

國學院大學學術情報リポジトリ

エボラ出血熱文献のビブリオメトリックスとテキスト・マイニングを利用した計量的分析：
小特集図書館と書物の世界

メタデータ	言語: Japanese 出版者: 公開日: 2023-02-05 キーワード (Ja): キーワード (En): 作成者: 青木, 仕, Aoki, Manabu メールアドレス: 所属:
URL	https://doi.org/10.57529/00000459

エボラ出血熱文献のビブリオメトリックスと テキスト・マイニングを利用した計量的分析

青木 仕

I. はじめに

I-1. エボラ出血熱の現状

エボラ出血熱のウイルスが初めて発見されたのは1976年6月スーダンのヌザラという町である。最初に罹患した男性がザイールのエボラ川近くの出身であることから、このウイルス性疾患はエボラ出血熱と命名された。その後エボラ出血熱は、アフリカ大陸で突発的に流行し、致死率は発症国によって50%~90%と非常に高い疾患である。そして、4年前の2014年の爆発的な流行は記憶に新しい⁽¹⁾。エボラ出血熱患者の累計患者数と死亡数を表1.に示す。表1.は、エボラ出血熱の国別症例数および致命率である。エボラ出血熱の発症国は、中央アフリカの3カ国。ギニア、シエラレオネ、リベリアが大半を占めている⁽²⁾。2015年3月での感染者数は約2万4千人、死者数は約1万人に及んでいる⁽³⁾。

表1.から当時感染伝播が継続中の3カ国の確定患者は、シエラレオネが最も多く11,841人、リベリアが9,602人、ギニアが3,429人である。死亡者数はリベリア4,301人、シエラレオネ3,747人、ギニア2,263人である。致命率は、ギニアが最も高く66%、次にリベリア45%、シエラレオネ32%である。感染伝播が終息したアフリカの国では、ナイジェリアは確定患者20人死亡者数8人致命率40%で

表1. エボラ出血熱の国別症例数および致命率
(2014年3月~2015年3月25日現在)

		確定例	死亡例	致命率
感染伝播が継続中	ギニア	3,429	2,263	66%
	シエラレオネ	11,841	3,747	32%
	リベリア	9,602	4,301	45%
感染伝播は終息	アフリカ			
	セネガル	1	0	0%
	ナイジェリア	20	8	40%
	マリ	8	6	75%
	その他			
	英国	1	0	0%
	スペイン	1	0	0%
	米国	4	1	25%
合計		24,907	10,326	41%

WHO : Ebola situation report - 25 March 2015 より引用.

あった。マリは確定患者8人死亡者数6人致命率75%と高い致命率である。セネガルからも1人の確定患者が出ている。アフリカ以外の国としては、確定患者が米国4人、英国とスペインからそれぞれ1人が出ておりその伝播率は脅威である。WHOは2015年5月にリベリア、11月にシエラレオネ、12月にはギニアの流行の終息を宣言したが、その後も2016年3月まで散発的な患者の発生が認められた。エボラウイルスを伝播させる宿主動物は、当初チンパンジーやゴリラが考えられていたが、現在フルーツコウモリの可能性が有力である。エボラ出血熱を含めた感染症は、我が国での発症はみられていないが、今日のグローバル社会化と共に国境を越えた感染の拡大「パンデミックス」が憂慮されている⁽⁴⁾。

I-2. ビブリオメトリックスとテキスト・マイニングとは

本論の研究手法であるビブリオメトリックス（計量書誌学）手法は、科学の実体を定量的に数値化して提示できる手法として現在情報学の世界で広く認知され実践されている。今日のようにビブリオメトリックスの研究が注目されてきた背景には、Priceの著書「リトル・サイエンス、ビッグ・サイエンス」に起因するところが大きい。この著書のなかで彼は、科学の指数的成長の発見と文献の計量的研究の意義を述べている⁽⁵⁾。ビブリオメトリックスという言葉を最初に使ったPritchardは、「ビブリオメトリックスとは記録された情報を数量化し分析することによって、記録されたコミュニケーションのさまざまな局面を分析し、記録されたコミュニケーションの過程や学問領域の発展の流れを明らかにする研究手法である」と述べている⁽⁶⁾。

ビブリオメトリックスによる代表的な法則は、雑誌論文の分散の関係をグラフ化して示したBradfordの法則⁽⁷⁾⁽⁸⁾、科学者の論文の生産性を数値化したLotkaの法則⁽⁹⁾、テキスト中の語彙の出現頻度についてその分布を提示したZipfの法則⁽¹⁰⁾が有名である。

このように、ビブリオメトリックスは情報の生産、伝達、利用について分析するための手段として活用でき、その利用によって実証的なデータを提示できる手法である。今日の図書館情報学分野においても活用されている。

次に本論の二つ目の研究手法であるテキスト・マイニングは、文章を対象とし、大量のテキストデータの中から、大切な情報を抽出し分析することを可能にする手法である。本論では、論文中のタイトルのフリータームをテキスト・マイニングにより主要用語を抽出し、用語間の関連を作図する。このような学術論文を対象にシソーラス用語の分析と共に論文のタイトルのフリータームをテキスト・マイニングで解析した研究は、著者が平成21年（2009年）にアスベスト文献を対象に報告した論文があるのみである⁽¹¹⁾。

今回用いた二つの研究手法のビブリオメトリックスとテキスト・マイニングの併用によりエボラ出血熱研究の量的な経時的変化や質的な内容分析が行えるもの

と考える。さらに、テキスト・マイニングにより作図された用語間の関係からエボラ出血熱研究の重要語が抽出されるものと思われる。その結果、本研究によりエボラ出血熱の国際的な研究動向と我が国のエボラ出血熱に関する研究状況を明らかにできるものとする。

II. 調査内容・分析方法

本論ではビブリオメトリック
スとテキスト・マイニングによる論文タイトルの内容分析を用いてエボラ出血熱の学術論文を分析する。

ビブリオメトリックスによるエボラ出血熱のシソーラス用語の分析には、洋文献は米国のNational Library of Medicine作成のシソーラス用語MeSH (Medical Subject Headings) タームを和文献は日本の医学中央雑誌刊行会作成の医学用語シソーラス第8版を使用する。エ

ボラ出血熱のMeSHターム Hemorrhagic Fever, Ebolaを医学用語シソーラスMeSHのツリーストラクチャー (樹木状の階層構造) を表2.に示す。表2.はMeSHによるエボラ出血熱タームの階層構造である。エボラ出血熱 (Hemorrhagic Fever, Ebola) は、Virus Disease (ウイルス性疾患) の下位RNA Virus Infections (RNAウイルス感染症) の下位Hemorrhagic Fevers, Viral (ウイルス性出血熱) と共にFiloviridae Infections (フィロウイルス感染症) の下位概念として位置づけられている。

本論の調査項目は、1.エボラ出血熱文献数の推移、2.エボラ出血熱洋文献の収載雑誌上位ランキングとその主題分野およびインパクトファクター、3.エボラ出血熱和文献の収載雑誌上位ランキングとその主題分野、4.洋文献のMeSHターム数とフリーターム出現頻度数の比較、5.和文献のシソーラス用語数とフリーターム出現頻度数の比較、6.エボラ出血熱の被引用文献数上位論文ランキング、7.エボラ出血熱文献の被引用文献数第1位の研究者についての7項目である。

エボラ出血熱文献タイトル中の自然語の出現頻度の分析は、テキスト・マイニングによりフリータームの解析を行った。本論のテキスト・マイニングソフトに

表2. MeSHによるエボラ出血熱の階層構造

Hemorrhagic Fever, Ebola
All MeSHemorrhagic Fevers, Viral (1978-1995)
All MeSH Categories
Diseases Category
Virus Diseases
RNA Virus Infections
Hemorrhagic Fevers, Viral
Hemorrhagic Fever, Ebola
H Categories
Diseases Category
Virus Diseases
RNA Virus Infections
Mononegavirales Infections
Filoviridae Infections
Hemorrhagic Fever, Ebola

はトレンドサーチ2008（社会情報サービス社）を使用した。トレンドサーチのコンセプトマッピングは形態素解析の手法を採用し、抽出されたタームの関連性をもとに平面上にビジュアルにマッピングする分析処理法である。タームの関連性をバネ（スプリング）の強さとみなし、そこから物理的なシミュレーションを行い、タームをお互いに引っ張り合わせることによって適正な配置（平衡状態）を可能にする。関連度が高いタームは近くに、低いタームは離れて配置され作図される⁽¹²⁾⁽¹³⁾。なお、トレンドサーチのアルゴリズムは製作会社の特許であるため非公開である。

調査日は、2015年2月25日である。これらの2つの方法により抽出された文献を基礎データとしてエボラ出血熱文献を分析した。

Ⅲ. 調査結果

Ⅲ-1. エボラ出血熱文献数の推移

米国のNational Library of Medicineが1971年に作成したMEDLINEは世界で最も権威のある医学分野のデータベースである。1997年から名称をPubMedとして無料でWeb上から利用できる。2011年からは携帯端末用PubMed Mobilからも利用が可能である。洋文献の検索にはPubMedを用いた。和文献の検索には、有料ではあるが日本国内の医科大学をはじめ病院など国内の医療機関が概ね導入している医学中央雑誌刊行会作成の医中誌Webを用いた。

エボラ出血熱文献は、PubMedから洋文献1,437件、医中誌Webから和文献206件が抽出された。

PubMedと医中誌Webに記載されている、エボラ出血熱文献数の推移を図1.に示す。縦軸がデータベースに登録された年間の文献数、横軸が年度である。

洋文献は、1977年に世界の臨床医学分野を代表するLancet誌に掲載された最初の報告から今日まで継続的に報告されている⁽¹⁴⁾。PubMedによると特に1996年から掲載論文数の増加が読み取れる。エボラ出血熱は今日まで10回の流行が記録されている。なかでも、2014年の流行は最大規模であり、それに伴い洋文献数も年間350件を超え、多くの報告がなされていた。和文献は、1979年に北村敬氏が日本の自然科学雑誌を代表する「自然」誌に発表した「野生ウイルスとパンドラの箱エボラ出血熱流行の記録」という論題の文献が最初である⁽¹⁵⁾。

和文献数は1997年から年間10件前後で推移していた。

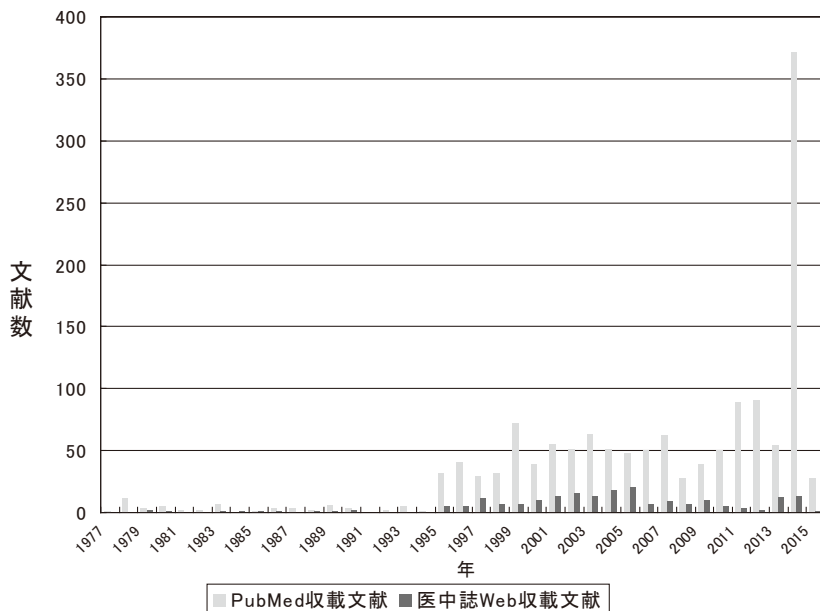


図1. エボラ出血熱文献数の推移

Ⅲ-2. エボラ出血熱洋文献の収載雑誌上位ランキング

エボラ出血熱洋文献の収載雑誌上位ランキングを表3.に示す。表3.は、左からエボラ出血熱洋文献の収載雑誌順位、雑誌名、収載件数、各雑誌の主題分野をNLMC分類法で表示し、次にIF (Impact Factor) 影響度を表示している。エボラ出血熱洋文献の収載雑誌の第1位は感染症学分野 (WC) のJ Infect Disが113

表3. エボラ出血熱洋文献の収載雑誌上位ランキング

順位	タイトル	件数	主題分野(NLMC)	IF(2013)
1.	J Infect Dis	113	WC	5.778
2.	J Virology	68	QW	4.648
3.	Lancet	58	WB	39.207
4.	Science	48	QH	31.477
5.	Nature	45	QH	42.351
6.	N Engl J Med	36	WB	54.420
7.	Virology	33	QW	3.278
8.	MMWR Morb Mortal Wkly Rep	31	WA	-----
9.	BMJ	28	WB	16.378
10.	Wkly Epidemiol Rec	25	WA	-----
合計		400誌		1437件

件でトップであり、第2位はウイルス学分野（QW）のJ Virology 68件、第3位は総合医学分野（WB）Lancet 58件であった。4位はScience 48件、5位Nature 45件とScienceとNatureは世界の自然科学分野（QH）を代表する総合誌がランクされた。6位はN Engl J Med 36件で現代最もインパクトファクターの高い総合医学雑誌がランクされた。7位Virology 33件は第2のJ Virologyと共にウイルス学分野（QW）のコア・ジャーナルである。

8位のアメリカ疾病対策センター（CDC）が発行するMMWR Morb Mortal Wkly Rep（疾病率と死亡率の週刊レポート）31件と10位のスイス発行のWkly Epidemiol Rec（WHOの週刊疫学記録）25件は、いずれも感染者数などの疫学統計を報告する雑誌である。かつて、エボラ出血熱と同様にAIDSが世界中を驚愕させた。32年前、著者がAIDS文献を調査した際にもMMWR Morb Mortal Wkly RepとWkly Epidemiol Recの両誌は上位に抽出された雑誌である⁽¹⁶⁾。

エボラ出血熱洋文献全体の1,437件からは、400雑誌が抽出された。

Ⅲ-3. エボラ出血熱と文献の収載雑誌上位ランキング

医中誌 Web で検索された、エボラ出血熱と文献の収載雑誌上位ランキングを表4.に示す。表4.は左からエボラ出血熱文献の収載雑誌順位、雑誌名、収載件数、各雑誌の主題分野を表示している。

1位は、「臨床と微生物学」であり主題分野（WB:医学一般、

表4. エボラ出血熱と文献の収載雑誌上位ランキング

順位	タイトル	件数	主題分野(NLMC)
1.	臨床と微生物	11	WB, QW
2.	化学療法の領域	8	QD
2.	日本臨床	8	WB
4.	最新医学	7	WB
5.	医学のあゆみ	5	WB
5.	感染症	5	WC
7.	Infection Control	4	WC
7.	公衆衛生	4	WA
9.	感染・炎症・免疫	3 (ほか 10)	WC
18.	感染と消毒	2 (ほか 26)	WC
44.	熱帯	1 (ほか 66)	WC
合計		113誌	206件

QW:微生物学)11件、2位は、「化学療法の領域」(QD:化学)と「日本臨床」(WB:医学一般)8件、4位「最新医学」(WB:医学一般)7件、5位「医学のあゆみ」(WB:医学一般)と「感染症」(WC:伝染病)5件、7位「Infection Control」(WC:伝染病)と「公衆衛生」(WA:公衆衛生学)4件が上位を占めていた。そして、それら上位の雑誌のエボラ出血熱文献の記事区分について調査すると、エボラ出血熱についての解説記事を多く掲載していることが判明した。上位の雑誌の主題分野は、(WB:医学一般)と(WC:伝染病)が多くを占め、(WB:医学一般)と共に感染・炎症・免疫、感染と消毒などの(WC:伝染病)分野の雑誌がエボラ出血熱と関連が深いことが判明した。

Ⅲ-4. 洋文献のMeSHターム数とフリーターム出現頻度数の比較

エボラ出血熱洋文献に付与されていたキーワードについて調査し、エボラ出血熱文献の内容分析を行う。PubMedには、一つの論文に対してシソーラス用語（統制語）のMeSHタームが約20語前後付与されている。そのうち内容を端的に表現したメジャーMeSHタームが5語前後含

表5. 洋文献のMeSHターム数とフリーターム出現頻度数の比較

順位	Mj-MeSHターム	件数	論文タイトル中のフリーターム	回数
1.	Disease Outbreaks	180	Ebola	939
2.	Ebolavirus	68	Virus	519
3.	Hemorrhagic Fever, Ebola	66	Fever	219
4.	Viral Vaccines	19	Hemorrhagic	160
5.	Global Health	15	Diseases	158
5.	International Cooperation	15	Infection	122
5.	Marburgvirus	15	Viral	91
8.	Ebola Vaccines	14	Against	81
8.	Epidemics	14	Haemorrhagic	79
8.	Virus Internalization	14	Outbreak	77
8.	Virus Replication	14		
384種類のターム				

まれている。この調査では、メジャーMeSHタームと論文タイトル中のフリーターム（自然語）の出現頻度数上位10タームを比較し、エボラ出血熱文献の内容分析を行う（表5.）。表5.は、洋文献のMeSHターム数とフリーターム出現頻度数の比較である。左からメジャーMeSHタームと論文タイトル中のフリータームの順位とその件数を表示してある。

メジャーMeSHタームは、1位Disease Outbreaks（疾病の発生）180件と多くを占めていた。2位は生物としてのEbolavirus（エボラウイルス）68件、3位はMeSHタームでもあるウイルス性疾患としてのHemorrhagic Fever, Ebolaが66件を占め、上位3位までのメジャーMeSHタームがエボラ出血熱文献に多く付与されていることがわかる。4位はViral Vaccines（ウイルスワクチン）19件、5位Gloval Health（世界的な健康）とInternational Cooperation（国際協力）とMarburgvirus（マーブルグウイルス）15件と続いていた。洋文献1,437件に対して384種類のメジャーMeSHタームが抽出された。

論文タイトル中のフリータームは、論題中に同じフリータームが複数回付与されている場合もある。そのためメジャーMeSHタームのような文献の件数ではなく、出現頻度の回数としてカウントしてある。

論文タイトル中のフリータームは、1位Ebola（エボラ）939回、2位Virus（ウイルス）519回、3位Fever（発熱）219回、4位Hemorrhagic（出血）160回、5位Diseases（疾病）158回が多くタイトル中に使われていることがわかる。

Ⅲ-5. 和文献のシソーラス用語数とフリーターム出現頻度数の比較

次に、医中誌Webで検索されたエボラ出血熱の和文献について付与されてい

たキーワードと論文
タイトル中のフリー
タームからエボラ出
血熱文献の内容分析
を行う。和文献の
キーワードについて
は、医中誌 Web で
使用されている医学
用語シソーラス第8
版を用いる。シソー
ラス用語と論文中の
フリーターム出現頻
度上位語の比較を表
6.に示す。

表6. 和文献のシソーラス用語数と
フリーターム出現頻度数の比較

順位	第8版シソーラス用語	件数	論文タイトル中の フリーターム	回数
1.	出血熱－エボラ	54	感染症	159
2.	エボラウイルス	16	ウイルス	77
3.	出血熱－ウイルス性	13	エボラ出血熱	72
4.	新興・再興感染症	11	出血熱	45
5.	マールブルグ病	10	感染	24
6.	人畜共通感染症	8	対策	24
7.	感染症予防	6	エボラウイルス	20
7.	輸入感染症（予防）	6	新興再興	20
9.	ウイルスワクチン	4	マールブルグ病	15
9.	ウイルス感染症	4	診断	15
9.	ウイルス被膜タンパク質	4		
88種類のシソーラス用語				

医学用語シソーラスの1位は「出血熱－エボラ」54件で、ウイルス性疾患としての用語である。2位は「エボラウイルス」16件、3位「出血熱－ウイルス性」13件、4位「新興・再興感染症」11件、5位「マールブルグ病」10件と続き、エボラ出血熱文献の和文献には、生物としてのウイルスの概念についての用語が上位を占めていた。和文献206件に対して88種類のシソーラス用語が抽出された。

タイトル中のフリータームは、1位「感染症」159回、2位「ウイルス」77回、3位「エボラ出血熱」72回、4位「出血熱」45回、5位「感染」と「対策」が共に24回と続いていた。これらのフリータームがエボラ出血熱文献の和文献の論題に数多く用いられた用語として抽出された。

Ⅲ－6. エボラ出血熱の被引用文献数上位論文ランキング

次に、エボラ出血熱の被引用文献数上位論文ランキングを検索し、エボラ出血熱研究において重要な論文を抽出する。この調査は、世界の研究者から最も信頼性が高く評価されている被引用解析データベースのWeb of Scienceを用いる。Web of Scienceは、自然科学分野のノーベル賞候補者の選定に用いられるデータベースである。Web of Scienceを用いてエボラ出血熱の被引用文献上位論文ランキングを示す（表7.）。表7.は、被引用数の高い上位10論文の書誌事項を表示したものである。

表7.から抽出されたこれらの文献は、エボラ出血熱研究で数多く引用されているコア論文である。表7.からGeisbert, TWが執筆した論文が1位、3位、6位と上位を占めていた。この結果からGeisbert, TWが今日のエボラ出血熱研究において研究者間で最も注目されているキーパーソンであることがわかる。そして、論文の書誌事項から2位、3位、4位、5位、8位、10位の雑誌は、雑誌

表7. エボラ出血熱の被引用文献数上位論文ランキング

順位	書誌事項	被引用件数
1.	Geisbert, TW. Pathogenesis of Ebola hemorrhagic fever in cynomolgus -- Am J Pathol 2003;163(3):2347-70.	215
2.	Bray, M. A mouse model for evaluation of prophylaxis and therapy of Ebola -- J Infect Dis 1998;178(3):651-61.	164
3.	Geisbert, TW. Mechanisms underlying coagulation abnormalities in ebola -- J Infect Dis 2003;188(11):1618-29.	150
4.	Ksiazek, TG. Clinical virology of Ebola hemorrhagic fever(EHF) --- J Infect Dis 1999;179(Suppl.1):177-87.	143
5.	Khan, AS. The reemergence of Ebola hemorrhagic fever, Democratic Republic of the Congo. 1995 J Infect Dis 1999;179(Suppl.1):76-86.	141
6.	Geisbert, TW. Pathogenesis of Ebola hemorrhagic fever in primate models -- Am J Pathol 2003;163(6):2371-82.	138
7.	Towner, JS. Isolation of Genetically Diverse Marburg Viruses from Egyptian Fruit Bats. PLoS Pathog 2009;5(7):e1000536	133
8.	Villinger, F. Markedly elevated levels of interferon(IFN)-gamma, IFN-alpha - J Infect Dis 1999;179(Suppl.1):188-91.	122
9.	Towner, JS. Rapid diagnosis of Ebola hemorrhagic fever by reverse - J Virol 2004;78(8):4330-41.	118
10.	Bwaka, MA. Ebola hemorrhagic fever in Kikwit, Democratic Republic of the Congo: Clinical observations in 103 patients. J Infect Dis 1999;179 (Suppl.1):1-7	115

上位ランキングでも1位であったJ Infect Dis誌が占めていた。その結果から、J Infect Dis誌はエボラ出血熱論文を掲載するコア雑誌であることも判明した。なお、4位、8位、10位の論文はJ Infect Dis誌の179巻Suppl1号1999年であり、この号はエボラ出血熱の特集号として発行されている。このことから、この特集号にはエボラ出血熱に関する重要な論文が数多く掲載されていたことがわかる。

Ⅲ-7. エボラ出血熱文献の被引用文献数第1位の研究者

次に、表7.「エボラ出血熱の被引用文献数上位論文ランキング」から1位、3位、6位の論文を執筆しているGeisbert, TWについて調査した。Dr. Tom W Geisbertは米国の病理学者でエボラ出血熱、ラッサ熱研究の第一人者であり、それらのワクチンの開発にも努めている。

彼は、2014年5月現在、124本の論文を執筆しており、それらは合計7,117回引用され、論文ごとの平均引用回数は、57.4回と非常に高い引用率を誇っている。

現在、University of Texas Medical BranchでMicrobiology and Immunology (細菌・免疫学) 講座に所属しており、2000年にUniformed Services University of the Health Sciencesで博士号を取得している。このように、被引用解析データベースのWeb of Scienceを用いると、ある主題分野のコアな雑誌や論文と共

たものが図3.である。

図3.のテキスト・マイニングによるエボラ出血熱和文献のタイトル中のフリータームのネットワーク図から、「感染症」、「ウイルス」、「出血熱」の3用語が、エボラ出血熱和文献の中心的な用語になっていることが判明した。

IV. 考察

雑誌に掲載された文献の解析を主な調査対象とするビブリオメトリックスは、論文数分析「学術情報の生産統計」と引用文献分析「学術の情報の消費・利用統計」の大きく二つの研究手法に大別できる。今回ビブリオメトリックスの論文数分析によって、エボラ出血熱文献の1.推移、2.洋雑誌上位ランキング、3.和雑誌上位ランキングを提示した。次にシソーラス用語の調査からは、論文のキーワードとして付与されているシソーラス用語（統制語）の解析により論文の内容を端的に分析した。4.洋文献のMeSHタームとフリータームの比較、5.和文献のシソーラス用語とフリータームの比較を提示した。エボラ出血熱文献は、PubMedから洋文献1,437件、医中誌Webから和文献206件が抽出された。

エボラ出血熱文献数の推移においては、エボラ出血熱を記した最初の文献は1976年に発表され、2014年の大規模な流行に伴いエボラ出血熱への関心も高まり、洋文献数は2014年の一年間で350件を超えていた。

掲載誌上位ランキングにおいては洋文献が400誌抽出され、1位「J Infect Dis」113件、2位「J Virology」68件、3位「Lancet」58件と続き、感染症、ウイルス学などエボラ出血熱に関連の深い分野の雑誌が抽出された。それらの雑誌と共に「Nature」や「Science」と言った自然科学分野を代表する雑誌も抽出された。そして、医学分野の中で最も論文の採択率が低く、投稿しても掲載されるのが難しいと言われている総合医学雑誌の「N Engl J Med」などのインパクトファクターの高い雑誌が上位を占めていた。その結果からエボラ出血熱に対する関心の高さが窺えた。このことから、エボラ出血熱が最近の医学界のみならず自然科学分野からも非常に関心の高いテーマであったことが推測できる。

PubMedのMeSHタームは2017年現在115,845語で統制されている⁽¹⁷⁾。医中誌Webは現在医学用語シソーラス第8版（2015年改訂）で統制されており、ディスクリプタ29,862語（同義語を含めて346,430語）から構成されている。PubMedおよび医中誌Webの両データベースはインデクサーの手作業により、一つ一つの論文にその内容が索引されている⁽¹⁸⁾。論文が著者によって同じような事柄が異なる言葉で表現されていても、このシソーラス（統制語）に基づき統一されたキーワードを付与することにより論文の内容を的確に表現し、漏れの少ない検索に役立っている。そのため、フリータームでの検索よりシソーラス用語を用いての検索の方が、求める論文を的確に検索する可能性が高いと言える。

MeSH ターム上位ランキングでは全体で384タームが抽出され、1位「Disease Outbreaks」180件、2位「Ebola virus」68件、3位「Hemorrhagic Fever, Ebola」66件と続いていた。そして、論文タイトル中のフリータームは1位「Ebola」939回、2位「Virus」519回、3位「Fever」219回に比べ、エボラ出血熱より具体的に表現する用語が抽出された。和文献においても同様であり、シソーラス用語ランキングでは全体で88種類のシソーラス用語が抽出され、1位「出血熱-エボラ」54件、2位「エボラウイルス」16件、3位「出血熱-ウイルス性」と続いていた。論文タイトル中のフリータームは1位「感染症」159回、2位「ウイルス」77回、3位「エボラ出血熱」72回とシソーラス用語に比べエボラ出血熱を表現する平易な用語が付与されていた。

被引用解析データベース Web of Science を用いて、被引用文献数上位ランキングを調査することによりエボラ出血熱研究のコア雑誌と論文と共に中心的な研究者を見いだすことができた。さらに、エボラ出血熱の重要論文は J Infect Dis 誌に掲載されている論文が数多く抽出され、J Infect Dis 誌がエボラ出血熱研究のコア雑誌であることが見いだされた。

被引用文献数上位論文ランキングからは Geisbert, TW が執筆した論文が1位、3位、6位を占め、彼がエボラ出血熱研究の第一人者であることが判明した。

次に、テキスト・マイニングでは、シソーラスに未収録の新しい医学用語やテキスト中の生の用語が抽出でき、執筆者がタイトルで用いた自然語の分析が行える。テキスト・マイニング解析は、大量のテキストデータから価値のある情報を取り出す手法の一つとして今日注目されている。テキスト・マイニング解析では関連度の高い語は近くに、低い語は離れて配置され図中に表示される Barabasi & Albert⁽¹⁹⁾ Newman⁽²⁰⁾。

エボラ文献の論題中のフリータームをテキスト・マイニングにより、出現頻度の高い用語間の関係がマッピングされた。洋文献では「Diseases」、「Fever」、「Virus」が、和文献では「感染症」、「ウイルス」、「出血熱」がエボラ出血熱文献のフリータームから中心の用語になっていることが視覚的に判明された。テキスト・マイニング解析によりネットワーク図にするとフリーターム間の関係が表現でき、その関係が鮮明に提示できることが可能になった。

2018年3月（現在）、PubMedでビブリオメトリックスの手法によるエボラ出血熱に関する文献を検索すると、わずか8件のみであった。検索された8件のエボラ出血熱に関する論文では、論文数の推移、国別生産数、研究者間のネットワークなどの記載のみであり⁽²¹⁾⁽²²⁾⁽²³⁾⁽²⁴⁾⁽²⁵⁾⁽²⁶⁾⁽²⁷⁾⁽²⁸⁾、本研究のような医学用語の分析に関する論文は見出せなかった。また、テキスト・マイニングによるエボラ出血熱に関する文献をPubMedで検索すると、10件の論文が抽出されたが、これらの論文はツイッターなどのソーシャルメディアによるエボラ出血熱の流行の伝播に関する解説記事等であり、本研究のような学術論文を対象にした医学用語の解

析を行った論文は発表されていなかった。

この結果から、本研究はエボラ出血熱文献をビブリオメトリックスとテキスト・マイニングを用いて学術論文を分析した初めての報告であると考えられる。

V. おわりに

ビブリオメトリックスの手法を用いた調査研究は、今日のコンピュータ機能の向上と普及や簡便な統計ソフトの開発などにより手軽に個人で実施できる環境が整ってきている。しかも、莫大な調査研究費も必要としない研究手法である。しかし、ビブリオメトリックス研究の注意点としては、研究の目的に応じて測定項目を選択することが必要である。つまり、現在実施している研究は、どのような特徴がある内容を解明しようとしているのかを常に念頭におきながら分析することが重要となる。ビブリオメトリックスによって抽出されたデータは、解釈や考察を十分に結して結論を導き出すことにより、科学的に実証性があり、周囲を納得させるだけの有用な資料となり得る。

ビブリオメトリックスを用いた研究は、大学の評価、学術論文の国際比較、学術研究の動向の分析、図書館経営や学術コミュニケーションの分析など多くの分野で今後も活用されることが期待できる⁽²⁹⁾。ビブリオメトリックスは現代の図書館の運営においても、資料の購入時の選定や資料の廃棄時期の設定、貸し出し記録による利用頻度の高い図書「ベストリーダー」の選定など多方面で活用されている。

最後に、本研究のようにビブリオメトリックスとテキスト・マイニング2つの研究手法を併用することにより、種々の研究テーマについてもその内容分析と計量的分析や経時的な主題変遷の研究など多面的なアプローチが可能である。今後このような研究分野の発展と共に2つの研究手法を活用した研究の報告を期待したい。

本稿は平成27年（2015年）7月18日・19日の両日、北海道大学で開催された第32回医学情報サービス研究大会における発表に加筆したものである⁽³⁰⁾。

参考文献

- (1) 加藤康幸. エボラ出血熱：西アフリカにおける過去最大の流行. 日本内科学会雑誌. 2017;106(3):405-8.
- (2) 有馬雄三, 島田智恵. 西アフリカのエボラウイルス病発生状況. ウイルス. 2015;65(1):47-54.
- (3) 高田礼人. エボラウイルス研究の現状と最新の知見. 医学のあゆみ. 2016;258(7-8):803-10.
- (4) 島田智恵. 国境を越えて広がる感染症-エボラウイルス病-. 臨床と微生物. 2015;42(3):225-9.
- (5) Price, Derek J. de Solla. Little Science and Big Science. New York, Columbia University Press, 1963. 島尾永康訳. リトル・サイエンス ビッグ・サイエンス. 創元社, 1963.
- (6) Pritchard, A. Statistical bibliography or bibliometrics. Journal of Documentation, 1969;

- Vo.25, p.348.
- (7) Bradford, S.C. Sources of Information on specific subjects. *Engineering*. 1934;137:85-6.
 - (8) Bradford, S.C. Documentation. London. Lockwood & Son. 1953; p144-59.
 - (9) Lotka, A.J. The frequency of distribution of scientific productivity. *Journal of the Washington Academy of Science*. 1926; 16(12):317-23.
 - (10) Zipf, G.K. *Human Behavior and the Principle of Least Effort*. Cambridge, MA. Addison Wesley. 1949.
 - (11) 青木 仕, 青木きよ子. わが国のアスベスト研究の分析: 文献中のシソーラス用語とタイトル中のフリータームの解析. *順天堂医学*. 2009;55(4):478-86.
 - (12) 渡部 勇: 緩い協調: 強調フィルタリングシステム. *情報処理学会研究会報告 (ヒューマンインターフェース)*, 1991;91:179-86.
 - (13) 神田陽治, 渡部 勇, 三木和男, 他: グループ発送支援システムGrIPS. *人口知能学会誌*, 1993;8:65-74.
 - (14) No authors listed. After Marburg, Ebola. *Lancet* 1977;1(8011):581-2
 - (15) 北村 敬. 野生ウイルスとバンドラの箱エボラ出血熱流行の記録. *自然*.1979;34(3):76-7.
 - (16) 青木 仕. AIDS 文献の計量的分析-NLM Literature Searchを利用して-.*医学図書館*,1986;33(4):327-37
 - (17) 阿部信一. Medical Subject Headings (MeSH) 2017年版の変更点. 2017;38(1/2):1-15.
 - (18) 医学中央雑誌刊行会.医学用語シソーラス第8版. [internet]. http://www.jamas.or.jp/service/service_o/thesaurus.html [accessed 2018-03-14]
 - (19) Barabasi AL, Albert R. Emergence of scaling in random networks. *Science*. 1999;286:509-12.
 - (20) Newman MEJ. The structure and function of complex Networks. *SIAM Review*. 2003;45:167-256.
 - (21) Hegel C, Weidemann F, Gauch S, Edwards S, Tinnemann P. Analysing published global Ebola Virus Disease research using social network analysis. *PLoS Negl Trop Dis*. 2017 oct 9; 11(10):e0005747.
 - (22) Sweileh WM. Global research trends of World Health Organization's top eight emerging pathogens. *Global Health* 2017 Feb 8; 13(1):9.
 - (23) Garg KC, Kumar S. Bibliometrics of global Ebola Virus Disease research as seen through Science Citation Index Expanded during 1987-2015. *Travel Med Infect Dis*. 2017 Mar-Apr; 16:64-5.
 - (24) Yi F, Yang P, Sheng H. Tracing the scientific outputs in the field of Ebola research based on publications in the Web of Science. *BMC Res Notes*. 2016 Apr 15; 9:221.
 - (25) Zhao XY, Sheng L, Diao TX, Zhang Y, Wang L, Yanjun Z. Knowledge mapping analysis of Ebola research. *Bratisl Lek Listy*. 2015; 116(12):729-34.
 - (26) Cruz-Calderon S, Nasner-Posso KM, Alfaro-Tolosa P, Paniz-Mondolfi AE, Rodriguez-Morales AJ. A bibliometric analysis of global Ebola research. *Travel Med Infect Dis*. 2015 Mar-Apr; 13(2):202-4
 - (27) Baiyewu O. Outbreak of Bbola virus disease in the West Africa. *Afr J Med Med Sci*. 2014 Jun; 43(2):84.
 - (28) Dias MB, Reyes-Gonzalez L, Veloso FM, Casman EA. *Proc Natl Acad Sci USA*. 2010 May 25; 107(21):9556-61.
 - (29) 青木 仕. 医学情報における情報の解析-新しい情報環境の確立を目指して-. *情報の科学と*

技術, 2003;53(1):34-41.

- (30) 青木 仕. エボラ出血熱文献のビブリオメトリックスによる分析. 第32回医学情報サービス
研究大会抄録集, 2015. P.18.