義が検索できるシステムは、技術的にはこのアプローチを採用したものとと言える^{*20}。

いずれのアプローチを適用すべきは、公開した LOD の用途や規模等によっても変わると思われる。臨床医学オントロジーのように大規模かつ多様な概念定義を含むオントロジーの LOD 化において、これらのアプローチの特徴を比較・検討することは、今後の重要な検討課題である。疾患オントロジーの構築作業は継続しており、定義内容の更なる精査や対象とする疾患の拡充、疾患連鎖以外の定義内容（疾患の属性など）の追加や異常状態オントロジーの整備 [山縣 12, Yamagata 12] などの作業が進められている。また、疾患連鎖 LOD を含めた臨床医学オントロジーの公開に用いる id 管理と URI の設計も、今後の課題として残っている。

なお、公開された疾患連鎖 LOD は、Linked Open Data チャレンジ Japan2012 において「ライフサイエンス賞」を受賞し、臨床医学に基づいた知識を体系化した非常に有用なデータセットとして高い評価を得ている。現在、マウスの表現型データベース [榎屋 13] など、他のデータと疾患連鎖 LOD を連携させることによる応用システムの開発も進めており、今後の更なる活用が期待される。

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。これらの研究・開発に参加された方々（システム実装を担当された加藤敦丈氏、他）に感謝致します。特に、疾患オントロジー記述のご尽力と臨床医学的な知識の提供を頂いた東大病院の臨床医の先生方（大友夏子先生、林亜紀先生、松村貴由先生、桜井亮太先生、寺田さとみ先生、脇嘉代先生ら）に心よりの感謝致します。

*16 <http://www.spatial.redlands.edu/sds/>

*17 <http://www.bioontology.org/>

*18 <http://biosciencedbc.jp/>

*19 <http://www.med-ontology.jp/>

*20 実装レベルで内部的に SPARQL クエリを生成しているか否かは、システムが採用しているデータベースの種類によると思われる

◇ 参 考 文 献 ◇

- [Assem 06a] van Assem, M., Gangemi, A., Schreiber, G., (eds.): RDF/OWL Representation of WordNet, W3C Working Draft 19 June 2006, <http://www.w3.org/TR/2006/WD-wordnet-rdf-20060619/>
- [Assem 06b] van Assem, M., Gangemi, A., Schreiber, G.: Conversion of WordNet to a standard RDF/OWL representation, Proc. LREC (2006)
- [Cowell 10] Cowell, L. G. and Smith, B: Infectious Disease Ontology. Infectious Disease Informatics, Chapter 19, Sintchenko V., 373-395 (2010)
- [Dimitrios 11] Dimitrios A. Koutsomitropoulos, Ricardo Borillo Domenech, Georgia D. Solomou: A Structured Semantic Query Interface for Reasoning-Based Search and Retrieval, In Proc. of ESWC, Lecture Notes in Computer Science, vol.6643, pp.17-31 (2011)
- [Gruber 92] Gruber, T.: A Translation Approach to Portable Ontology Specifications, in Proc. of JKAW'92, pp. 89-108 (1992)
- [Koide 06] Koide, S., Morita, T., Yamaguchi, T., Muljadi, H., Takeda, H.: RDF/OWL Representation of WordNet 2.1 and Japanese EDR Electric Dictionary, 5th International Semantic Web Conference (ISWC2006), Poster (2006)
- [小出 11] 小出 誠二, 武田 英明, 大向 一輝: WordNet 日本語化への LOD アプローチ, 人工知能学会研究会資料, SIG-SWO-A1103-05, (2011)
- [古崎 08] 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, 人工知能学会研究会資料, SIG-SWO-A802-09 (2008)
- [古崎 11] 古崎 晃司, 他: 臨床医学オントロジーにおける疾患連鎖モデルの考察, 人工知能学会研究会資料, SIG-SWO-A1102-08 (2011)
- [Li 12] Li, N., Raskin, R., Goodchild, M., Janowicz, K.: An Ontology-Driven Framework and Web Portal for Spatial Decision Support. Transactions in GIS, Vol.16, No.3, pp.313-329 (2012).
- [榎屋 13] 榎屋 啓志, 古崎 晃司, 大江 和彦, 溝口 理一郎 (大郎): コンテキストに依存した定性値を扱う生物表現型統合データベースの試作, 2013 年度人工知能学会全国大会 (第 27 回) 資料集, 311-2 (2013)
- [溝口 05] 溝口 理一郎: オントロジー工学, オーム社 (2005)
- [Mizoguchi 09] Mizoguchi, R., Kou, H., Zhou, J., Kozaki, K., Imai, T., Ohe, K.: An Advanced Clinical Ontology, In Proc. of ICBO, Buffalo, NY, pp.119-122 (2009)
- [Mizoguchi 11] Mizoguchi, R., Kozakil, K., Kou, H., Yamagata, Y., Imai, T., Waki, K., Ohe., K.: River Flow Model of Diseases, Proc. of ICBO2011, Buffalo, USA, pp.63-70 (2011)
- [Nuzzolese 11] Nuzzolese, A. G., Gangemi, A., Presutti., V.: Gathering Lexical Linked Data and Knowledge Patterns from FrameNet. In Proc. of the 6th International Conference on Knowledge Capture (K-CAP), pp.41-48, Ban, Alberta, Canada (2011)
- [大江 10] 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)
- [Osborne 09] Osborne, J. D., et al.: Annotating the human genome with Disease Ontology. BMC Ge-nomics 10(1):S6 (2009)
- [Scheuermann 09] Scheuermann, R. H., Ceusters, W., and Smith, B.: Toward an Ontological Treatment of Disease and Diagnosis. Proc. of the 2009 AMIA Summit on Translational Bioinformatics, pp.116-120, San Francisco, CA (2009)
- [玉川 11] 玉川 奨, 森田 武史, 山口 高平: 日本語 Wikipedia からプロパティを備えたオントロジーの構築, 人工知能学会論文誌, Vol.26 No.4 pp.504-517 (2011)
- [Tamagawa 12] Tamagawa, S., Morita, T., Yamaguchi, T.: Extracting Property Semantics from Japanese Wikipedia, In Proc. of 2012 International Conference on Active Media Technology (2012)
- [Whetzel 11] Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., Musen, M. A.: BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications, Nucleic Acids Research, Vol.39, suppl.2, W541-545 (2011)
- [山縣 12] 山縣 友紀, 国府 裕子, 古崎 晃司, 今井 健, 大江 和彦, 溝口 理一郎: 異常状態オントロジーとその応用, 2012 年度人工知能学会全国大会 (第 26 回) 資料集, I12-R-4-3 (2012)
- [Yamagata 12] Yamagata, Y. and et.al.: Ontological Modeling of Interoperable Abnormal States, In Proc. of the 2nd Joint International Semantic Technology Conference (JIST2012), Lecture Notes in Computer Science, vol.7774, pp.33-48, Nara, Japan, Dec.2-4 (2012)

〔担当委員: 山本 泰智〕

2013 年 9 月 2 日 受理

著者紹介



古崎 晃司(正会員)

1997年大阪大学工学部電子工学科卒業。2002年同大学院工学研究科博士後期課程修了。同年、化学工学会嘱託研究員、同年12月大阪大学産業科学研究所助手、2008年同准教授、現在に至る。博士(工学)。オントロジー工学の基礎理論、オントロジー構築・利用環境の設計・開発、セマンティック Web, Linked Data, 医療, 環境など各種領域におけるオントロジー開発・応用に関する研究に従事。Linked Open Data チャレンジにおいて、ライフサイエンス賞(2013年)、アプリケーション部門優秀賞(2014年)を受賞。2014年、オープンデータ・アプリコンテスト技術賞受賞。情報処理学会、電子情報通信学会、International Association for Ontology and its Applications、各会員。



山縣 友紀(正会員)

大阪大学産業科学研究所知識科学研究分野、特任助教。大阪大学大学院医学研究科医科学専攻修士課程修了。科学技術振興機構バイオインフォマティクス推進センターにてシグナル伝達パスウェイオントロジー開発後、知財関連業務等を経て、現在、大阪大学産業科学研究所知識科学分野にて臨床医学オントロジーなど生命科学分野のオントロジー開発・応用に関する研究に従事。日本医療情報学会、日本バイオインフォマティクス学会、日本弁理士会各会員。



国府 裕子(正会員)

1999年大阪大学医学部保健学科卒業。2004年同大学院医学系研究科博士課程修了。同年、独立行政法人産業技術総合研究所人間福祉医工学部門特別研究員、2007年大阪大学産業科学研究所知識システム研究部門特任助教、2013年独立行政法人情報通信研究機構脳情報通信融合研究センター技術員、現在に至る。保健学博士、医療情報学、数理統計学を基礎とするコンピュータによる医療支援を軸とし、臨床医学オントロジー構築やマルチモダリティ脳機能計測法の開発、再生医学領域における心筋梗塞モデルを用いた画像評価等の研究および産学官連携組織、脳情報通信研究センターの運用業務に従事。2003年生体医工学シンポジウムベストアワード賞、2012年医療情報学会第28回医療情報学連合大会研究奨励賞、2014年度情報通信研究機構成績優秀表彰、優秀賞受賞等。



今井 健(正会員)

1999年東京大学理学部数学科卒業、2005年同大学院情報学環・学際情報学府修士課程修了。2005年博士課程修了。博士(学際情報学)。同年~2008年同医学部附属病院特任助教。医療分野の自然言語処理とオントロジー構築に関する研究でそれぞれ2004年・2007年医療情報学連合大会若手最優秀賞受賞。2008~09年米国 Mayo Clinic 客員研究員(兼任)、2009~東京大学大学院医学系研究科助教を経て、2012年より東京大学医学部附属病院助教。日本医療情報学会評議員。日本医療情報学会、情報処理学会、言語処理学会、日本医療・病院管理学会各会員。



大江 和彦(正会員)

1984年東京大学医学部医学科卒業、東大病院外科研修医、1989年東京大学医学系大学院博士課程(第一基礎医学医療情報学)退学、東大病院中央医療情報部助手、1990年同講師、1991年医学博士、1994年同助教、1997年東京大学大学院医学系研究科医療情報経済学分野教授、2003-2005年東大病院副院長兼任、2011-2013年日本医療情報学会理事、副会長。専門は医療情報の標準化、医学知識の計算機処理、医療情報システム。日本医療情報学会評議員。同学会、情報処理学会、各会員。



溝口 理一朗(正会員)

1977年大阪大学基礎工学研究科博士課程修了。同年、大阪電気通信大学工学部講師、1978年大阪大学産業科学研究所助手、1987年同研究所助教、1990年同教授、2013年10月北陸先端科学技術大学院大学教授、現在に至る。工学博士。パターン認識関数の学習、クラスタ解析、音声の認識・理解、エキスパートシステム、知的学習支援システム、オントロジー工学の研究に従事。1985年 Pattern Recognition Society 論文賞、1988年電子情報通信学会論文賞、1996年人工知能学会創立10周年記念論文賞、1999年、2006年 ICCE Best paper Awards、2005年大川出版賞(オントロジー工学)、2006年、2013年人工知能学会論文賞、2010年教育システム情報学会論文賞受賞。人工知能学会編集委員会委員長、教育システム情報学会理事、同編集委員長、J. of Web Semantics Editors-in-Chief、Intl. AI in Education (IAIED) Soc. President、APC of AACE President、人工知能学会会長、Vice-President of Semantic WebScience Association(SWSA)を歴任。現在、Associate editors of IEEE TLT and ACM TiiS。