

症例検索システムの試行運用に基づいた利用状況に関する基礎的検討

荒牧, 英治
奈良先端科学技術大学院大学研究推進機構

岩尾, 友秀
奈良先端科学技術大学院大学研究推進機構

若宮, 翔子
奈良先端科学技術大学院大学研究推進機構

伊藤, 薫
奈良先端科学技術大学院大学研究推進機構

他

<https://hdl.handle.net/2324/7183414>

出版情報 : Japan Journal of Medical Informatics. 38 (4), pp.245-256, 2018-10-30. Japan Association for Medical Informatics

バージョン :

権利関係 : © 2018 Japan Association for Medical Informatics

資料

症例検索システムの試行運用に基づいた利用状況に関する基礎的検討

荒牧 英治^{*1} 岩尾 友秀^{*1,2} 若宮 翔子^{*1} 伊藤 薫^{*1}
 矢野 憲^{*3} 大江 和彦^{*4}

診療文書は、後に分析することを想定して収集されていないため、二次利用が困難な場合が多い。このため、診療文書に基づく症例検索システムを試行的に臨床現場へ導入した報告例はあるものの、数年以上の長期間に渡り実運用した結果を報告した事例は少ない。本稿では、約7年にわたり運営してきた日本内科学会の症例検索システムの利用状況について、クエリ検索回数、傾向、経時的変化に関するログを分析した。分析結果から、運用年月を経るにつれて利用回数が増加しており、病名を中心とした検索ではコンディショニングや稀少疾患など様々なクエリに対して、一定数の検索結果を返す性能を備えていることが確認できた。複合名詞など文字数が多いクエリについては課題が残るものの、クエリのサジェスト機能や、様々な用途に応じてガイドできるようなインターフェースを備える等改良の余地があり、今後の発展が期待できる。

■キーワード：症例報告、症例検索システム、自然言語処理、データベース

A Fundamental Study on User Utilization Based on a Trial Operation of the Medical Case Retrieval System: Aramaki E^{*1}, Iwao T^{*1,2}, Wakamiya S^{*1}, Ito K^{*1}, Yano K^{*3}, Ohe K^{*4}

Since clinical documents are not supposed to be analyzed later, the secondary use of them is often difficult. For this reason, although there have been a few case retrieval system trials introducing it to clinical sites, there are few cases in which it has been practically used over a long period of several years or more. In this paper, we analyzed the frequency of query retrieval, the trends of query keywords and the chronological changes of them by analyzing the access logs of the case retrieval system of the Japanese Society of Internal Medicine, which was developed by us and had been operated for about seven years. As a result, we confirmed that the frequency of query increases as the operational year passes by, and the search function provided on disease names had the ability to return a certain number of the results for various search

^{*1}奈良先端科学技術大学院大学 研究推進機構
〒630-0192 生駒市高山町 8916-5

^{*2}京都大学大学院 情報学研究科

^{*3}東京工業大学 工学院情報通信系

^{*4}東京大学大学院 医学系研究科

E-mail : aramaki@is.naist.jp

受付日：2018年3月5日

採択日：2018年7月31日

^{*1}Nara Institute of Science and Technology, Institute for Research Initiatives

8916-5 Takayama-cho, Ikoma, Nara, 630-0192, Japan

^{*2}Graduate School of Informatics, Kyoto University

^{*3}Department of Information and Communications Engineering, Tokyo Institute of Technology

^{*4}Graduate School of Medicine, The University of Tokyo

queries about common and rare diseases. Although sometimes it does not work for queries with many characters such as compound nouns, we think that there is a room for further development by providing a query suggestion function and interface for suitable guide by assuming various patterns of the use cases.

Key words : Case report, Case retrieval system, NLP, Database

1. 緒 論

近年、人工知能が社会変革のトリガー技術として注目されている。とりわけ、少子高齢化や医師不足などの人的リソースに関する問題を抱える日本において、医用人工知能に対する期待は高い。しかし、医用人工知能の開発には、良質の診療データが大量に必要である。例えば、がんの画像診断を行うならば様々な種類の様々なステージの画像、検査値から患者の予後の推定ならば、両者がペアとなったデータといった具合に、想定するタスクに応じたデータを準備する必要がある。これらを研究開始時から前向きに構築することも可能であるが、すでに日常診療で集積されたデータを変換し、機械学習の教師データとすることも重要である。

一方で、電子カルテや退院サマリなどの多くの診療文書は、後に分析することを想定して収集されていないことに加え、後に分析する内容が多種多様かつ変幻自在であり、事前に分析に必要な情報を網羅、整理することが難しいため、二次利用が困難な場合が多い。したがって、分野を限定した基礎的検討¹⁾や、試行として臨床現場へ導入した報告^{2,3)}はあるもの、数年以上の長期間に渡る実運用について報告した事例は少ない状況である。本稿では、はじめに症例報告について述べたあと、「症例くん」について述べ、実運用にて使用された検索キーワードであるクエリ（以降、クエリと呼ぶ）について議論する。

2. 目 的

日本内科学会をはじめとする学会で報告される症例報告は500字程度であるもののフォーマットがほぼ統一されており、検索が比較的容易であ

る。また、その多くは診断が確定しているため、必要に応じて検索できれば臨床現場に情報を提供できると考えられている。このような背景から、東京大学と千葉大学の協力により「PINACO」（類似症例検索）と「症例くん」（クエリ検索）と呼ばれる2つの検索システムが開発され、日本内科学会で2011年から運用されてきた。前者の「PINACO」については既報があるが⁴⁾、後者の「症例くん」については、体系的なレビューがなされてこなかった。これは、入力クエリが1つのキーワードのみであり、情報が少ないことが主な理由であったが、数単語のクエリであっても、長期の実運用により得られる傾向は、貴重なフィードバックである。そこで本研究では、7年にわたり運用を続けてきた日本内科学会のWebベースの症例報告検索システムである「症例くん」の開発および利用状況に関して分析することで、新たな知見や課題点について言及することを目的とする。

3. 方 法

1) 症例報告の構成

図1に、症例検索システム「症例くん」で表示した症例報告の一例を示す。日本内科学会の症例報告は、①地方会（地方会名、回数）、②著者情報（著者名、著者所属、連絡先など）、③分類（例えば、循環器など）、④クエリ、⑤演題、⑥本文という項目から構成されている。

2) 症例報告の概要

日本内科学会では、北海道から九州まで全国10支部において年間35回の地方会が行われており、症例報告が発表されている。この際の症例報告は、演題抄録として電子化され、蓄積されている。収載数は2017年10月現在で4.5万件にの

家族性骨形成不全症に伴う弁膜症による心不全の1例

toshiyakojima-circ@umin.ac.jp

小島 敏弥, 藤生 克仁, 齋藤 幹, 山下 尋史, 平田 恭信, 永井 良三

東京大学循環器内科

【分類】 循環器

【キーワード】 心不全, 心臓弁膜症, 家族性骨形成不全症

【年齢・性別】 82 male

症例は82歳、男性。経過8年の慢性心不全があり、連合弁膜症、慢性心房細動を認めていた。また、青色強膜、外傷性骨折の反復、家族歴から臨床的に家族性骨形成不全症と考えられていた。2007年5月に呼吸困難にて入院。来院時、心駆出率は保たれていたが重度の僧帽弁閉鎖不全症を認めており、連合弁膜症、慢性心房細動を基礎疾患とする慢性心不全が呼吸器感染症により増悪したものと診断した。入院後、カルペリチド、フロセミド、アルブミン製剤投与にて心不全は一時改善した。しかし、その後原因不明の胸水増加を来し、呼吸不全にて永眠となった。病理解剖では胸水の原因として悪性疾患は否定的であり、肺実質には慢性心不全の変化は乏しく、急性の変化と考えられた。また僧帽弁のバルーン状の弁逸脱と腱索の延長を認め、組織学的には僧帽弁及び腱索にmyxoid changeが見られた。虚血性心疾患の所見は見られず、冠動脈の狭窄も有意ではなかった。家族性骨形成不全症による連合弁膜症とそれによる急性心不全、呼吸不全という病態が考えられ、非常に特徴的な症例であった。

図1 症例報告の一例

ばり、多くの専門領域をカバーしている。

3) 症例検索システム「症例くん」の特徴

本節では、症例検索システム「症例くん」の特徴について述べる。現在、日本内科学会では症例報告を検索可能な2つの検索システムが導入されている。1つは、千葉大学により開発された「PINACO (Patient Inquiry in Naika-gakkai Archives by COntext; ぴなこ)」である⁵⁾。もう1つは、東京大学により開発された「症例くん」である。

「PINACO」は類似症例検索システムであり、症例報告から名詞を抽出し、その頻度などを分析して類似度の高い順に症例を呈示する。これにより単にクエリを入れる以上に類似度の高い症例報告を検索することができる。

一方で「症例くん」は、クエリベースの全文検

索システムであり、短いクエリを含む症例を検索し、呈示する。それぞれの特徴について表1に示す。このように、両者は目的を異にしているため、相補的なシステムとして共存している。

4) 「症例くん」の構成

本節では、症例検索システム「症例くん」の構成、および、入力インターフェースについて述べる。症例検索システムの検索アルゴリズムは、基本的に症例報告の全文検索をベースとしており、それに対して同義語展開などの機能を加えた3つのモジュールから構成されている。それぞれについて下記で説明する。

■ UIモジュール: 検索インターフェース、サーバサイドのCGI (Common Gateway Interface) で構築されている。なお、開発言語にはPerlを

表1 「症例くん」と「PINACO」の比較

| | PINACO | 症例くん |
|----------|------------------------------|---------------------------------|
| 入力 | 症例文書 (1,600文字以内) | クエリ |
| 出力 | 類似症例 | 入力されたクエリを含む症例 (and 検索) |
| 検索アルゴリズム | TF/IDF 法 | 全文検索 (Geta-soc) |
| 形態素解析 | Mecab+独自辞書 | 形態素解析なし (事前に病名抽出した結果を使うオプションあり) |
| 入力の言い換え | なし (入力は文章を想定しており、言い換える必要がない) | 標準病名や独自辞書による言い換えあり |

表2 症例検索システムのサーバ構成

| | |
|-------------|---|
| Model | HP DL60 Gen9 E5 v4 4LFF SAS |
| OS | CentOS 6.8×64 |
| CPU | Xeon E5 2640 v4 2.40 GHz 1P/10C CPU |
| メインメモリ容量 | 64 GB |
| RAID コントローラ | HP Smart P440/4 GB FBWC |
| RAID 構成 | RAID5 |
| ストレージ | HP 450 GB (15krpm SC 3.5型 12G SAS HDD×4) |
| バックアップ媒体 | HDJ-U8.0 (8 TB) |

| | |
|------------------|--|
| 症状キーワード検索 | <input type="text" value="肺炎"/> アンドをスペースで連結してください。 (例) 「発熱 男性 心筋梗塞」 <input type="button" value="検索"/> |
| 患者オプション | <input type="text" value=""/> 歳 <input type="radio"/> 以上 <input type="radio"/> 未満 <input checked="" type="radio"/> 年齢区別なし <input type="radio"/> 男性 <input type="radio"/> 女性 <input checked="" type="radio"/> 性別区別なし |
| 言語処理オプション | <input type="radio"/> 辞書なし <input checked="" type="radio"/> 同義語言い換え (万病辞書) <input checked="" type="radio"/> 文字列検索 <input type="radio"/> 病名限定検索 |

図2 「症例くん」のインターフェース

用いた。サーバサイトのサーバ構成を、表2に示す。また、図2に、症例検索システム「症例くん」の検索インターフェースを示す。症例に関するクエリを入力することで、クエリに合致した症例を表示する。下記に、各オプション機能について説明する。

■**検索モジュール**：全文検索部分は、汎用連想計算エンジン (GETA: Generic Engine for Transposable Association)⁶⁾ を用いている。GETAは、文書データを対象として、文章間や単語間の類似度を内積によって高速計算するツールである。図2で示したように、患者オプションや言語処理オプションの項目を備えており、それぞれ年齢、性別、そして後述する同義語言い換えによる検索を実現している。

■**同義語展開モジュール**：同義語展開部分については、標準病名マスター⁷⁾の索引語と登録した語をベースに同義語を展開する。標準病名マスターとは、広く疫学研究やそれを応用した医学研究などに用いられている病名体系であり、およそ25,000の病名からなる。表3に、同義語言い換えの事例を示す。例えば、クエリに「痛み」を入力した場合、同義語言い換えには標準病名「疼痛」を含んだ5個の文字列に言い換えられる。クエリ「血尿」では、標準病名「血尿」に加えて、検査値情報「尿潜血 (3+)」にも言い換えている。このように、標準病名に近い症状や病名を言い換え候補として対応させることで、症例報告で類似症例を広く検索することができる。なお、同義語言い換え機能には、機械学習を用いた病名抽出機

表3 クエリと同義語言い換え例

| クエリ | 同義語言い換え例 |
|--------|--|
| 痛み | 疼痛, 痛み, 激痛, 体動時痛, 電撃痛 |
| 糖尿病 | 糖尿病, 血糖コントロール不良 |
| 血尿 | 血尿, 尿潜血 (3+) |
| 腫瘍 | 腫瘍, 肉芽腫病変 |
| 悪性腫瘍 | 悪性腫瘍, 悪性軟部腫瘍, 腫瘍塞栓症 状, low grade malignancy, 低悪性度 腫瘍 |
| 肺炎 | 肺炎, 肺臓炎, 市中感染 |
| バージャー病 | 閉塞性血栓血管炎, BUERGER 病 ビュルガー病, 血栓閉塞性動脈炎 |
| 狭心症 | 虚血性胸痛 |
| NASH | 非アルコール性脂肪性肝炎, 非アル コール性脂肪肝炎 |

能⁸⁾と病名に関する辞書リソースを用いている。万病辞書は、医療従事者が記載した電子カルテや退院サマリから、症状や病名に関連する語を広く抽出した医療言語処理に適した辞書リソースであり、およそ13万語句の語彙を備えている。辞書リソースを用いる構造とすることで、拡張性の高い検索エンジンを実現している。これら病名抽出機能を備えたソフトウェア⁹⁾、辞書リソース¹⁰⁾を奈良先端科学技術大学院大学が一般公開している。

5) 症例報告の前処理：性別と年齢

「症例くん」を用いるためには、年齢や性別など構造化されていない情報を抽出する必要がある。これは症例本文中の「症例は82歳、男性」といった表現から抽出する。性別に関しては、男性、女性といった単語で抽出を行う。年齢の抽出に関しては、「歳」と「才」の直前に出現する数字を抽出する。複数の年齢表現が得られた場合には、複数の症例を扱う症例シリーズとして、「その他」に分類する。

4. 結果

前章で述べた症例検索システムは、日本内科学会のウェブサイトでは2011年度から試行運用されている。本章ではその利用実態から得られた結果について報告する。はじめに、図3、図4に、地方会、および診療科ごとの報告数をそれぞれ示す。また、図5に年度ごとの報告数の変遷を示す。

次に、検索回数の変遷について図6に示す。時期によって大きな変動があるが、運用期間を経るにつれて総じて検索回数が増加する傾向が見られ、多い時期には月に1,200アクセスを記録している。次に、時間帯別の検索回数を図7に示す。なお、検索回数に関しては2011年から2017年までの平均値を用いた。図8に検索時のクエリ数

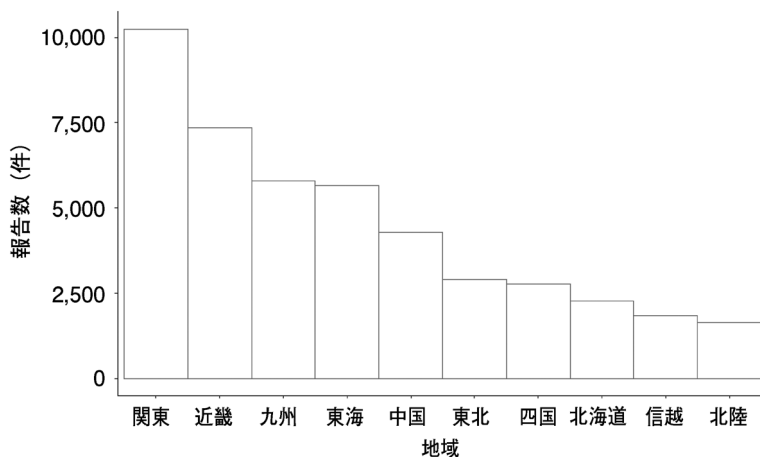


図3 地方会ごとの報告数

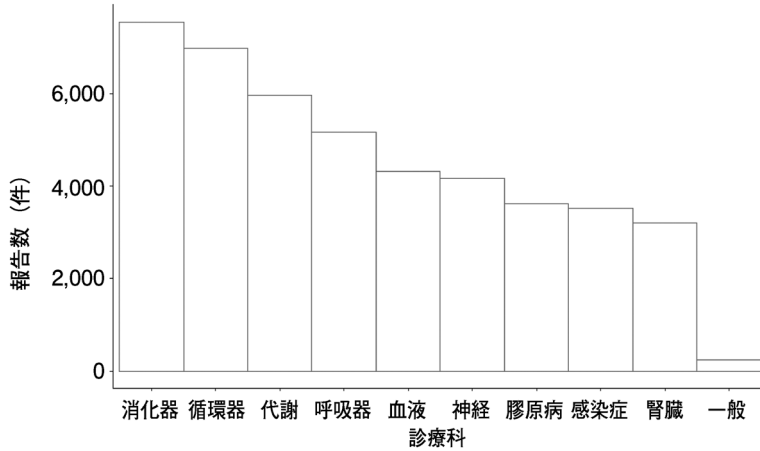


図4 診療科ごとの報告数

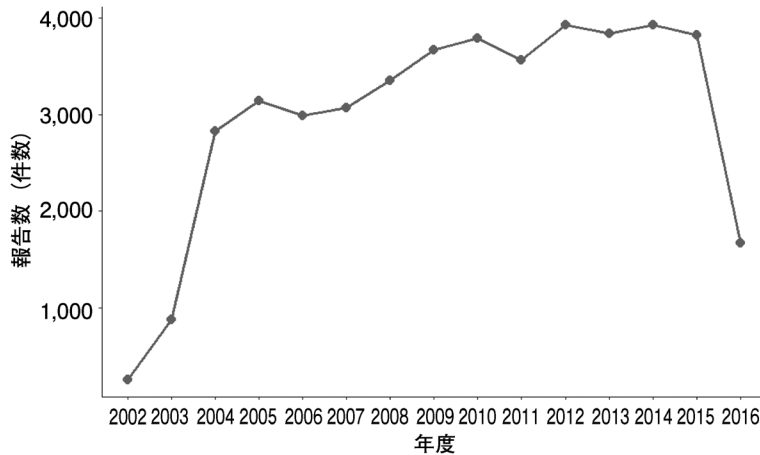


図5 年度ごとの報告数

の統計を示す。ここでいうクエリ数は、入力されたクエリを構成する単語数である。例えば、クエリが「心筋梗塞 and 若年」の組み合わせの場合は2となる。結果的に、クエリ数1が多くの割合を占めた。

最後に、クエリに関して表4に2011年度から2014年度、表5に2015年度から2017年度に渡るクエリの上位30件の変遷を示す。また、これらクエリを病名、症状、診療行為の3種類のカテゴリに分類し、表中の各クエリの先頭にカテゴリごとにマーキングした。ただし、各クエリに異なる

カテゴリが含まれているものや、分類が難しいクエリに関してはマーキングしていない。なお、この分類作業は医療情報技師免許を有する著者1名の主観により手作業で実施した。

5. 考察

本章では、「症例くん」の利用状況を解析することで得られた結果について考察する。

1) 症例報告の傾向

図3から、大都市に存在する地方会ほど報告数が多い結果となり、これは医療機関の数や規模が

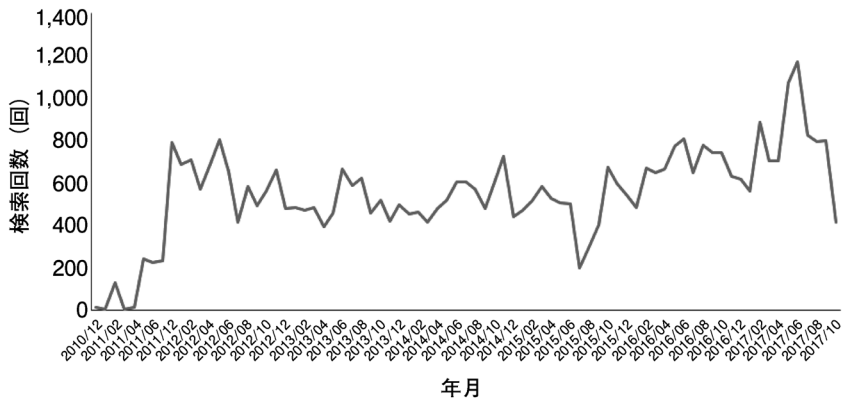


図6 検索回数と年月

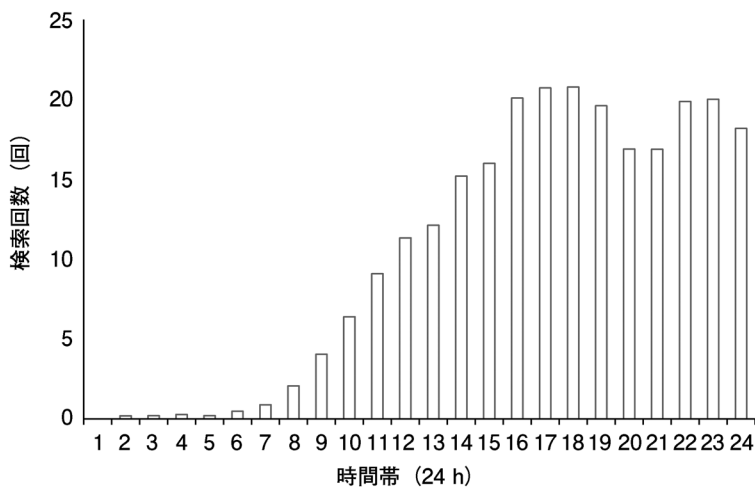


図7 検索回数と時間帯

反映されていると考えられる。また図5から、年度ごとの報告数は、運用期間を経るごとに着実に増加傾向であることが見て取れる。

2) 検索回数の変化 (時期と時間)

図6で示した検索回数の変遷から、学会会員に限定されたサービスとしては頻りに利用されている状況が分かる。また、図7で示した結果から、1日の時間帯の中で16時から19時の夕方、22時から23時前後の深夜の時間帯にピークがあり、診療時間外に利用されていることが分かる。その一方で、日中での検索も少なくなく、診療時間の合間にも利用されている状況が伺える。

3) クエリ数

図8から、クエリ数3以内が全体のおよそ90%を占めており、利用者は比較的少ないクエリ数で検索を行うケースが多いことが分かる。クエリ数4以上についても散見されるが、これは文をそのまま検索するなど類似症例の検索を試みたと推測される。

4) クエリ傾向

表4、表5で示した結果から、利用者は病名そのもので検索するケースが多いことが分かる。特に「糖尿病」、「心筋梗塞」、「肺炎」、「心不全」といったコモンディーズが各年度の上位に多く出

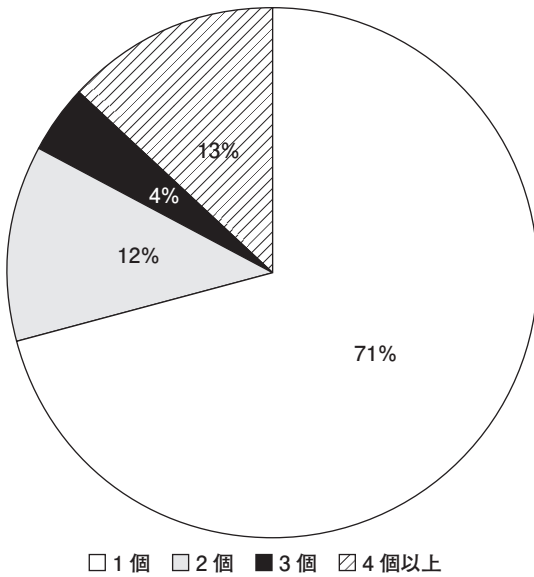


図8 検索回数とクエリ数

現している。これらのクエリを「症例くん」で検索してみたところ、ヒット数はそれぞれ以下のように膨大にあり、さらに絞り込まれていくと考えられる。

- 糖尿病 (6,915 件)
- 心筋梗塞 (1,493 件)
- 肺炎 (5,110 件)

この結果から、症例報告においても、珍しい病名ばかりでなく、コモンディジーズについてもシステムが情報提供に寄与している実態が伺える。

その一方で、コモンディジーズ以外にも「後天性血友病」、「IgG4」等の難治性の病名も上位に出現している。これらについても、以下のように100件以上の症例がヒットし、治療方針の決定などに貢献している可能性がある。

- 「後天性血友病」(709 件)
- 「IgG4」(156 件)

病名以外のクエリとしては、「肝臓」(2011年度)といった部位がある。これらが少ないのは、システム名の「症例くん」が病名のみを扱っていることを連想させている可能性があり、今後の検討を要する。

また、「歩行時のふらつき」(2012年度)のように「AのB」型の名詞句での検索も出現していた。このクエリにマッチする症例は139件あるが、これを分割した検索ヒット症例数は、同じく「症例くん」を用いた検索結果では以下のように多く見られた。

- 「歩行時」(380 件)
- 「ふらつき」(839 件)
- 「歩行 and ふらつき」(287 件)

しかしながら、このケースでは本来検索しなかった症例がマッチしなかった可能性があり、「AのB」タイプのクエリについては、ばらした結果を参考として表示するなど、今後の対応が必要である。このように、特定の疾患だけでなく様々な病名に関して検索されていることから、症例検索システムとしての利用範囲は比較的広いと思われる。

5) クエリの経時的変化

運用期間を経るにつれて症状や病名が主なクエリとして使われるようになっていく。徐々に病名を中心とした検索システムとしての使用方法が固定化されてきたことが伺える。「症例くん」は病名のみの特化したシステムではないため、本当に医薬品名に関する検索要求がないのか調査が必要である。

また、2013年度からはクエリとして「冠動脈解離 and 女性」といった病名と患者の特性を組み合わせたものや、「乳酸アシドーシス and アルコール」のような病名とその原因の組み合わせ等が見られるようになった。さらに2015年度からは、「アミロイド and トシリズマブ」などの病名と治療法の組み合わせ等が上位に入っている。このように、病名だけでなくそれと関連する用語を組み合わせることで、有用な症例を探る検索方法が模索されてきたことが見て取れる。その一方で、「むくみ」「低血圧」などの症状に関する検索は少ない。

全期間を通じてクエリの大半を病名が占めていることがわかる。一方で、症状と診療行為に関しては、全期間を通じて使用頻度が少なく、年度ご

表4 クエリ上位30件(2011年度~2014年度)

(*表中 ■:病名, ◆:症状, ▲:診療行為)

| 順位 | 2011年度 | 件数 | 2012年度 | 件数 | 2013年度 | 件数 | 2014年度 | 件数 |
|----|-------------------------|----|--------------------|----|-----------------|----|--------------|----|
| 1 | 肝臓 | 82 | ◆歩行時のふらつき | 79 | ■大動脈炎症候群 PET | 46 | ■糖尿病 透析 | 62 |
| 2 | ■心不全 | 41 | ■ミトコンドリア糖尿病 | 76 | ■IgG4 | 40 | 内科 | 53 |
| 3 | ■インスリンノーマ HbA1c | 41 | ■本態性高血圧症 | 66 | ■血小板減少 胸腹水 | 38 | ■日本紅斑熱 | 45 |
| 4 | ■消化管アミロイドーシス | 37 | ■後天性血友病 | 43 | ■原発性免疫不全症候群 | 31 | ■巨細胞性動脈炎 | 25 |
| 5 | 関東 | 37 | ■レジオネラ肺炎 | 34 | ■血球貪食症候群 | 29 | ■若年 脳梗塞 | 21 |
| 6 | ■グリコアルブミン | 32 | ■後腹膜線維症 | 33 | ■アルコール性ケトアシドーシス | 25 | ■肺炎 | 21 |
| 7 | インスリン 低血糖 | 30 | ■反応性関節炎 | 32 | ■siadh | 20 | ■糖尿病 | 21 |
| 8 | ■反応性低血糖 | 23 | ■低アルブミン血症 | 31 | ■心破裂 | 19 | ■下垂体機能低下症 | 20 |
| 9 | ■RVOT | 22 | ▲ロキソニン | 31 | ■ネフローゼ症候群 感染 | 19 | ■一酸化炭素中毒 | 20 |
| 10 | ■原発性アルドステロン症 | 21 | ■高血圧性脳症 | 28 | ■PN | 19 | ■鉄欠乏性貧血 | 18 |
| 11 | 妊娠 ステロイド | 18 | ■アレルギー性肉芽腫性血管炎 | 27 | ■関節リウマチ 悪性リンパ腫 | 17 | ■溶血性貧血 | 17 |
| 12 | ■肺高血圧症 | 18 | ■筋サルコイドーシス | 27 | ■髄外 骨髄腫 | 17 | ■気腫性膀胱炎 | 16 |
| 13 | ■インスリン抗体 GA | 15 | ■顎骨壊死 | 26 | ■サルコイドーシス 高Ca | 17 | ■蛋白漏出性胃腸症 | 16 |
| 14 | ■ホジキン | 14 | ■上腸間膜静脈血栓症 | 24 | ■腎アミロイドーシス | 16 | ◆めまい | 16 |
| 15 | ■SLE | 14 | ■感染性心内膜炎 | 24 | ■側頭動脈炎 | 16 | ■SAPHO 症候群 | 15 |
| 16 | ▲トシリズマブ | 13 | ■高血圧 | 22 | ■冠動脈解離 女性 | 15 | ■百日咳 | 14 |
| 17 | ■慢性リンパ性白血病 | 12 | ■SLE+ 血小板減少 | 22 | ■上腸間膜動脈閉塞 | 15 | ■急性散在性脳脊髄炎 | 14 |
| 18 | ■コレステロール塞栓症 | 12 | ■低血糖脳症 | 21 | ■乳酸アシドーシス アルコール | 15 | ■日本脳炎 | 14 |
| 19 | ■TSH レセプター抗体 +バセドウ病 | 12 | ■粘液水腫 | 21 | ■非閉塞性腸管虚血症 | 15 | ■側頭動脈炎 | 13 |
| 20 | ◆フェリチン | 12 | ***** | 21 | ■ACTH 単独欠損症 | 14 | ■安定狭心症 | 13 |
| 21 | ■咽頭炎 | 12 | ■アミロイドーシス ネフローゼ症候群 | 20 | ■脾腫 | 13 | ■レジオネラ肺炎 | 13 |
| 22 | kannzou+ | 10 | ■アルコール性ケトアシドーシス | 19 | ■肺癌 脳 転移 | 13 | ■急性前立腺炎 | 13 |
| 23 | ■アミロイドーシス | 10 | ■IgG4 | 18 | ■皮疹 発熱 胸水 | 13 | ■慢性下痢 | 13 |
| 24 | ■原発性硬化性胆管炎 | 9 | 高血圧 妊娠 | 17 | アルコール 呼吸 | 13 | 低血糖 高齢者 | 13 |
| 25 | ■自己免疫性溶血性貧血 +抗リン脂質抗体症候群 | 9 | ■劇症1型糖尿病 | 17 | ■脂肪織炎 | 13 | ■後天性血友病 | 13 |
| 26 | ■異所性 副甲状腺機能亢進症 | 8 | ■肺炎 | 17 | ■本態性高血圧症 | 13 | ■ Dengue 出血熱 | 12 |
| 27 | ■異常ヘモグロビン | 8 | ▲サラソスルファピリジン | 17 | ■潜在性甲状腺機能 | 13 | ■心タンポナーデ | 12 |
| 28 | ▲ゲムシタピン | 8 | ■発作 褐色細胞腫 | 17 | ■偽性副甲状腺機能低下症 | 13 | ■たこつぼ心筋症 | 12 |
| 29 | ***** | 8 | ■慢性骨髄性白血病 | 17 | ■後腹膜線維症 | 12 | ■感染性心内膜炎 | 12 |
| 30 | ■肺門リンパ節腫脹 | 7 | ■溶血性尿毒症症候群 | 17 | ■後天性血友病 | 12 | ■たこつぼ | 12 |

「*****」と記載した箇所は、個人名が記載されていたため、公表するには不適であることからマスクしている項目

表5 クエリ上位30件(2015年度~2017年度)

(*表中 ■:病名, ◆:症状, ▲:診療行為)

| 順位 | 2015年度 | 件数 | 2016年度 | 件数 | 2017年度 | 件数 |
|----|---------------|----|----------------|-----|----------------|----|
| 1 | ■心内膜炎 | 52 | ■心内膜炎 | 110 | ■心筋梗塞 | 44 |
| 2 | ■低血圧 | 46 | ■A型肝炎 | 40 | ▲クリオフィルトレーション | 24 |
| 3 | ■胸鎖関節炎 | 37 | ■糖尿病 | 37 | ■帯状疱疹 | 22 |
| 4 | ■細菌性髄膜炎 | 28 | ■心筋梗塞 | 32 | ■糖尿病 | 21 |
| 5 | ■インスリン受容体異常症 | 23 | ■たこつぼ型心筋症 | 31 | ■ST上昇型心筋梗塞 | 21 |
| 6 | ■レジオネラ肺炎 | 22 | ■Tafro | 25 | ■IgG4 | 21 |
| 7 | ■アナフィラキシー | 21 | ■後天性血友病 | 25 | ■悪性リンパ腫 | 20 |
| 8 | ■横紋筋融解症 | 20 | ■急速進行性糸球体腎炎 | 24 | ■IgA血管炎 | 19 |
| 9 | ■自己免疫性肝炎 | 19 | 狭心症 心筋梗塞+喫煙 | 23 | ■孤立性上腸間膜動脈解離 | 19 |
| 10 | ■たこつぼ型心筋症 | 18 | ■血球貪食症候群 | 21 | リウマチ性多発筋痛症 透析 | 18 |
| 11 | ■脊髄梗塞 | 18 | ■急性胆管炎 | 20 | ■多系統萎縮症 | 18 |
| 12 | ■多発性骨髄腫 | 17 | ■SLE | 19 | ■急性骨髄性白血病 | 17 |
| 13 | ■インフルエンザ | 16 | ■抗リン脂質抗体症候群 | 19 | ■狭心症 | 17 |
| 14 | 血液培養 陽性 | 16 | ■つつが虫病 | 18 | ■川崎病 | 17 |
| 15 | ■皮膚筋炎 | 15 | ■アルコール性心筋症 | 18 | ■周期性四肢麻痺 | 17 |
| 16 | ■シェーグレン症候群 | 15 | ■家族性地中海熱 | 18 | ■糖尿病性ケトアシドーシス | 16 |
| 17 | ■周期性四肢麻痺 | 14 | ■E型肝炎 | 18 | ■溶連菌感染後糸球体腎炎 | 16 |
| 18 | ■腹水 細菌性 | 14 | ■急性腎障害 | 17 | IVCフィルター+出血 | 16 |
| 19 | ■上腸間膜動脈閉塞症 | 14 | ■睡眠時無呼吸症候群 | 17 | ■化膿性脊椎炎 | 16 |
| 20 | ■血球貪食症候群 | 13 | ■好酸球性胃腸炎 | 17 | ■脳梗塞 | 15 |
| 21 | ▲アミロイド トシリズマブ | 13 | ■肺炎 | 16 | IVCフィルター+後腹膜出血 | 15 |
| 22 | ■肺胞出血 | 13 | ■肺結核 | 15 | ***** | 15 |
| 23 | ■低カリウム血症 | 13 | ■ACTH単独欠損症 | 15 | ▲イオン交換樹脂 | 15 |
| 24 | ■肺動静脈瘻 | 13 | ***** | 15 | ■後天性血友病 | 15 |
| 25 | ■低音障害型感音難聴 | 13 | ■家族性高コレステロール血症 | 14 | ■リウマチ性多発筋痛症 | 15 |
| 26 | ■結核性腹膜炎 | 12 | ***** | 14 | ■IgG4関連下垂体炎 | 14 |
| 27 | ■ベーチェット病 | 12 | ■橋本病 | 14 | ■副腎不全+DIC | 14 |
| 28 | アシクロビル 腎不全 | 12 | ◆片頭痛 | 13 | ■冠攣縮 心室細動 低体温 | 14 |
| 29 | ■低血糖 | 12 | ■Churg 喘息 | 13 | ■歯周病+糖尿病 | 14 |
| 30 | ■クモ膜下出血 | 12 | ■血管内リンパ腫 | 13 | ■家族性地中海熱 | 14 |

「*****」と記載した箇所は、個人名が記載されていたため、公表するには不適であることからマスクしている項目

とにこれといった特徴的な傾向は見られなかった。

前述したように、クエリの変遷から、本システムは基本的に病名をクエリとして用い、場合によってその他の原因、患者の特徴、診療行為等を組み合わせるといった使用方法がほぼ定着しつつあることが見て取れる。しかし、本システムの開発当初は、上記の用い方以外にも以下のような使用シナリオを想定していた。

■診断支援シナリオ：患者の症状から想定した診断はあるものの、見落とし防止のために、症状から、あり得る他の診断の可能性を確認する。

■副作用検索シナリオ：医薬品の副作用を検索し、起こりうる副作用かどうかを確認する。

上記のようなシナリオでの利用のされ方は、現在のクエリログからは推測されず、インターフェース改善などガイドも必要であると思われる。

6. 結論

本研究では、症例報告検索システム「症例くん」のシステムに関する詳細と、約7年間に渡る利用状況について報告した。一般のWeb検索システムと同様のクエリのための検索、簡素なインターフェース、症例報告に特化した年齢、性別などの特徴の利用により実用性を高めたことにより、学会員限定のサービスとしては頻回に利用され、現在も利用が増加している。

現在の利用状況は、病名を中心とした検索に終了しているものの、コモンディジーズや稀少疾患など様々な病名を主とするクエリに対して、一定数の検索結果を返すことに成功している。一方で、複合名詞などの長いクエリに対しては、本来の対応する検索結果のごく一部しか返せていないケースや、症状や医薬品などをクエリに用いた検索は利用者に周知されていない可能性がある。本来の症例報告の利用可能性を考えれば、さらに発展の余地があり、今後は適切なクエリのサジェスト機能、副作用検索、診断支援など様々な用途の利用を補助するインターフェースの設計などが課題で

ある。

利益相反

本研究の実施にあたっては、全共著者に関して「利益相反」は存在しない。

謝辞

本研究は、主に国立研究開発法人日本医療研究開発機構 (AMED)、「人工知能による総合診療診断支援システムの開発」(代表：永井良三，科研費番号：JP171k1010019)の助成によって実施された。

協力いただいた日本内科学会に感謝いたします。

参考文献

- 1) 斎藤恵一，納富一宏，藤本哲男. 自己組織化マップと自然言語処理を用いた症例検索に関する基礎的検討. バイオメディカル・ファジィ・システム学会大会講演論文集 2000；13, 0：57-58.
- 2) 岡本和也，竹村匡正，黒田知宏，長瀬啓介，吉原博幸. 文脈に基づく類似診療文書検索システム. 生体医工学日本エム・イー学会誌 2006；44, 1：199-206.
- 3) 岡本和也. 京都大学医学部附属病院における自然言語処理の試み. Seagaia Meeting in Tokyo, 2013.
[http://www.seagaia.org/~sg2013/_src/sc773/okamoto.pdf (cited 2018-May-1)].
- 4) 日本内科学会事務局. 地方会演題検索システムのご案内. 一般社団法人日本内科学会, 2005.
[http://www.naika.or.jp/info/info_kensaku/ (cited 2018-May-1)].
- 5) 土井俊祐，木村 隆，関根正樹，鈴木隆弘，高林克己，田村俊世. 学会ホームページにおける類似症例検索システムの実運用と評価. 生体医工学 2011；49, 6：870-876.
- 6) 日立製作所，国立情報学研究所，東京工業大学，北陸先端科学技術大学院大学，国文学研究資料館. 汎用連想計算エンジン (GETA). 情報処理振興事業協会, 2007.
[<http://geta.ex.nii.ac.jp/geta.html> (cited 2018-May-1)].
- 7) 医療情報利活用推進部門標準病名マスター担当. ICD10 対応標準マスター. 医療情報システム開発センター, 2018.

- [<http://www2.medis.or.jp/stdcd/byomei/index.html> (cited 2018-May-1)].
- 8) Aramaki E, Yano K, Wakamiya S. MedEx/J. A One-scan Simple and Fast NLP tool for Japanese Clinical Texts, Studies in Health Technology and Informatics. *MEDINFO eHealth-enabled Health* 2017 ; **217** : 285-288.
- 9) 荒牧英治. 日本語病名抽出システム MedEx/J. 奈良先端科学技術大学院大学ソーシャル・コンピューティング研究室, 2017. [<http://mednlp.jp/MEDEX/index.html> (cited 2018-May-1)].
- 10) 荒牧英治, 若宮翔子. 万病辞書. 奈良先端科学技術大学院大学ソーシャル・コンピューティング研究室, 2017. [<http://mednlp.jp/DIC/index.html> (cited 2018-May-1)].
-