

# 臨床医学オントロジーに基づく医学電子辞書の自動生成

## Automatic generation of medicine electronic dictionary based on Clinical Medical Ontologies

周俊<sup>\*1</sup> 国府裕子<sup>\*1</sup> 古崎晃司<sup>\*1</sup> 今井健<sup>\*2</sup> 大江和彦<sup>\*2</sup> 溝口理一郎<sup>\*1</sup>  
Jun ZHOU, Hiroko KOU, Kouji KOZAKI, Takeshi IMAI, Kazuhiko OHE, Riichiro MIZOGUCHI

<sup>\*1</sup> 大阪大学産業科学研究所  
The Institute of Scientific and Industrial Research (ISIR),  
Osaka University

<sup>\*2</sup> 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

In order to explain the content of Clinical Medical Ontologies plainly, the technology to express contents by natural language is important. So in this paper we developed "Natural language explanation generation function" as a basic technology expressing defined contents of diseases utilizing Clinical Medical Ontologies. Furthermore, as to the searched diseases according to names and symptoms, we developed an electronic dictionary provides reconversion of explanation according to patients and doctors, using the same ontology.

### 1. はじめに

近年、日本における医療情報の発達は、様々な異なる領域のトピックを医療分野に持ち込んだ。加えて、日本政府は、IT新改革戦略の重点課題として世界最先端の国民健康情報基盤の構築をあげており、医療情報の電子化を推進してきた。このような背景により、日本の臨床現場に適したオントロジーの構築が必要とされ、厚生労働省の支援のもと、筆者らは大規模な臨床オントロジー開発を進めている[大江 08, 国府 08]。臨床医学オントロジーは様々な知識処理アプリケーション開発において、体系化された医療知識を必要に応じて適切に提供するという中核技術としての重要な役割を果たす。このためオントロジーの可視化、マイニング、アライメントといったオントロジーを利用するための技術やツールの開発も盛んに行われている[溝口 05]。より良いオントロジーはその構築自体に高い価値があり、その価値を十分に活かすにはオントロジーを適切に利用する基盤技術の開発が重要となる。

例えば、臨床医学オントロジーで定義された概念定義やそれらの関係性を適切に提示することで、疾患定義の説明提示、医学電子辞書の自動生成、医学生が疾患概念構造を理解することを支援するための新しいタイプの電子医学教科書(医学知識ナビゲータ)[国府 09]といった応用システムの開発が期待される。このようなアプリケーションの実現のための基盤技術の1つとして、臨床医学オントロジーの内容を分かりやすく説明するために、自然言語を用いた説明文生成技術[Bontcheva 09]が重要である。

そこで本研究では、臨床医学オントロジーを用いて、疾患概念の定義内容を自然言語で説明する「自然言語説明文生成機能」を開発した。さらに、その自然言語説明機能を用いて、同一のオントロジーから利用者に応じて説明内容の詳細度を切換えて提供できる医学電子辞書の自動生成システムを開発した。以下、2章では医学知識を自然言語で説明する際に検討すべき点を述べる。続く3章では、本システムが利用した臨床医学オントロジーの概要を示し、4章では本研究で開発した自然言語説明文生成機能について述べる。そして5章では本機能を用いた応用システムの一例として開発した医学電子辞書について述べ、6章で本報告の総括と今後の課題を述べる。

### 2. 自然言語による医学知識の説明

先に述べたように、臨床医学オントロジーを用いた応用システムの1つとしてオントロジーを用いた知識提供が挙げられる。その際、提供する知識をどのメディアで表出するかは重要な課題である。ここでいうメディアとは、知識を表現する媒体のことで、たとえば自然言語、グラフ、テーブルなど、どのメディアを選択するかで説明のわかりやすさに大きく影響する。特に、自然言語は人間にとって理解しやすく、多くのシステムでの利用が想定される。本研究では、臨床医学オントロジーで定義されている疾患概念を対象として、オントロジーに基づく説明文章生成機能を開発する。

医学のような専門性の高い知識を、人に分かりやすく説明するためには、相手に応じて説明文のレベルを変えて提示する必要がある。例えば同じ疾患でも、患者に説明する際に使う文章と、医師がその疾患について学習したり、知識を確認したりするために用いる文章とは、注目する点や必要な内容が異なる。よって、良い自然言語説明文生成機能には、同じオントロジーから患者と医師といった利用者に応じて、提供する説明文章の切換えが可能であることが望まれる。そこで我々は、臨床医学オントロジーにおける疾患概念の定義内容を用いて患者と医師といった利用者に応じて説明に必要な情報を取り出し、異なるレベルの説明文章を出力可能な自然言語説明文生成機能を開発する。この機能は、様々なアプリケーションから利用可能な汎用的なもので、疾患概念説明文自動生成し提供するWebシステム、医学電子辞書、電子医学教科書などの複数のシステム開発に有用であると考えられる。

### 3. 臨床医学オントロジー

本臨床医学オントロジーの基本方針として、概念の本質構造を捉えた汎用性・共通性を追求した概念化、本質を捉えた統一的視点による概念化と多角的視点による対象依存の概念化の峻別、多様な概念に対応可能な統一された概念記述フレームの導入などを重視している[国府 08]。また上位オントロジーには、その哲学的考察が深いと評価されるグアリーノの上位オントロジーを参考にしつつ、表現や属性などの扱いをより厳密に考察された溝口らの上位オントロジー<sup>1</sup>[Mizoguchi 09]を基盤とした。

連絡先: 古崎晃司, 大阪大学産業科学研究所,  
567-0047 大阪府茨木市美穂が丘 8-1,  
Tel: 06-6879-8416, E-mail:kozaki@ei.sanken.osaka-u.ac.jp

<sup>1</sup> [http://www.ei.sanken.osaka-u.ac.jp/hozo/onto\\_library/upperOnto.htm](http://www.ei.sanken.osaka-u.ac.jp/hozo/onto_library/upperOnto.htm)

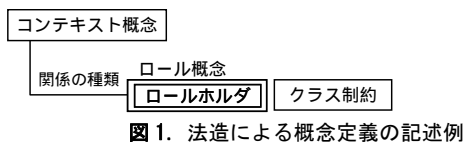


図 1. 法造による概念定義の記述例



図 2. 臨床医学オントロジーの疾患記述

本臨床医学オントロジーの構築には、ロール概念を中心としたオントロジー基礎理論に基づいて開発されたオントロジー構築・利用環境「法造」<sup>1)</sup>を用いた。法造では、オントロジーにおける概念定義は複数のスロットにより表され、各スロットはロール概念、ロールホルダー（以下、RH と記す）とクラス制約（以下、CC と記す）の組み合わせで記述される(図 1)[古崎 02]。疾患は多数の異常状態が主病態や症状、原因といった役割(ロール概念)を担ったもの(RH)の組み合わせであると捉えられ、法造上では図 2 のように表される。

例えば、糖尿病では、「持続的高血糖状態」(CC)という異常状態が主病態(ロール概念)の役割を担い「糖尿病性持続的高血糖状態」(RH)となることや、その原因となる異常状態が「インスリン作用の低下」(CC)であることが定義されている。そして、これらの異常状態はすべて「対象」の「属性」が「異常値」であるという OAV(Object/Parameter/Value)表現で記述される(図 2(a))。また原因となる異常状態を原因スロットで参照することにより、その疾患における異常状態(主病態や症状など)の因果関係が表されている(図 2(b))。糖尿病の例では「持続的高血糖状態」は「血液(O)」の「血中グルコース(A)」が「高い(V)」という枠組みで記述され、その原因は「インスリン作用の低下」であることが原因スロットで記述されている[国府 08]。

#### 4. オントロジーに基づく自然言語説明文生成機能

##### 4.1 自然言語説明文生成テンプレート

本研究では、臨床医学オントロジーから自然言語説明文を生成するテンプレートを用意し、それを生成する説明文の対象や詳細度に応じて組み合わせ適用することで、疾患の説明文を

表 1. 説明文生成テンプレート一覧

	患者用		医師用
	簡易版1	簡易版2	詳細版
上位概念の説明	○	○	○
1段目のスロット	○※	○	○
2段目以降のスロット(OAV表現, 原因)	×	×	○

※クラス制約 (CC) によりテンプレートを切り替える

<sup>1)</sup> <http://www.hozo.jp>

表 2. 上位概念の説明文生成テンプレート

疾患の説明用テンプレート(1)
[疾患概念] は, [上位概念] の一種である。 例) 1型糖尿病は糖尿病の一種である

生成する機能を実現した。用意したテンプレートは、

- (1) 疾患の上位概念の説明
- (2) 1段目のスロット(主病態, 症状など)の説明
- (3) 2段目以降のスロット(異常状態の OAV 表現および、原因)の説明

の 3 種類である。これらのテンプレートを組み合わせることで、利用者に応じた説明文の詳細レベルを切り換える(表 1)。以下、それぞれのテンプレートの詳細を述べる。

##### (1) 上位概念の説明文生成

オントロジーにおいては is-a(上位-下位)関係を用いて、概念の分類階層が表される。疾患概念を説明する際には、その上位概念を用いて、その疾患がどのような分類に属するかを説明することができる。例えば、「1型糖尿病」の上位概念は「糖尿病」であるが、この情報を用いて「1型糖尿病は糖尿病の一種である」と説明することができる(表 2)。

臨床現場においては、専門家はそれぞれの専門分野に応じた視点や目的で疾患概念を分類する。よって、実用的な説明文生成システムにおいては、このような複数の視点に応じた分類階層で疾患を説明することが求められる。例えば、内科医に(1型)糖尿病を説明する時には、「(1型)糖尿病は代謝性疾患の一種である」と説明し、患者に対してはなじみが深い一般的な言葉(概念)での説明が分かりやすいので「糖尿病は生活習慣病の一種である」という簡単な説明に切り換えることが好ましい。本研究で利用する臨床医学オントロジーは、このような視点に応じた分類階層の動的生成が可能な概念構造を提供しており、別途開発を進めている動的な分類階層生成機能を本研究の説明文生成機能と組み合わせることでこれを実現できる。

##### (2) 1段目のスロットの説明文生成

疾患概念の 1 段目のスロットには、その疾患が持つ基本的な定義が記述されている。これを用いて概念を説明する場合、そ

表 3. 1 段目のスロットの説明文生成テンプレート

	適用条件	疾患の説明用テンプレート(2)
(2)-1	RH の定義あり	[疾患概念] は, [ロール概念] として [CC] をもち、それは [RH] と呼ばれる。
(2)-2	RH の定義なし	[疾患概念] は, [ロール概念] として [CC] をもつ。

表 4 1 段目のスロットの説明文生成例

RHがある	糖尿病は、主病態として持続的高血糖状態をもち、それは糖尿病性持続的高血糖状態と呼ばれる。 糖尿病 p/o 1. 主病態 [+] 糖尿病性持続的高血糖状態 持続的高血糖状態[RH]
RHがない	虚血性心疾患は、主病態として心筋虚血状態をもつ。 虚血性心疾患 p/o 1. 主病態 [+] 心筋虚血状態[RH]

のスロットにおいてロールホルダー (RH) が定義されている場合とされていない場合<sup>1</sup>があり、それぞれ利用されるテンプレートが異なる。表 3 に、それぞれの説明文生成テンプレートを示す。また、このテンプレートを用いて、糖尿病および虚血性心疾患の説明文を生成する具体例を表 4 に示す。

ここで、テンプレート(2)-1 を用いた説明文を生成する際、クラス制約として参照している概念 (CC) が抽象度の高い概念であると患者向けには言葉の意味が分からず、わかりにくい説明になる場合がある。例えば「糖尿病」の定義において「多尿」は「属性値関連異常」をクラス制約 (CC) とするロールホルダー (RH) として定義されているが、患者向けの簡易な説明文としては、「多尿」という RH のみを用いることが好ましいと思われる。一方、「糖尿病性持続的高血糖状態」といったより特殊化された RH である場合は、その CC (この例では「持続的高血糖状態」) のみを説明に用いることが適していると考えられる。どちらの説明文を採用するかはオントロジーの定義内容に依存するが、本研究で用いた臨床医学オントロジーではドメイン依存性の分離を厳密に行なっているため、これを説明文の切り替えの1つ基準として用いた。表 5 のこの条件を加えた説明文生成テンプレートを示す。

表 5. クラス制約のドメイン依存性で分けられたテンプレート

	適用条件	疾患の説明用テンプレート(2)
(2)-1'	CC がドメイン依存	[疾患概念]の[ロール概念]は[CC]である。
	CC がドメイン非依存	[疾患概念]の[ロール概念]は[RH]である。

(3) 2 段目以降のスロットの説明文生成

疾患にける 2 段目以降のスロットには、主に主病態や症状といった異常状態のより詳細な定義内容として、OAV 表現による異常状態の定義と、その異常状態の原因が定義されている (図 3)。これらの説明文を生成するテンプレートを表 6 に示す。

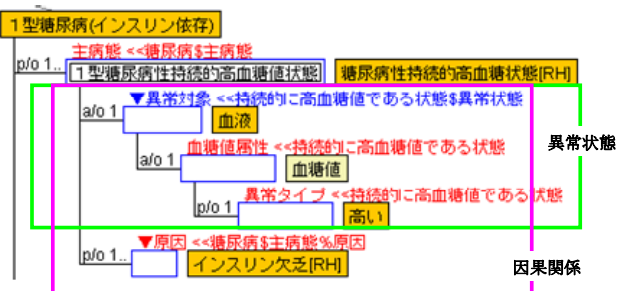


図 3. 1 型糖尿病の異常状態と因果関係の記述方法の例

表 6. 異常状態の定義 (OAV) と原因の説明テンプレート

疾患の説明用テンプレート(3)
[疾患概念]の[ロール概念]は[RH] ( [原因]を原因とする, [異常対象(O)]の[属性(A)]が[異常タイプ(V)]状態にあることを意味する[CC]) である。
例) 1 型糖尿病(インスリン依存)の主病態である 1 型糖尿病性持続的高血糖値状態は、インスリン欠乏を原因とする血液の血糖値の高い状態にあることを意味する続的高血糖値状態である。

<sup>1</sup> 厳密には、ロールホルダーの名前の定義が省略されている場合を指す。

表 7. 簡易版 1, 簡易版 2, 詳細版の説明文章の対比

レベル	生成された自然言語説明文
簡易版1	1型糖尿病(インスリン依存)は糖尿病の一種である。主病態は1型糖尿病性持続的高血糖値状態であり、症状として多尿、頻尿、体重減少をもつ。
簡易版2	1型糖尿病(インスリン依存)は糖尿病の一種である。主病態は1型糖尿病性持続的高血糖値状態と呼ばれる持続的高血糖値状態である。症状は多尿と呼ばれる属性値関連異常、頻尿と呼ばれる属性値関連異常、体重減少と呼ばれる属性値関連異常である。
詳細版	1型糖尿病(インスリン依存)は代謝性疾患のうちの糖代謝異常を生じる疾患の一種である。1型糖尿病(インスリン依存)の主病態は1型糖尿病性持続的高血糖値状態(インスリン欠乏を原因とする、血液の血糖値の高い状態である持続的高血糖値状態)である。症状は多尿(属性値関連異常)、頻尿(属性値関連異常)、体重減少(属性値関連異常)である。

※表中の太字は臨床医学オントロジーから取り出した情報を表す。

4.2 テンプレートの組合せによる説明レベルの切り替え

前述した各テンプレートの組み合わせにより、本研究で開発したシステムでは、疾患への説明を想定した簡易版①および②、医師などを対象にしたより詳細な説明を想定した詳細版の3つのレベルを切り替えて説明文を生成することができる。これを用いて、利用者は必要や目的に応じた説明文の提供を受けることができる。

疾患概念の1段目のスロットは主病態、症状など、疾患の基本的な定義内容と考えられる。よって、これらの1段目のスロットのみを用いて生成した説明文の内容は患者むけの簡易な説明レベルで用いられる。ここで、1 段目の説明にクラス制約で参照している概念のドメイン依存性を考慮してより簡易な説明を行うテンプレート(2)-1'を利用したものを簡易版①、これを考慮しないテンプレート(2)-1 を利用したものを簡易版②とした。さらに、2 段目以降のスロットの定義内容も利用したテンプレート(3)を加えて説明文を生成したものを詳細版とした。表 7 に「1型糖尿病(インスリン依存)」を具体例として各レベルで出力した説明文を示す。

5. 臨床医学電子辞書の自動生成

4 章で述べたオントロジーに基づく自然言語説明文生成機能を用いて、臨床医学オントロジーから医学電子辞書を自動生成するシステムを試作した。ここではその概要を述べる。

5.1 医学辞書で説明される内容

臨床医学電子辞書のコンテンツを生成するのに先立ち、一般的に疾患を対象とした医学辞書として必要とされる説明内容を知るために、現在 WEB で公開されている疾患に関する医療知識を提供するサイトや既存の医療電子辞書から疾患についてどのような医用語を説明しているかを調査した。その結果、主な説明項目は疾患の主病態、原因、症状などであることが分かった。そのほかには、診断、治療、合併症などが説明されているサイトや辞書もあるが、現状の臨床医学オントロジーではそれらの内容が定義されていないため、本システムで説明文を生成す

表 8. 医療辞書で必要とされる疾患の説明項目と対応状況

疾患の説明項目	対応状況		対応に必要な拡張		
	オントロジー	説明文生成機能	オントロジー	テンプレート	追加機能の開発
疾患概要	×	×	—	○	○
主病態	○	○	/	/	/
原因	○	○	/	/	/
症状	○	○	/	/	/
徴候	×	×	○	○	—
危険因子	×	×	○	○	—
発生機序	×	×	—	○	○
合併症	×	×	○	○	—
病態生理	×	×	○	○	—
診断	×	×	○	○	—
治療	×	×	○	○	—
検査方法	×	×	○	○	—
予防	×	×	○	○	—
予後	×	×	○	○	○

ることはできない。ただし今後、臨床医学オントロジーを拡張すると共に、必要な文章生成テンプレートを追加すれば、これらの多くの説明文生成にも対応できると思われる(表 8)。さらに要約文生成や因果連鎖を用いた文章生成といった追加機能を開発することで、より複雑な説明項目についても対応が見込まれる。

## 5.2 臨床医学電子辞書の実装

図4に本研究で開発した、臨床医学オントロジーに基づく医学電子辞書の実行画面を示す。本システムは、法造で構築されたオントロジーを処理する Java API である HozoCore<sup>1</sup>を用いて開発された Web アプリケーションである。表示される説明文は、すべて臨床医学オントロジーから自動生成されたもので、3 章で述べた疾患説明文生成テンプレートの組み合わせにより、利用者の要求に応じて 3 つのレベルの説明文を切り替えることができる。また、疾患名や症状から疾患を検索する機能など、電子辞書としての基本的な機能は実装されている。本システムは専門家による Informal な評価によって、きわめて好評を得ており、今後の説明対象の拡張や機能改良により更なる有用性の向上が期待される。特に、臨床医学オントロジーの拡充に伴い、電子辞書としてのコンテンツが自動的に拡充される点が、本システムのきわめて有用な点と言える。

## 6. まとめと今後の課題

本研究では、オントロジーから自然言語説明文を生成する機能を開発し、それを用いて臨床医学オントロジーに基づく医学電子辞書の自動生成を試みた。本オントロジーは、医療の専門家とオントロジー研究者、情報科学の専門家などが議論に参加し、その基盤構造を定義したものである。さらに、今後のオントロジー拡充も継続して進められており、本システムが対象とする説明内容も充実する見通しを得ている。

今後の課題としては、オントロジーの拡充とあわせた説明文生成テンプレートの追加と、より良い説明文生成のため考察が必要とされる。現状では、疾患の説明を対象としたテンプレート

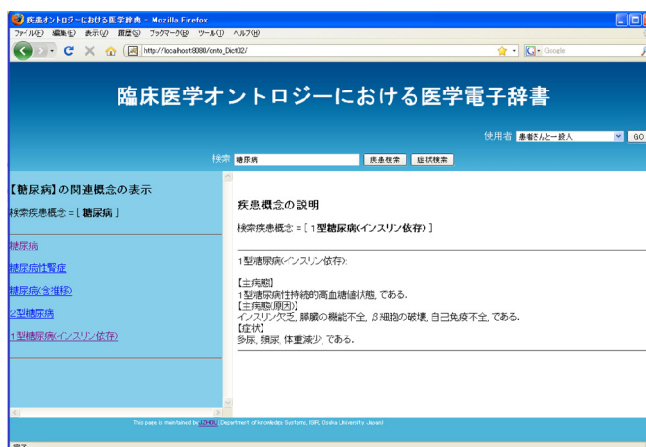


図 4. 臨床医学オントロジーに基づく医学電子辞書の実行画面

を留意しているが、その一部を改良することで人体構造など他の内容を説明するためのテンプレートの作成も進めている。また説明文生成の際に、患者など専門外の利用者にも理解しやすい言葉を用いた説明、例えば易しい言葉への変換などの対応が必要とされる。これ以外にも、他の応用システムでの利用を想定し、より柔軟性があり利用価値の高いシステムを目指した改良を進めていきたい。

## 謝辞

本研究は、厚生労働省、医療情報システム開発普及等委託研究費、「平成20年度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」の一環として行なわれた。

## 参考文献

- [大江 08]大江和彦: 医療知識基盤としての臨床医学オントロジー, セマンティック Web コンファレンス, 2008.
- [国府 08]国府, 周, 古崎, 今井, 大江, 溝口: 臨床医療オントロジーの構築に関する基礎的な考察, 第 22 回人工知能学会全国大会, 2E3-01, 2008.
- [国府 09]国府, 周, 太田, 今井, 大江, 古崎, 溝口: オントロジーに基づく医療知識ナビゲータ, 第 23 回人工知能学会全国大会, 2009.
- [Bontcheva 09] Natural Kalina Bontcheva and Brian Davis: Language Generation from Ontologies, Semantic Knowledge Management, pp.113-127, 2009
- [古崎 02]古崎, 来村, 池田, 溝口: 「ロール」おとび「関係」に関する基礎的考察に基づくオントロジー記述環境の開発, 人工知能学会論文誌, Vol.17, No.3, pp196-208, 2002.
- [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.: Yet Another Top-level Ontology : YATO, Proc. of InterOntology 2009, Keio Univ. Feb (2009)

<sup>1</sup> <http://ontsupport.enegate.jp/ontology/modules/> (法造によるオントロジー構築技術サポートサイト) にて公開されている。