

[事例実践論文]

健康危機管理支援ライブラリー(H-CRISIS)の ユーザビリティ改善に向けた情報検索機能強化の取組み An Effort to Enhance Information Retrieval Functions to Improve Usability of H-CRISIS

小森 賢一郎[†], 冨尾 淳[†]
Kenichiro KOMORI, Jun TOMIO

[†] 国立保健医療科学院 健康危機管理研究部

[†] Department of Health Crisis Management, National Institute of Public Health

要旨

国立保健医療科学院が運営する健康危機管理支援ライブラリー(H-CRISIS)は、健康危機管理情報のリポジトリおよび提供を目的とするウェブサイトである。筆者らは、記事へのアクセス性向上につながる検索機能の実装を目的とし、保健医療分野のウェブサイトにある検索機能を調査した。その中で H-CRISIS 未搭載機能として「絞り込み検索」が確認されたため、当該機能を実装した。また、タブレット端末など利用時の操作性を考慮し、タグ(キーワードなどが記されたボタン)を活用した機能(「検索語推薦機能」)も実装した。タグとするキーワードは、H-CRISIS 中のあるカテゴリに属する記事全体を形態素解析し、一定回数以上出現する形態素を採用した。機能の追加により、目的記事の検索が容易になり、同時に記事数把握が可能になった。

Abstract

Health Crisis and Risk Information Supporting Internet System (H-CRISIS), provided by the National Institute of Public Health, is a website that aims to store and provide information on health risk management. In order to implement a search function that would lead to better accessibility of these articles, we analyzed the search functions of related sites provided by databases and websites in the health and medical fields. Among the functions provided by each site, "filtering by metadata" was identified as a function not yet provided in H-CRISIS, and therefore "filtering by metadata" was added to H-CRISIS. In addition, a search function using tags (buttons with keywords, etc.) ("keyword suggestion search") was implemented in order to facilitate the use of smartphones and tablet-type terminals. The keywords used for the tags were determined using morphological analysis. Based on the above studies, we implemented a narrowed search and a tag search on the site.

1. はじめに

国立保健医療科学院が提供する健康危機管理支援ライブラリーシステム(以下 H-CRISIS) ウェブサイト[1]では COVID-19 の流行以降、公開記事数が格段に増加し 2024 年現在サイト上では 3 万件を超える公開記事がある。H-CRISIS が情報を大量に保持し記事が日々追加される中で、ウェブサイトのユーザビリティ、すなわち使いやすさを保つ必要がある。H-CRISIS ではユーザビリティ向上のため、記事を探し出すための新たな機能を、ウェブサイトの一部領域にて試験的に実装し、検索機能を強化した。本稿では、H-CRISIS の特徴を踏まえた新たな記事検索機能を試験的に導入する背景と目的、導入する機能の検討、実装結果について述べる。

2. 背景と目的

2.1. 健康危機管理と DHEAT について

健康危機管理とは、「医薬品、食中毒、感染症、飲料水その他何らかの原因により生じる国民の生命及び健康の安全を脅かす事態に対して行われる健康被害の発生予防、拡大防止、治療等に関する業務であって、厚生労働省の所管に属するもの」[2]などを指し、防災についても同指針において「災害に起因する健康危機については、厚生労働省防災業務計画に沿った総合的かつ計画的な対策の推進に努める」[2]として、範疇の一つとして取り扱われる。具体的な取り組みとしては、感染症などの拡大防止に係る活動や大規模災害時における地域住人の健康維持に係る活動、これらの活動を円滑に実施する

[事例実践論文]

2024 年 7 月 5 日受付, 2024 年 11 月 29 日改訂, 2025 年 1 月 17 日受理

© 情報システム学会

ための平時における体制構築に関する活動などが挙げられる。特に大規模災害時には、被災都道府県及び保健所において指揮調整機能などの業務を補助するために、災害時健康危機管理支援チーム(以下 DHEAT) が自治体単位で組織し派遣されており、2024 年 1 月に発生した能登半島地震では日本全国から DHEAT が派遣され様々な活動に従事した。

2.2. H-CRISIS の概要と本稿の目的

H-CRISIS は、健康危機管理に関する情報を提供するウェブサイトであり、厚生労働省や地方公共団体などが発信した健康危機管理関係情報、健康危機事例のデータベース、バイオテロ関連疾患の臨床診断や検査方法の情報、CBRNE テロ(化学,生物学,放射線,核兵器,爆発などを用いたテロ)発生時における医療従事者・現場対応者向けの情報を主に提供している。同サイトは 2002 年より国立保健医療科学院にて運営されており、健康危機管理に関するデジタルアーカイブとしての機能を有している。想定する利用者の一例が先ほど挙げた DHEAT であり、「災害時健康危機管理支援チーム活動要領の改正(DHEAT に係る協議会の設置及び保健所現状報告システム等の運用) について」(健健発 0328 第 2 号)においても、国立保健医療科学院の役割として「DHEAT の活動に係る必要な情報の提供」をするため H-CRISIS の運用と管理を行う、とされている。加えて、CBRNE テロが発生した際の対応が求められる職員や、地域保健医療に携わる保健所や行政機関の職員といった、業務として健康危機管理に取り組む方も、利用者として想定している。一例として、ある地域で共通する症状を訴える人々が多数発生した際に、発生原因の仮説を立てる記述疫学の過程において、同業務にあたる保健所など機関の担当者が H-CRISIS を活用して過去の健康危機対応事例などを参照することを想定している。これらの専門職だけでなく、健康危機管理に関する取り組みを知りたい研究者や純粋に興味のある方など、健康危機管理に関する様々な利用者が存在するものとしてサイトは運営されている。また、H-CRISIS は健康危機管理に関する情報を系統的に収集し記録する点で、リポジトリとしての機能を有する。

目的とする情報へ到達するための機能として、H-CRISIS にはカテゴリに登録されている記事一覧を表示する機能(カテゴリ記事一覧検索)や、アルファベット社が提供するサービスである google カスタム検索エンジンを活用したサイト内検索機能がある。しかしこれら既存の検索機能では、例えば「医療安全」に関するカテゴリの記事は 2024 年現在 2,500 件を超えているが、ある単語に関して検索する際に「医療安全」カテゴリ内記事に絞って探す、というような探し方はできない。ユーザビリティは、日本産業規格(JIS)で「特定のユーザが特定の利用状況において、システム、製品又はサービスを利用する際に、効果、効率及び満足を伴って特定の目標を達成する度合い」と解釈・定義[3]されている。この定義を踏まえると上記検索機能のみ提供する H-CRISIS は、カテゴリに含まれる数千件の記事、もしくはある単語に関する数百件の記事から、目的とする情報を探す行為は、効率や効果を伴って目標を達成しやすいとは必ずしも言えず、ウェブサイトのユーザビリティが高いとは言えない。ユーザビリティが低いと、想定利用者が目的とする健康危機管理情報に到達することが困難、もしくは途中で諦めることにつながるため、検索機能の改善によりユーザビリティを高める必要がある。また、菊池ら[4]が提唱した検索に対する要求段階の中で、現在の検索機能は初期段階にあたる「曖昧性を持った汎用目的の機能」に回答するものでしかない。今後、記述疫学などの一環として機能要求があった際に、同要求段階における「データ記述で利用される語彙や概念構造を反映したメタデータによる精緻な問い合わせ処理」という第二段階、さらには第三段階の要求も考えられる。

本稿では、H-CRISIS のユーザビリティを高めるために、既存の検索機能とは異なる記事を探し出す方法を検討し、実際にウェブサイトでの機能提供を行った過程を示す。

3. 機能の検討

ここでは、具体的に実装すべき機能について言及する。なお、今回は構造化が行われている健康被害危機管理事例データベース(以下、事例 DB)に絞って、試験的に検索機能を実装する。後述の想定する検索機能実装にはデータの構造化が必要であるが、H-CRISIS のデータは部分的にしか構造化されていないため、本稿ではデータの構造化が行われている事例 DB のみを対象としたためである。

3.1. 事例 DB の課題と実装する機能の検討

事例 DB は、全国の地方衛生研究所から提供された健康危機対応事例に関する情報の提供を主な目的とする。2024 年現在 1,535 件の記事が報告書形式で公開されているが、既存の検索機能ではサイト内個別領域を検索する精度が高いとは言えない。例えば、事例 DB 上で腸管出血性大腸菌(O-157)による死亡事例を既存検索機能から絞りこもうとする場合、検索式「site:http://~」によって検索領域を絞ったとし

ても、死亡事例を分類するための適切な検索式が無い場合、検索結果からリンクを辿って記事を個々に閲覧の上判断する必要がある。また、「ウイルス」のような言及頻度が高い事象を検索すると、該当記事が非常に多い上に、個別記事の閲覧による判断が必要となる。これらはサイト内検索機能の仕様に由来するものであるため、既存の検索機能に依存しない、独自のスキーマに基づいて情報を絞り込む機能が求められる。独自のスキーマを活用することは、「メタデータによる精緻な問い合わせ処理」への対応にもつながる。つまり、何らかのカテゴリ内分野についての各種検索条件を組み合わせた検索要求に対し、データベースを活用して精緻な結果を返すことが可能となる。

また、既存の検索機能はパソコンでのウェブサイト利用を想定したものであるが、スマートフォンやタブレット型端末などタッチパネル式機器を想定した検索機能も検討するべきである。具体的にはタグ検索のように、あらかじめ検索候補となる単語と、単語に索引されたページ情報を用意することで、タッチやクリックによる操作(以下、タッチ式操作)のみで検索できる機能を意味する。タッチ式操作で完結する検索機能があれば、DHEAT など業務として健康危機管理に取り組む利用者が、業務開始前後の移動中などにタッチパネル式端末から H-CRISIS を閲覧するような利用にも対応可能となる。

以上より、独自のデータベースを用いて情報を絞り込むことが可能であり、少ない操作回数で検索可能な機能の導入が、H-CRISIS のユーザビリティを高めることになると考えられる。

3.2. 関連事例とユーザビリティを高める検索機能の検討

本節では、関連事例として用語定義や各種ウェブサイトで提供されている検索機能を参照し、既存の検索機能の特徴やユーザビリティを高める検索機能を検討する。

図書館情報学用語辞典第 5 版によると「検索」とは、「いずれ利用されることを想定して蓄積された情報の中から、指定あるいは想定された属性に合致する情報を取り出す行為や操作」[5]である。また記事などの文章に対し、「特定部分に容易にアクセスできるよう、アクセスの手がかりとなる語(見出し語)を一定の規則(五十音順やアルファベット順など)で排列し、各見出しのもとに該当する情報の所在指示を記載したリスト」を、「索引」という[5]。索引が存在するため、指定あるいは想定された属性に合致する情報を素早く取り出すことが可能となる。既存の検索機能において、カテゴリ別の記事のみを表示する機能の場合、当該カテゴリに登録された記事 ID、記事スラッグ(URL の一部)、記事タイトル、記事作成日、記事本文、当該カテゴリ登録記事数、そして記事が登録している当該カテゴリのカテゴリ ID とその名称を、それぞれ索引として利用している。一方でサイト内検索機能は、活用元の google カスタム検索エンジンが、何を索引化し、どのようにテキスト検索しているのか、各々情報公開範囲が限定的なため厳密な索引対象や検索方法を示すことができない。ただし、検索結果から推察すると、検索文字列に関連する記事 URL、記事タイトル、記事作成日、記事本文については、索引化していると考えられる。

保健医療分野で検索を主目的とするウェブサイトは、PubMed や医中誌 Web など論文検索が可能なサイトや、厚生労働省法令等データベースサービスなど法令を検索するサイトがある。表 1 では、これら代表的な検索サイトおよび事例ウェブサイトで提供されている検索機能についてまとめたものである。なお全文検索とは、例えば「健康危機管理に関係する」という文章を検索した際に、同文を含むページ情報一覧を表示するといった機能である。また、類似検索はある単語に類似する単語や文章を含む検索結果を表示する機能、絞り込み検索はカテゴリ検索や日付など関連情報と AND 検索や OR 検索などの検索条件を組み合わせる機能を、それぞれ意味する。類似検索や本文検索は、google によるサイト内検索機能によって既に H-CRISIS 上に実装しているが、絞り込み検索は実装していない。表 1 やユーザビリティの観点から、絞り込み検索機能は保健医療に関係するウェブサイトとして実装することが適切と考えられる。

健康危機管理以外の各分野におけるウェブサイトで提供されている検索機能としては、タグを活用した検索がある。これは、単語や期間、分野などを事前に用意し、画面に表示された任意の文字列(タグ)を選択すると、文字列に対して索引されたページ情報一覧を表示する、という検索機能である。World

表 1 健康危機管理情報を扱うウェブサイト上の検索機能(*は登録不要だが機能有)

	PubMed	医中誌web	厚労省法令等DBサービス	世界保健機関(WHO)
URL	https://pubmed.ncbi.nlm.nih.gov/	https://login.jamas.or.jp/(法人版)	https://www.mhlw.go.jp/hourei/index.html	https://www.who.int/home/search
カテゴリ検索	○	○	—	—
全文検索	—	—	○	○
類似検索	○	○	—	—
絞り込み検索	○	○	○	○
要登録	—*	○	—	—

Wide Web 上のコンテンツサービス全般において、カテゴリとタグの違いや関係性が定まっているとは言い難いが、本稿では H-CRISIS が採用する CMS である WordPress が意図するカテゴリとタグの違いを採用する。つまり、カテゴリが階層的性質を有しておりサブカテゴリを作成することでより具体的なセクションとして整理可能であり、一方でタグが非階層的であり横断的に事象を分類可能である、という趣旨[6]である。なお、事例 DB のカテゴリ分類は情報提供元である地方衛生研究所が設定した分類を元に設定したものであり、表 1 の医中誌 Web と PubMed におけるカテゴリと対応していない。本稿では WordPress のタグ機能の定義のうち、文字入力を省略して選択的に入力することでユーザビリティに寄与する UI の側面に焦点を当て、そのような UI を「タグ検索」と呼ぶことにする。タグを用いた検索は、地理空間情報や各種統計などオープンデータを提供するサイト[7],[8]で導入されており、LIFULL HOME'S[9]など電子商取引サイトや Instagram などソーシャルネットワークサービス(SNS)などでも導入されている。人々がタグを用いる動機の一つは、情報を端的に要約するためとされている[10]。また、タグの類似概念でハッシュタグというものがある。カテゴリと同様に明確に違いが定まっているとは言い難いが、Instagram などの SNS 上では利用者が自身の選好でキーワード文字列を設定し、同じキーワード文字列が設定されている投稿を分類する機能である。ハッシュタグに関連する情報を閲覧することで、状況に応じた専門的な学習を可能とし認知的成長をサポートするという研究[11]が存在する。これらは SNS 上でのハッシュタグに関する働きであり、タグについても同様の効果が得られる確証は無い。しかし、情報の知覚・解釈の側面では、発信者はタグおよびカテゴリなどを含むタグの類似概念を用いて情報を端的に要約し人々に当該情報を伝えるために、受信者も内容を端的に知るために、それぞれ役立てていることが期待できる。実際に多量のデータを扱うサイトで導入されていることや、認知的な側面での有用性を踏まえると、タグ検索の導入により H-CRISIS のユーザビリティが高まると考える。さらに、絞り込み検索とタグ検索を併用して使う利用場面を想定し、両検索を組み合わせたような検索機能を実装する。つまり、分野や期間を指定する機能と、画面に表示された任意の文字列の選択により検索語が指定できる機能を、ユーザーインターフェース(以下、UI)を工夫して実現する。ただし、後者の機能は厳密にはタグ検索ではないので「検索語推薦機能」と呼び、詳細は後述するが絞り込み検索に追加して提供する。

以上を踏まえ、H-CRISIS の新規検索機能として、絞り込み検索、タグ検索を導入することとした。

4. 検索機能の実装

4.1. 絞り込み検索の導入

これまでの議論から健康危機管理情報を検索する場合に絞り込み検索として適切な機能は、(1)カテゴリを単独あるいは複数指定できること、(2)情報が発信された日や事例発生日など期間を指定できること、(3)AND 検索と OR 検索の指定ができること、以上 3 点が想定される。一方で、概要や背景などの文章形式の情報は、後述のタグ検索の活用によって絞り込み検索に適合させる。つまり、絞り込み検索として適切な機能として、(4)被害者数を範囲指定できること、(5)事例における死者の有無を指定できること、(6)画面に表示された任意の文字列の選択により検索語が指定できること、以上が考えられる。また事例 DB では、事例名(件名)、事例 ID、インシデント原因別の分野カテゴリ名、管轄地方衛生研究所名、報告者の所属・氏名、事例終息状況、事例発生日、事例終息日、発生地域、発生規模、患者被害報告数、死亡者数、原因物質、記事全文に加え、後述するタグ検索実装を通じて得られたタグ化した単語を、DB の項目として取り扱う。一方で、類似検索に用いる単語の類似性をベクトル形式で表した情報や、ハッシュタグに対応する項目は無い。その点で、表 1 の PubMed や医中誌 Web のような類似検索機能を提供できる体制を現在の事例 DB は有していない。一方で、今回は実装していないが、全文検索を実装可能とする素地があると言える。

また、絞り込み検索の UI も検討する必要がある。絞り込み検索自体は、検索したい単語と検索時に方法を指示するための検索式を、組み合わせることで検索可能である。しかし、検索式の複雑さやタッチパネル式機器での利用を踏まえると適切ではない。そこで、絞り込み検索の条件が指定しやすいような UI が重要となる。具体的なウェブサイトは割愛するが絞り込み検索 UI としては、物件検索サイトなどで提供されている UI などがある。この種の UI には具体的な名称が存在しないので、便宜的に絞り込み UI と呼ぶ。物件検索サイトの絞り込み UI では、月家賃などの数値を範囲で指定や、二階以上に存在する物件など特性を条件に設定するかをボタン式で指定が、それぞれ可能である。これは、(1)から(6)の機能をタッチ式操作で実現する際に、絞り込み UI により実現可能なことを示唆している。そこで、絞り込み UI によって、(1)から(6)の機能を実現するものとする。

4.2. タグ検索およびタグランダム表示機能の導入

タグ検索の導入にあたり、どのような単語をタグにするべきかを検討した結果、事例 DB の各記事に登場する文章を形態素解析し、その結果得られた単語を活用するものとした。図書館情報学用語辞典によると形態素解析とは、「文を、意味を持つ最小単位である形態素に分割し、品詞や活用、見出し語などを判別・付与する処理」を意味する。この解析により、事例 DB の各記事に登場する単語を集計することが可能となる。今回はオープンソースの日本語形態素解析および単語分割エンジンである MeCab[12]による形態要素解析を行い、単語を抽出した。この際辞書データとして、「Neologism dictionary for MeCab v0.0.7」を用いた。抽出データに対し、例えば指示語である「これ」や、「0」など、文書の要約や、端的な表現としての機能を果たせない形態素を削除し、表記ゆれを一部統合するクリーニング処理を実施した。さらに形態素の登場回数を数えるが、この時形態素の登場回数計数方法として、対象とする文献群全体の形態素登場回数を合計する term frequency と、文献別の登場回数である document frequency がある。今回、document frequency により形態素を数えた際、専門的用語の形態素が登場する文献数が非常に少ないケースが多かった。後述する形態素登場回数をタグ併記した際に、このような形態素タグの注目度や他タグと比較したタップ頻度が下がるのではないかと懸念し、term frequency を採用した。その結果、登場回数 10 回以上の形態素がおよそ 2,700 あった。クリーニングを踏まえて、形態素の登場回数、および全形態素に占める形態素の累積割合をグラフ化したものが図 1 である。このグラフでは、形態素の登場回数を縦棒グラフで表し、形態素全登場回数を分母とする累積の形態素登場回数割合を登場回数の多い順に線グラフで表している。また下部の各形態素は、各数値および割合に該当する形態素からランダムに選出したものである。具体例として、「児童」という形態素は 195 回登場したが、この登場回数は全形態素中上位 5% に入り、多い順の累計割合でおおよそ 40% となる。事例 DB の形態素解析結果から、メタデータの種類としてタグ化する単語を検討する。今回、基本的に登場回数が一桁の単語は採用しないものとする。一方で、グラフ中の「違法ドラッグ」などは登場回数が 4 回と比較的少ないが、登場頻度 10 回未満の原因物質や症例などでも、健康危機管理情報を網羅する観点から重要な形態素がある。そこで、a. 登場回数が 10 回以上である形態素、b. 10 回未満ながら重要性が高いと判断した形態素、以上 2 つの基準によってタグ化する形態素を決定する。なお b については、健康危機管理の研究者（実務経験 5 年）である筆頭著者が、登場回数 10 回未満の形態素一覧を見て、経験的に特に重要と判断した形態素を採用した。a に該当した形態素数は 2191、b に該当した形態素数は 9 であった。また、登場頻度を利用者に対し明確化するために、事例 DB における形態素登場回数をタグに併記する。以上のタグ検索機能を、情報受信者がタグを介して情報を得ることで学習のサポートになるという認知的な側面を踏まえて、閲覧者に向けて新しい情報の認知のためにランダムにタグを表示させる形式で実装した。

同時にこれらタグの文字列と登場回数を活用した検索語推薦機能を絞り込み検索に加えることで、前節で適切とした絞り込み機能の「(6) 画面に表示された任意の文字列の選択により検索語が指定できること」が可能となる。この時、ある単語の検索語推薦機能単体で得られる検索結果は、同一単語のタグ検索による検索結果と同じである。

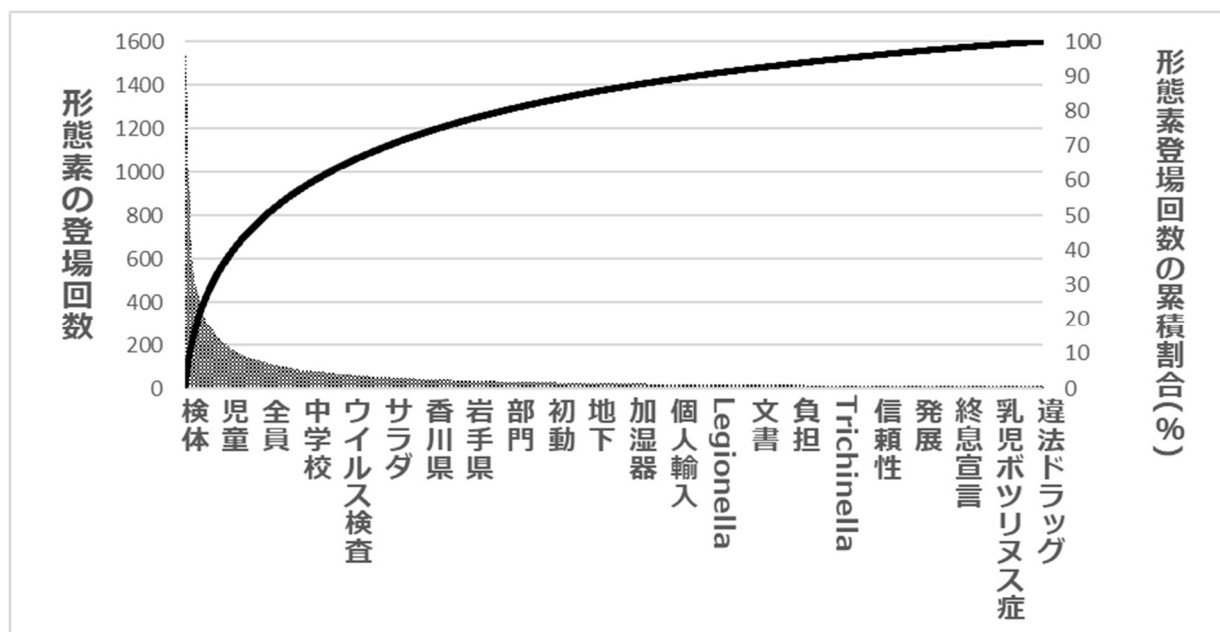


図 1 事例 DB 全記事の形態素解析結果(下部の形態素はランダムに選出)

なお、既存の検索機能2点(カテゴリ記事一覧検索、サイト内検索機能)と、検索語推薦機能を含む絞り込み検索、タグ検索についての仕様を、表2にまとめた。

表2 各検索機能の仕様

	データソース数 (2024年3月末時点)	索引化対象要素	テキスト検索の方式
カテゴリ記事 一覧検索	31,210	各カテゴリに登録された記事ID, 記事スラッグ, 記事タイトル, 記事作成日, 記事本文, 当該カテ ゴリ登録記事数, 記事が登録している当該カテ ゴリのカテゴリID, カテゴリ名称	—
サイト内検索機能	31,210	記事URL, 記事タイトル, 記事作成日, 記事本文 (以上は検索結果からの推測, 詳細は不明)	googleカスタム検索 に準拠
絞り込み検索	1,535	「カテゴリ記事一覧検索」の索引化対象要素, 分野 名, 患者数, 死亡者の有無, タグ化したキーワード, キーワードが登場する記事ID, キーワード登場回数	OSSのRDBMSに実装 されているSQL構文で 各テーブルを検索
タグ検索	1,535	タグ化したキーワード, キーワードが登場する記事 ID, キーワード登場回数	OSSのRDBMSに実装 されているSQL構文で 各テーブルを検索

4.3. 実装した検索機能

H-CIRISIS に検索機能を実装し、図2の絞り込みUI形式のUIを設けた。図2の説明を通じて、実装した機能について説明する。

H-CRISIS のウェブサイトアクセスし、「健康被害危機管理事例データベース」のタブから、事例DBにアクセスする。事例DB ページ内の「詳細検索」ボタンを押すと、図2のUIがポップアップする。このUIでは、検索したいキーワード、分野、期間、確認患者数、死亡者の有無について、入力および選択できる。また、選択したいキーワードは、「頻出語タグ」の各ボタンを押すと、検索窓に当該単語が自動入力される。これらのボタンは、前章の方法にてタグ化した形態素と、その登場回数を表示する。図2はUIのイメージであり、①「下痢」というキーワードを記事に含む、②「ウイルス性食中毒」カテゴリに該当する、③2000年以降、④確認患者数が10人以上、⑤死亡者の有無は問わない事例、以上①から⑤の条件全てを満たす記事を検索するよう指定している。この条件で検索すると、2024年現在、15件の記事見出し一覧結果が返される。各リンクを辿ると、各事例についての報告書が閲覧可能である。詳細検索機能は、事例DB内いずれのページからも利用可能である。

また、事例DB ページの右下に、頻出語例をランダムに表示するタグ検索の領域を設けた。このスペースに表示されるタグを押すと、タグと関係する記事見出し一覧が表示される。ランダムに情報を与えることについては、人間の創造力を刺激する可能性、先入観的な思考パターンを中断させること、新しい連想を引き起こすことなど、認知的な側面での有用性が指摘されている[13]。専門職や研究者などが、健康危機に関するタグとしてメタデータを視認することで、新規情報を認知させ検討に良い影響を与えることを目指した。

4.4. 実装した検索機能の評価

本稿で提案した検索機能の性能を評価する。既存の検索機能の場合、「事例データベース」という一単語で検索すると、2024年時点で約3,190件の候補が表示された。この時「約」を用いた概数値であるのは、既存検索機能であるgoogleサイト内検索が11ページ以上の検索結果を表示できず、概数値としてシステムがカウントした数値を記載しているためである。この検索条件で、確認可能な検索結果が100件、その中で事例DBに関連する記事が7件、関連しない記事が93件、事例DBの記事数が1,535件、事例DBの記事で同検索条件の検索結果として得られなかった記事が1528件であった。従ってこの時の適合率は $7/(7+93)=0.07$ 、再現率は $7/(7+1,528)\approx 0.005$ であった。一方、絞り込み検索で条件を指定せず検索した場合、1,535件の検索結果が得られる。これらの記事は全て事例DBに関係する記事であり、同時に事例DBを意図して検索した際に想定する記事全てを含むため、適合率および再現率ともに1であった。

絞り込み検索について、期間2件(2010年以降および2023年以降)と分野名3件(細菌性食中毒、化学

① 下痢

検索条件: and or

頻出語タグ (三角クリックで全件表示/一部表示切り替え) ▼

検体(1307) 菌(1070) 施設(1048) 食品(995) 分類(748) 感染(728) 調理(710) 発症(652) 環境(640) 喫食(622) 情報(618) 衛生(592)

集団(588) 汚染(588) ウイルス(588) 報告(551) 分析(503) 対策(478) 下痢(481) 状況(454) 感染症(448) 従事(447) 対応(445) 保健(427)

血清(401) 株(388) 試験(387) 衛生研究所(386) 依頼(385) 細菌(372) O157(369) ノロウイルス(359) 終息(351) 提供(350) 研究所(347)

② 分野名

細菌性食中毒(272) ウイルス性食中毒(122) 細菌性感染症(213)

ウイルス性感染症(199) リケッチア感染症(7) 原虫・寄生虫・衛生動物(27)

自然毒等による食中毒(143) 化学物質による食品汚染(144) その他(258)

③ 時期: 2000 年 ~ 西暦 年

④ 患者数: 10 人 ~ 人

⑤ 死亡者の有無: 全て / 有り / 無し

検索

図 2 実装した検索機能用ユーザーインターフェース(UI)

物質による食品汚染、リケッチア感染症)の各検索結果数群(平均 146 件)と、既存検索において絞り込み検索と同じ効果を期待できる検索式(期間の場合「after:2010」など、分野名の場合分野名を直接入力)から得られた各検索結果数群(平均 2, 598 件)を、Wilcoxon の符号付き順位和検定によって比較した。結果は $p < 0.05$ であり統計的に有意、つまり絞り込み検索が既存検索と比較してより少ない検索結果数をほぼ確実に得られることが確認できた。タグ検索および検索語推薦機能についても同じ検定方法で、タグ化した形態素の中で登場頻度が最も多い上位 10 単語に関し、既存検索と、タグ検索および検索語推薦機能各々の検索結果数サンプル 2 群 (既存検索の検索結果数平均 5,811 件、タグ検索などの検索結果数平均 134 件) を比較した。結果は $p < 0.05$ であり統計的に有意、つまりタグ検索および検索語推薦機能についても既存検索と比較してより少ない検索結果数をほぼ確実に得られることが確認できた。以上より新規検索機能が、利用者が意図する正確な記事群を示すことを可能とし、「メタデータによる精緻な問い合わせ処理」へ対応したと考える。

5. まとめ

本稿では、H-CRISIS の特徴と既存検索機能の問題点に言及し、ユーザビリティを高めることを目的として、事例 DB を対象に新たな検索機能を試験的に実装するものとした。ユーザビリティを高めるという観点から、健康危機管理情報を提供する代表的なウェブサイトの検索機能提供状況から絞り込み検索を、健康危機管理以外のウェブサイトでタッチ式操作のみで検索可能であり認知的側面でも有用であるタグ検索を、それぞれ新規検索機能として実装を決めた。絞り込み検索に適切な機能を 6 件定義し、それらを満たす UI として絞り込み UI を採用した。タグ検索のために事例 DB を形態素解析し、解析結果を踏まえて、登場頻度などの観点からタグ化する形態素を決定した。検索画面を踏まえて、実装した新規検索機能の使い方と、ランダムに頻出語を表示する機能に言及した。実装した機能を活用した検索結果を例示し、「メタデータによる精緻な問い合わせ処理」へ対応したことを示した。

今後の展望として、サイト全体への新規検索機能拡充を計画している。一方で課題として、実際にユーザビリティが高まったかどうかを、アンケートなどを通じて調査していない点がある。また、今回タグとして採用した形態素について、出現頻度の桁数スケールにより採用単語を選択したが、タグの選出

形態素や採用量基準について十分な事前検討を行っているとは言えない。加えて、形態素登場回数数の計数方法を **term frequency** としたが、タグ上のキーワード横に表示される数値と、該当記事数の数値の差があるため、ウェブサイト上で説明しているものの直感的には違和感があるかもしれない。その点を踏まえ、表示数値に採用する指標を、ユーザの意見を踏まえつつ検討したい。H-CRISIS サイト全体の検索は検索に対する要求段階の第三段階である「メタデータ発生源や研究データの管理主体の複数組織化、それらの連携を考慮した問い合わせ処理」に該当する。しかし、タグ検索機能のサイト全体への導入には困難性が伴う。理由の一つが記事固有 ID の重複である。現状 H-CRISIS の各領域自体は同じ CMS で扱っている一方で領域別に情報ソースが異なり、領域別に記事固有 ID が設定されている。タグ検索機能は記事固有 ID を要するが、領域間で固有 ID が重複している例がある。以上についてサイト全体の検索機能の仕様を検討中である。更なるユーザビリティの改善や課題抽出のために、今後改めて既存の研究を参照して H-CRISIS に適切な条件を検証し、利用者を交えた十分な議論による意見集約が考えられる。

謝辞

本研究の実施にあたり、国立保健医療科学院健康危機管理研究部 田中克典氏、山地香代氏の事務的支援を得ました。また、検索機能のサーバ上への展開について株式会社 R102 の技術的支援を得ました。感謝申し上げます。

参考文献

- [1] 国立保健医療科学院, “H-CRISIS (健康危機管理支援ライブラリー),” <https://www.niph.go.jp/h-crisis/>, 2002, 2025 年 2 月 12 日参照。
- [2] 厚生労働省, “厚生労働省健康危機管理基本指針,” 2001.
- [3] JIS Z 8521:2020, “人間工学—人とシステムとのインタラクション—ユーザビリティの定義及び概念,” 2020.
- [4] 菊地伸治, 田辺浩介, 坂本浩一, 高田安裕, 傳法春樹, 門平卓也, 谷藤幹子, “研究データ管理における検索機能の実装とその変遷,” 情報システム学会誌, Vol.19, No.1, 2023, pp.52-70.
- [5] 日本図書館情報学会用語辞典編集委員会, “デジタル版『図書館情報学用語辞典 第 5 版』,” 2020, 2025 年 2 月 12 日参照。
- [6] WordPress Foundation, “Categories vs. Tags: What’s the difference?,” <https://learn.wordpress.org/tutorial/categories-vs-tags-whats-the-difference/>, 2025 年 2 月 12 日参照。
- [7] 一般社団法人 社会基盤情報流通推進協議会, “G 空間情報センター,” <https://front.geospatial.jp/>, 2024 年 11 月 10 日参照。
- [8] 福岡市, “自治体オープンデータ,” <https://www.open-governmentdata.org/>, 2024 年 11 月 10 日参照。
- [9] 株式会社 LIFULL, “【ホームズ】 タグから探す | 新しい生活スタイル・暮らしから探すならタグで検索,” <https://www.homes.co.jp/smp/chintai/ltag/>, 2024 年 11 月 10 日参照。
- [10] Rieger, D., & Klimmt, C., “The daily dose of digital inspiration: A multi-method exploration of meaningful communication in social media,” *New Media & Society*, Vol.21, No.1, 2019, pp.97-118.
- [11] Carpenter, J. P., Krutka, D. G., & Kimmons, R., “#RemoteTeaching & #RemoteLearning: Educator tweeting during the COVID-19 pandemic,” *Journal of Technology and Teacher Education*, Vol.28, No2, 2020, pp.151-159.
- [12] 京都大学情報学研究科, 日本電信電話株式会社コミュニケーション科学基礎研究所, “MeCab: Yet Another Part-of-Speech and Morphological Analyzer,” <https://taku910.github.io/mecab/>, 2025 年 2 月 12 日参照。
- [13] Malthouse, E., et al., “The influence of exposure to randomness on lateral thinking in divergent, convergent, and creative search,” *Cognition*, Vol.218: 104937, 2022.

著者略歴

小森 賢一郎 (こもり けんいちろう)

2020 年東京工業大学大学院総合理工学研究科博士課程修了。2022 年国立保健医療科学院健康危機管理研究部研究員、現在に至る。

富尾 淳 (とみお じゅん)

2009 年東京大学大学院医学系研究科博士課程単位取得退学。東京大学大学院医学系研究科公衆衛生学講師等を経て、2021 年より国立保健医療科学院健康危機管理研究部部長、現在に至る。