

Analysis of Healthcare Information on the Web and Discussion of Non-topical Classification

石田, 栄美
九州大学附属図書館研究開発室 : 研究員

<https://doi.org/10.15017/27253>

出版情報 : 九州大学附属図書館研究開発室年報. 2012/2013, pp.29-34, 2013-09. Kyushu University Library
バージョン :
権利関係 :



論文

ウェブ上で提供されている医療・健康情報の実態と 非主題カテゴリへの分類のための考察

石田 栄美[†]

<抄録>

ウェブ上の医療・健康情報の実態調査を行った。2語からなる検索語6種を用いてサーチエンジンで検索した4,762固有URLを対象にしたところ、ポータル、Q&A、ブログなどのサイトが多かった。この中から1,759URLを対象にページの内容を調査したところ、解説・情報提供、日記・雑談、Q&Aなどの内容が多かった。最後に、非主題カテゴリを設定し、1,759ページに非主題カテゴリごとのレベルを付与した。その結果、非主題カテゴリとして、リーディングレベル、客観性・主観性、信頼性に関するカテゴリが有効である可能性が高いことがわかった。

<キーワード> 医療情報、健康情報、ウェブ、非主題カテゴリ、自動分類

Analysis of Healthcare Information on the Web and Discussion of Non-topical Classification

ISHITA Emi

1. はじめに

高齢化社会の進展、健康への関心が高まる一方で、ウェブ(World Wide Web)の普及等により、誰もが比較的容易に情報の入手ができるようになってきた。このような背景の中で、近年、医療・健康情報に対するニーズが高まってきているといえる。公共図書館などでは、これらのニーズに応えるために、医療・健康情報の提供サービスを行っている館も増えている。たとえば、東京都立中央図書館では、健康・医療情報コーナーを設けており、ホームページ上でも健康・医療情報関係のブックリスト、雑誌リストなどを公開している¹⁾。その他、鳥取県立図書館をはじめとして、医療・健康情報コーナーや健康情報棚、闘病記文庫などを設けている図書館も多い。米国においても、その情報ニーズは高く、様々な方法により、健康・医療情報に関するサービスを実践している。米国の公共図書館を中心とした健康・医療情報に関する消費者、レファレンスサービス、インターネット上の情報源に関する実態、サービス事例などは『公共図書館員のための消費者健康情報提供ガイド』²⁾に詳しい。少し異なるサービスとしては、ワシントン州シアトルの公共図書館 Beacon Hill 分館が無料のヘルスケアのコンサルティングのセッションを月に1回設けている例などがある。これは、中国人を主とする英語が母国語でない人たちを対象に、SHIBA (Statewide Health Insurance Benefits Advisors) や

CISC (Chinese Information and Service Center) などとの連携・協力をしながら、保険制度に関するコンサルティングや英語以外の言語で対応可能な病院の紹介などを行っているということであった。情報そのものの提供にとどまらず、医療・健康に関わる多様なニーズに応えるための一例であるといえる。

ウェブは、サーチエンジンを用いることで、専門的な情報から一般的な情報まで多種多様な情報を得ることができるだけでなく、専門機関から一般のユーザーまで自由に情報発信ができるメディアである。現在では、ウェブそのものが一つの情報サービスとなっており、ウェブから情報を得ることが一般的な行為として行われるようになってきている。

本稿では、情報ニーズが高いと考えられるウェブ上の医療・健康に関する情報の提供実態を把握し、これらの情報により多面的にアクセスできるような方法を検討する。本研究では、その方法として、非主題カテゴリへの分類の可能性について考察する。

以下では、医療関係機関が作成しているウェブページでどのような内容が提供されていたかを既往調査から概観する。さらに、サーチエンジンから得られる検索結果を対象に、情報提供を行っているサイト、実際に提供されている内容などを調査した結果について述べる。最後に、非主題カテゴリへの分類に対して考察する。

[†] いた えみ 九州大学附属図書館研究開発室研究員 (〒812-8581 福岡市東区箱崎 6-10-1) E-mail:ishita.emi.982@m.kyushu-u.ac.jp

2. 医療機関のウェブページの内容分析

ウェブ上の医療・健康情報に関する提供実態は、少し前の調査ではあるが、日本の医療機関を対象に情報の内容を分析した既往研究がいくつかある。碓^{3,4)}や橋本ら⁵⁾は、『関東病院名簿 2001/2003 年版』に記載されていた東京都・神奈川県・千葉県の上記ホームページを開発していた病院を対象にページの内容調査を行った。その結果、橋本らは、患者への病院のアクセシビリティを高めるための情報や、病院のブランドづくりに関する情報の提供が進んでおり、一方で、医療情報の公開や患者教育情報は1割から2割程度の記載であったことを明らかにしている。なお、この研究での「医療情報の公開」とは、医師の経歴・学歴、人員体制、医療設備の説明、医療成績、費用情報、カルテの開示であり、「患者教育・健康教育」とは、疾患情報、予防・生活管理情報、治療方法情報、患者会情報、他の医学サイトへのリンク、地域連携情報をさしている。本研究で、注目している医療・健康情報は、「患者教育・健康教育」に含まれる情報に近い。

その他にも、歯科医療機関ホームページの掲載情報を調査した例⁶⁾や全国の保健所のホームページの内容を調査したもの⁷⁾などがある。自治体のホームページにおける外国人向け医療情報の提供状況を調査したもの⁸⁾もある。

これらの調査のほとんどは、医療機関を中心にしたものであり、2002年から2003年に発表されている調査事例が多い。新しいものでも2008年である。2002年ごろは、サーチエンジンが今ほど普及しておらず、ホームページに関して、会社や団体などがページを作成し情報を提供することが一般的で、現在のようにブログ、ツイッターを通じて一般の人々が膨大な情報を発信している状況ではなかったといえる。そのため、これらの調査結果と、現在のウェブ上で提供されている情報の様子とは大きく変わっていると推測できる。次章では、サーチエンジンの検索結果からアクセスできる情報に関する調査を行う。

3. ウェブ上で提供されている医療・健康情報

3.1. 医療・健康情報のウェブページの URL の収集

医療・健康情報については、様々な情報が存在すると思われる。本調査では、内容の多様性に注目するため、主題としては特定の疾患に限定して調査を行う。まずは、医療・健康情報が提供されていると考えられるウェブページの URL を収集した。URL の収集は、2013年2月から3月にかけて、サーチエンジン Google を用いて行った。検索語は、疾患名とその疾患と組み

合わせて用いられると考えられる語の2語を用いた。複数の語を用いたのは、偏りが生じる可能性はあるが、ひとつの検索語の場合だと多くの URL を得ることができないためである。実際に用いた検索語は、「ぜんそく 治療」「ぜんそく 症状」「ぜんそく 原因」「ぜんそく 予防」「小児ぜんそく」「ぜんそく 経験」である。それぞれの検索において、提示された検索結果の上位から URL を収集した。それぞれ894件、851件、951件、1037件、890件、842件の URL が得られ、合計すると5,465URL が得られた（この作業は、検索 API を用いずに手作業で行った）。得られた URL 集合には、重複が含まれていたため、それらを除去した結果、4,762 固有 URL を得た。

3.2. 医療・健康情報を提供しているサイト

上で得られた URL 集合に対し、サイトごとに集計した。検索結果に含まれたサイトは1,715種であり、その上位15件までを表1に示す。サイト名は、サイトに表示されている名称である。カテゴリは、著者がそれぞれのサイトを見て判断した。「Q&A」とは、質問者から寄せられた質問に対して回答者が回答できる専門のサイトである。質問者も回答者も、不特定多数の匿名の人が利用できるものである。「ポータル」とは、ぜんそくであれば、ぜんそくの症状、原因、対処法、病院の検索など様々な情報を提供しているサイトのことである。

表1から、もっとも多いのは、Yahoo!知恵袋であり、11位にも登場している。Yahoo!知恵袋での質問には、夜寝ているときは全くでない喘息の原因についてや、気管支喘息の原因や予防について質問している例などがあつた。15位の「AskDoctors」は、医師に相談できる Q&A サイトである。一般の人からの質問に対し、医師からの回答を得ることができる。このサイトには、その他に、病気・症状事典、お薬事典、病院検索などの項目がある。Q&A サイトの中でも、不特定多数の質問者・回答者によって成り立つ Q&A サイトである、OKWave、教えて goo などのサイトも上位にあつた。上位15位のうち Q&A サイトは303件(6.4%)を占めており、ブログよりも大きな割合を占めていた。

ポータルサイトに分類されたサイトのうち、最も多かったのは「ハッピー夫婦ドットコム」のサイトであり、すべて/zensoku 以下のディレクトリに存在していた。このサイトは、医療関係者ではない一般の人が運営しているサイトであり、掲示板を開発するなどして情報交換などをできる場を提供している。6位の「チェンジ喘息！」は、喘息の正しい知識と治療についての総合サイトであり、製薬会社2社が提供、医師が内容について監修している。喘息の治療法、喘息の薬、

表1 医療・健康情報を提供しているサイト

カテゴリ	サイト名	URL	ページ数(割合)	
1 Q&A	Yahoo! 知恵袋	detail.chiebukuro.yahoo.co.jp	132	2.8%
2 ポータル	ハッピー夫婦ドットコム 共働き夫婦の子供のぜんそくにすぐ効いた!生活術	happyfu-fu.com/zensoku	91	1.9%
3 Q&A	OKWave	okwave.jp	64	1.3%
4 ブログ	Livedoor ブログ	blog.livedoor.jp	53	1.1%
5 ブログ	Ameba ブログ	ameblo.jp	52	1.1%
6 ポータル	チェンジ喘息!	naruhodo-zensoku.com	49	1.0%
6 ポータル	喘息の総合情報サイト	zensoku.jp	49	1.0%
8 ポータル	病院なび momo	momo365.jp	48	1.0%
9 ポータル	喘息の原因と治療・総合館	ppaapp.com/zensoku	41	0.9%
10 販売	花粉症予防対策 2013 最新ランキング	kahunbousinm.blog.so-net.ne.jp	39	0.8%
11 Q&A	Yahoo! 知恵袋	chiebukuro.yahoo.co.jp	38	0.8%
11 ブログ	goo ブログ	blog.goo.ne.jp	38	0.8%
13 Q&A	教えて!goo	oshiete.goo.ne.jp	37	0.8%
14 販売	アマゾン	www.amazon.co.jp	34	0.7%
15 販売	Googl ブックス	books.google.co.jp	32	0.7%
15 ブログ	Yahoo! ブログ	blogs.yahoo.co.jp	32	0.7%
15 Q&A	AskDoctors	asthma.askdoctors.jp	32	0.7%

N = 4,762

予防のポイントなどの項目がある。「喘息(ぜんそく)の総合情報サイト」も製薬会社が管理・運営しているサイトであり、成人喘息、小児ぜんそくに関する情報提供や病医院検索を提供したりしている。「病院なび momo」は、全国の女性向け医院・クリニック・病院検索サービスであり、ヘルスケア関係の会社が運営しているサイトである。喘息を診療項目としている病院の情報が検索結果に多く含まれていたため、上位にランクインしたと考えられる。

以上の結果から、現在では、Q&A やブログなどのように不特定多数の人からの情報発信と、ポータルサイトのように、特定の疾患について総合的な情報を提供しているサイトが多いことがわかった。

3.3. 医療・健康情報のウェブページの内容

検索語に「ぜんそく 経験」「ぜんそく 原因」「ぜんそく 治療」を用いた検索結果のうち、それぞれ 842 件、951 件、60 件を対象に、重複を除いた 1,837 件のそれぞれのページに対し、内容に関するカテゴリの付与を判定者と著者の 2 名で行った。

調査時点で閲覧できなかったもの、動画などに直接リンクしていたものなどを除き、カテゴリを付与したものは 1,759 件であった。その内訳を表 2 に示す。この調査では、内容を基準に判断したため、サイトがポータルであっても、ページに質問とその回答が掲載されている場合は Q&A とした。

ページ数が最も多かった「解説・情報提供」とは、

ぜんそくに関する症状、原因、治療法などについて解説や説明などを行っているページである。表 1 に示したポータルサイト内のページが多いが、クリニックが開設しているサイトの中にも、診療対象であるぜんそくの説明をしているページがあった。その他にも、大きなポータルサイトではないが、ぜんそくに関する様々な情報の提供をしているサイトからのページも多かった。2 番目に多かったのは「日記・雑談」である。個人的な内容をブログ等に記述している場合はここに含めた。ただし、多くの人の個人的な体験・経験をまとめて紹介しているページなどは「体験談・経験談」に含めた。また、喘息の治療のための薬、プログラム、本、吸入器などの器具の紹介、購入を勧めるページは「商品・製品」に分類した。ぜんそくが治ったという経験を紹介し、治療プログラムの購入を勧めるページが目立った。

全体的に見てみると、疾患に関する解説や情報の提供を行っているページと個人的な情報の共有を目指した「日記・雑談」「掲示板」「体験談・経験談」などのページが多いことが明らかになった。これらのページの多くは、情報を共有し、患者会のようなコミュニティを形成することを目的としたものも多かった。つまり、ウェブ上の情報は、医療・健康に関する客観的な事実を提供するためのページと、個人の経験や体験を共有しようとするページが、主に存在するといえる。

表2 ページの内容
カテゴリ ページ数 (割合)

カテゴリ	ページ数	(割合)
解説・情報提供	486	27.6%
日記・雑談	383	21.8%
Q&A	184	10.5%
掲示板	90	5.1%
体験談・経験談	90	5.1%
医療機関紹介	87	4.9%
商品・製品	74	4.2%
ニュース	64	3.6%
本	53	3.0%
医師 Q&A	51	2.9%
その他	197	11.2%
合計	1,759	100%

4. 非主題カテゴリへの分類に関する検討

4.1. 非主題カテゴリへの分類の可能性

ユーザーがある主題に関する情報を見つけない場合、3章の実態調査でも示したように、既存のサーチエンジンを用いることで多種多様な情報を入手することは可能である。しかしながら、ほとんどのサーチエンジンは、ユーザーが入力した検索語とページに含まれる語の照合や主題的関連を重視しており、どのような読者を想定して書かれたページか、書かれている内容に信頼性はあるのか、客観的に書かれているのか、あるいは主観的に書かれているのかなどは、個々のページをみるまでわからない。たとえば、ぜんそくに関する2つのウェブページがあった場合を想定する。一つは医療機関や専門家がぜんそくの治療法について書いたものであり、もう一方はぜんそくの治療体験について書かれたものであったとする。この場合、ぜんそく患者は、信頼性が低かったとしても、気持ちを共有するために、同じ患者の体験を読みたいと思っているかもしれないし、治療方法を検討している者は、どのような治療法が存在するのか、より専門性の高い信頼できる情報を得たいと思っているかもしれない。つまり、ユーザーによって、同じ主題でも異なる観点から情報を求めているケースが考えられる。もし、テキストにこれらの側面からの情報を付与することができれば、ユーザーは、より多面的な角度から情報にアクセスすることが可能となる。そこで、本節では、主題以外の側面からの付加的な情報を付与すること、つまり主題以外のカテゴリへの分類の可能性について検討する。本稿では、これらをノントピカルなカテゴリ、すなわち、非主題カテゴリと呼ぶ。

以下では、いくつかの非主題カテゴリを対象に、医療・健康情報に対して、実際にカテゴリに関するラベルを付与した結果をもとに、これらのカテゴリが有効

かどうかを検証する。

4.2. 医療・健康情報に対する非主題カテゴリの付与

ウェブ上の情報に対して非主題カテゴリの分類を行っている先行研究を参考に、非主題カテゴリとして、「対象とする読者」「リーディングレベル」「客観性」「信頼性」を選定した⁹⁾。3.3節において調査対象としたURL集合に対して、それぞれのページを見ることにより判断した。「対象とする読者」については、一般人、専門家、子どものいずれかを付与し、他の3カテゴリについては、1から5段階の数字でレベルを付与した。各カテゴリの判断基準については、判定は3人で行ったが、最初に、いくつかのページを対象に判定を行い、その結果をもとに判定者間で合議を行うことにより、判断基準が一定になるようにした。その後、一つのページにつき1人の判定者が判断を行った。

「対象とする読者」の結果を表3に、「リーディングレベル」の結果を表4示す。1,759件中、一般人を読者対象に想定したと考えられるページが全体の96.2%であり、圧倒的に多かった。専門家を対象にしたものは65件(3.7%)あり、これらのほとんどは、研究機関の成果報告のページ、CiNiiの書誌情報、J-GLOBALの文献情報のページであった。子ども用のページと判断されたものは2ページしかなかったが、同じサイトのぜんそくの知識についてクイズ形式で学ぶコーナーのページであった。今回の調査対象には、検索語に小児ぜんそくを用いたこともあり、小児ぜんそくについてのページは多かったが、その多くは家族や親に向けてのページであり、子どもに向けてのページは少なかったといえる。

リーディングレベルについては、全体の84.8%が中程度の知識が必要というレベルに分類されていた。これらのほとんどは、対象とする読者が一般人と評価されたものである。高度な知識が必要と判断されたのは、CiNii等から提供されている論文や医師会が提供している症状の解説などで、専門用語がつかわれている場合が多かった。やや高度な知識と判断されたページは、医師によるQ&Aサイトなどが含まれていた。

客観性に関しては、1,759件のうち、主観的であると判断されたページが最も多く、464件(26.4%)であったが、やや客観的であるページやどちらでもないと判断されたページも同等の割合であった。主観的に書かれているページには、患者の体験談などが多く含まれていた。一方、客観的に書かれていると判断されたページは、医療機関や医師などが作成していると判断できるページが多く、情報提供を目的としたページと考えられるものも多かった。

表3 対象とする読者

ラベル	件数	割合
一般人	1,692	96.2%
専門家	65	3.7%
子ども	2	0.1%
合計	1,759	100%

表4 想定されているリーディングレベル

レベル	件数	割合
1: 知識を全く必要としない	11	0.6%
2: やや知識が必要	29	1.6%
3: 中程度の知識	1,492	84.8%
4: やや高度な知識	123	7.0%
5: 高度な知識	104	5.9%
合計	1,759	100%

表5 主観性・客観性

レベル	件数	割合
1: 主観的である	464	26.4%
2: やや客観的である	428	24.3%
3: どちらでもない	452	25.7%
4: やや客観的である	119	6.8%
5: 客観的である	296	16.8%
合計	1,759	100%

4.3. 非主題カテゴリへの分類に対する考察

本調査で設定した非主題カテゴリが医療・健康情報にとって有効であるか、また、それらを判断するための手がかりとしてどのようなものがあるかについて考察する。

対象とする読者は、そのほとんどが一般人であり、専門家が対象であると判断されたページはわずかである。実態調査からも、サーチエンジンから得られるウェブ上の情報のほとんどは一般の人を対象にしたページであったことから考えると、「対象とする読者」というカテゴリを設定する意義が大きいとは判断できない。

リーディングレベルについては、中程度の知識が必要と判断されたページが最も多いが、高度な知識、やや高度な知識が必要と判断されたページも存在し、ユーザーにとっては読みやすさとも関連するため、これらの付加的な情報が付与されていることは有益であるといえる。また、サイト名等から判断できないものも多かった。リーディングレベルの判定方法については、日本語の医療・健康情報に関する読みやすさの研究がある¹⁰⁾。この研究では、読みやすさに影響を与えるものとして、構文、語彙、テキスト構造をあげている。

これらの項目は、リーディングレベルを自動的に判断する際にも用いることができると考えられる。

主観性・客観性に関しては、他のカテゴリほどの偏りもなく、また、実態調査からも主観性を重視したページと客観性を重視したページがそれぞれある程度の割合を占めているため、このカテゴリを付与することは有効であると考えられる。どのように判定するかは、今後の検討が必要であるが、主観的であると判断されたページは、口語体で書かれているものなどが多かったため、文体を用いることが考えられる。

信頼性の判定に関しては、重要な観点ではあるが、実際に判断を行った判定者から非常に判断が難しいという意見があった。その分野の専門家でなければ、本当の信頼性の判断は難しい。ただし、信頼性を自動的に判定する手法がいくつか提案されている。宮森ら¹¹⁾は、信頼性を、情報内容、情報発信者、情報外観、社会的評価という4つの基準で捉えることを提案している。情報発信者については、発信者そのものの信頼度で評価しており、情報外観については、情報ソースや連絡先が明示されているかを基準にし、これらを組み合わせることによって信頼性の評価をおこなっている。寺島ら¹²⁾は、ウェブサイトの信頼性を判断する際の基準として、制作者、情報根拠、作成・改訂年月日、問い合わせ先の4項目を用いている。信頼性の判断については、ページの管理や制作をしている組織や作成者が明示されているか、何かの情報を提示する際に根拠を示しているか、作成年月日などが明示されているかなどの外的な基準を用いて行うほうが、ページの内容で評価するよりも有効であると考えられる。

5. おわりに

本稿では、ウェブ上で提供されている医療・健康情報の実態調査を行い、また、それらに対して、どのような非主題カテゴリを付与することが有効であり、また現実的であるかを検討した。その結果、リーディングレベルや主観性・客観性、信頼性は、非主題カテゴリとして有効であり、また、自動的に分類を行うために用いることができる手がかりも推測することができた。

本調査は、特定の疾患、特定のサーチエンジン、検索語を用いたものであり、ウェブ全体の様子を表現したものではない可能性もある。今後は調査対象を広げることにより、他の医療・健康に関する情報についても同様のことがいえるか、検証していく予定である。

謝辞

本研究は、JSPS 科研 22500220 の助成を受けたものです。

参考文献

- [1] 東京都立図書館, “健康・医療情報サービス,” <http://www.library.metro.tokyo.jp/tabid/408/Default.aspx> (参照 2013-06-12)
- [2] アンドレア・ケニヨン, バーバラ・カシーニ (著), 野添篤毅 (監訳), 公共図書館員のための消費者健康情報提供ガイド, JLA 図書館実践シリーズ 6, 日本図書館協会, 東京, 2007.
- [3] 碓朋子, “2001・2001 年の広告規制緩和前後での医療機関の情報提供活動と消費者が参考にする情報源の変化~医療機関ホームページの内容分析と母親の産科選択を中心に,” 経営と情報, vol.16, no.1, pp.27-42, Oct.2003.
- [4] 碓朋子, “医療機関が Web 上で提供する情報内容の変化—医療機関ホームページの 2000 年・2002 年の比較,” 病院, vol.62, no.9, pp.780-786, Sep.2003.
- [5] 橋本栄里子, 和田ちひろ, 碓朋子, “インターネット上の病院の情報発信内容に関する研究: 病院のホームページは患者に何を伝えているのか,” 医療と社会, vol.11, no.3, pp.69-86, 2001.
- [6] 外山敦史, 森田一三, 外山康臣, 中垣晴男, “インターネット検索エンジンに登録されている歯科医療機関ホームページ掲載情報の現状,” 口腔衛生学会雑誌, vol.52, no. 4, pp.500-501, Aug.2002.
- [7] 瀬戸山陽子, 中山和弘, “全国保健所ウェブサイトの情報発信内容とユーザビリティ, アクセシビリティ評価,” 日本公衆衛生学会雑誌, vol.55, no.2, pp.93-100, Feb.2008.
- [8] 佐々木空美, 長松康子, “自治体ホームページにおける外国人向け医療情報の提供状況,” 聖路加看護学会誌, vol.12, no.1, pp.25-32, Mar.2008.
- [9] E. Ishita, “Non-topical Classification for Healthcare Information,” Bulletin of IEEE Technical Committee on Digital Libraries, Vol.5, No.3, Dec.2009. <http://www.ieee-tcdl.org/Bulletin/current/Ishita/ishita.html>, (accessed 2013-06-12)
- [10] 酒井由紀子, “健康医学情報を伝える日本語テキストのリーダビリティの改善とその評価—一般市民向け疾病説明テキストの読みやすさと内容理解のしやすさの改善実験,” Library and information science, vol.65, pp.1-35, 2011.
- [11] 宮森恒, 赤峯亨, 加藤義清, 兼岩憲, 角薫, 乾健太郎, 黒橋禎夫, “情報の信頼性分析に向けた評価データおよびプロトタイプシステム WISDOM,” 情報処理学会研究報告. 自然言語処理研究会報告, vol.2007, no.76, pp.103-108, Jul.2007.
- [12] 寺島朝子, 町田絵里, 山形真一, 望月眞弓, 佐藤信範, 上田志朗, “医療関連情報の検索におけるインターネットの利便性および Web サイトの信頼性に関する調査: 高齢心筋梗塞患者に対する薬物治療について,” 医療情報学, vol.21, no.6, pp.435-443, Feb.2002.