

いわゆるハゲタカジャーナルに関する文献調査

A literature survey on so-called predatory journals

栗山正光

KURIYAMA, Masamitsu

首都大学東京 学術情報基盤センター

Tokyo Metropolitan University. Library and Academic Information Center

あらまし：学術文献データベース Scopus と Web of Science を利用して、ハゲタカジャーナルに関し、どのような文献が、いつ、どのような学術雑誌に発表されているか調査した。文献検索の過程で predatory という比喩的な語が検索の精度を落とすことが確認され、両データベースの検索結果の比較によって検索漏れや未収録データが発見された。Scopus も Web of Science も単独では網羅的な文献収集には不十分なことが明らかになった。

キーワード：ハゲタカジャーナル，ハゲタカ出版社，ハゲタカ学会，情報検索，Scopus, Web of Science, predatory journals, predatory publishers, predatory conferences, information retrieval

1. 研究の背景, 目的

近年、オープンアクセスの進展に伴って、いわゆるハゲタカジャーナル、ハゲタカ出版社 (predatory journals/publishers) が問題になっている。predatory とは捕食性とか略奪的という意味だが、米国コロラド大学デンバー校の図書館員ジェフリー・ビール(Jeffrey Beall)が、論文掲載料(APC)目当ての悪徳オープンアクセス出版をこう呼んで糾弾し、欧米で有名になった。これを「ハゲタカ」と訳して日本の大学図書館界に紹介したのは筆者だが、同じ時期にすでにこの言葉を使っていた人がいたかもしれない。いずれにせよ、その後、2018年には毎日新聞に取り上げられ、日本でも注目されるようになった。

本研究では学術文献データベースを利用して、ハゲタカジャーナルに関し、どのような文献が、いつ、どのような学術雑誌に発表されているか計量書誌学的な調査をし、また文献レビューを行うことによって、この問題への学術界の対応を探ろうと試みた。しかし、その過程で、この問題に関する文献を効率よく網羅的に収集することがなかなか困難で、情報検索の事例研究としても興味深いことが判明した。

そこで、今回は文献データベースにおける検索の試行錯誤の過程と検索結果を整理し、吟味することにより、情報検索上の課題を明らかにすることを目的とした。

2. 研究の方法

文献収集を行うためのデータベースとしてエルゼビア社の Scopus とクラリベイト・アナリティクス社の Web of Science(WoS)を選択した。どちらも学術文献データベースとして定評があり、信頼性が高いとされている。この二つのデータベースを、できるだけ効率的かつ検索漏れが生じないような検索語や検索式を探りつつ、同じ方法で検索した。その結果をダウンロードして Excel で読み込み、比較対照し、片方では検索されなかったものは再度他方を検索し、双方の未収録データおよび検索漏れを発見した。検索漏れについてはその理由を確認した。

これらの作業は 2019 年 9 月後半から 10 月上旬にかけて行った。

表 1 検索方法と検索結果

	Scopus ヒット数	うち適合	精度	WoS ヒット数	うち適合	精度
predatory AND 検索 ¹⁾	6,051	?		1,391	?	
predatory フレーズ検索 ²⁾	542	542	1.00	485	485	1.00
predatory 近接 1 語 ³⁾	26	17	0.65	15	12	0.80
predatory 近接 2 語	43	29	0.67	24	22	0.92
predatory 近接 3 語	22	4	0.18	4	2	0.50
predatory 近接 4 語	21	1	0.05	4	1	0.25
predatory 近接 5 語	20	0	0.00	4	2	0.50
predatory 近接 6 語	18	3	0.17	7	3	0.43
predatory 近接 7 語	20	2	0.10	4	1	0.25
predatory 近接 8 語	17	1	0.06	7	0	0.00
predatory 近接 9 語	16	0	0.00	6	0	0.00
predator* フレーズ ⁴⁾	9	3	0.33	5	2	0.40
predator* AND "open access" ⁵⁾	66	17	0.26	40	4	0.10
fake フレーズ ⁶⁾	23	19	0.83	20	17	0.85
questionable フレーズ	23	11	0.48	14	6	0.43
fraudulent フレーズ	38	33	0.87	28	27	0.96
pseudo フレーズ	5	2	0.40	4	1	0.25
計	909	684	0.75	671	585	0.87

1) predatory AND (journal* OR publish* OR publication* OR conference*)

2) "predatory journal*" OR "predatory publish*" OR "predatory publication*" OR "predatory conference*"

3) predatory NEAR/1 (journal* OR publish* OR publication* OR conference*)
以下 NEAR/9 まで. ヒット数は上の検索結果を除外した数字. 以下同様.

4) "predator* journal*" OR "predator* publish*" OR "predator* publication*" OR "predator* conference*"

5) "predator*" AND "open access" AND (journal* OR publish* OR publication* OR conference*)

6) "fake journal*" OR "fake publish*" OR "fake publication*" OR "fake conference*"

以下 fake の部分を置換

3. 検索方法と検索結果

表 1 に検索方法と Scopus, WoS それぞれの検索結果をまとめた. 検索フィールドはデフォルトにもなっているタイトル, 抄録, キーワードである. いろいろな語を試したところ, predatory の後に journal[s], publisher[s], publishing, publication[s], conference[s]などが続く場合が多いことがわかった. しかし, 1)のように, predatory とこれらの語(いずれも前方一致)との AND 検索を行うと, Scopus で 6,051 件, WoS で 1,391 件となり, ノイズが非常に多い. これは predatory が比喩的な語で, 本来の「捕食性」の意味で使われている生物学などの論文が大量にヒットしてしまうためである.

そこで 2)のようなフレーズ検索を行うと, 件数は激減したが, ざっと目で確認した限り, ノイズはゼロとなった. これだけでは検索漏れがあるので, 3)のような近接演算を 9 語まで行っ

た. NOT 検索で直上の検索結果を除外して当該距離だけのヒット数にしてある. predatory open access journals のように 2 語離れている場合が多い. 3 語以上離れている場合は少ないが, 全く無いわけではない. ちなみに 8 語でヒットしたのは”publisher boasts the ranking, this paper shows how these predatory magazines”という部分である (タイトルは”About the excesses of the university publication”).

4)のフレーズ検索を入れたのは“predator journals”というものが見つかったからである. ただし, predatory を predator*と前方一致にするとノイズが増える. 例として”... period of the predators. Published data show ...”といった風にピリオドをまたいで隣り合っているものがヒットする (これは検索システムの不都合とも感じられるが).

5)は predator* と journal*等との AND 検索ではノイズが多すぎるので, ”open access”という語も加えてみたものである. ヒット件数がそれなりに収まり, predatory との近接演算で見つからなかったものが見ついている. ただし, やはり結構ノイズはあるし, ”open access”を入れることで漏れてしまうものも多くあるので, これだけで済ますわけにはいかない.

さらに, ハゲタカを表すのに predatory 以外にも多くの言葉が使われている. ここではその中から fake, questionable, fraudulent, pseudo という 4 語を選び, 6)のような形のフレーズ検索を試みた. これらは predatory に比べると件数も少ないが, ビールが登場する以前の文献が検索されている.

2)のフレーズ検索以下を集計すると, Scopus ではヒット件数 909 件, そのうち適合するのは 684 件 (精度 75%), WoS ではヒット件数 671 件のうち適合 585 件 (精度 87%) という結果になった.

4. Scopus と WoS の比較

Scopus と WoS の検索結果をダウンロードし, Excel でマージする作業を行った結果, 表 2 に示すような件数となった. Scopus では 684 件, WoS では 585 件のうち 489 件が同じ文献で, それぞれ 71%, 84% を占める. Scopus からのみ検索されたのは 195 件だが, そのうち 25 件は実は WoS にも収録されており, 検索漏れである(#6). 従って Scopus のみに収録されているのは 170 件ということになる(#5). 同様に WoS のみで検索された 96 件のうち Scopus の検索漏れは 28 件だった(#8). ユニークタイトル, すなわち共通の文献を一つと数えた場合の合計は 780 件である(#9).

表 2 Scopus と WoS の件数まとめ

#1	Scopus検索結果	684
#2	WoS検索結果	585
#3	検索結果単純合計	1,269
#4	Scopus,WoS共通	489
#5	Scopusのみ収録	170
#6	Scopusヒット, WoS検索漏れ	25
#7	WoSのみ収録	68
#8	WoSヒット, Scopus検索漏れ	28
#9	Scopus,WoS重複削除後合計	780

検索漏れの理由は, いずれもタイトルには上述の検索に使用した語が入っておらず, Scopus にはついている抄録が WoS ではついていなかったり, Scopus では付与されていない predatory journals というキーワードが WoS では付与されていたりするといったものである. ただし, Scopus の場合, 多くのレコードに SciVal Topic Prominence のトピック Periodicals as Topic | Publishing | Predatory publishers が与えられている. これが検索項目になっていないため検索漏れとなってしまうのは釈然としない.

他の例では, Asadi, A [et al.] “Fake/Bogus Conferences”はタイトルをみただけで適合文献とわかるが, “fake conference*”というフレーズ検索ではヒットしないため Scopus で検索漏れとなった. WoS では fake journals というキーワードが付与されていたためヒットした.

特殊な例として, “Open Access: Is There a Predator at the Door?”というタイトルの記事が

ある。これは耳鼻咽喉科関係の医学雑誌十数誌の編集長が連名で各誌に掲載した論説で、Scopus では **publication** や **open access publishing** というキーワードが付与されているので、上記 5) の検索条件で 12 件がヒットする。WoS ではキーワードが付与されていないためヒットしないが、タイトルで検索すると Scopus を上回る 14 件がヒットする。掲載誌が違うので DOI もそれぞれ異なるが、文章は全く同じであり、大量の検索漏れと言うには抵抗がある。

また別の例だが、Scopus の Singh Chawla, D. “The undercover academic keeping tabs on 'predatory' publishing” と WoS の Chawla, DS “Sites warn against 'predatory' journals” とが同じものだと気づくのに随分時間がかかった。Nature の Correspondence の同一記事で DOI も同じだが、HTML 版と PDF 版でタイトルが異なり、Scopus は HTML 版、WoS は PDF 版のタイトルを採用している。著者の姓の取り方も違っているため、著者名とタイトル名だけでは別物としか見えない。この例ではそれぞれタイトルに **predatory publishing**, **predatory journals** という語があるので両方ともヒットするが、Beall, J. “Predatory journals: Ban predators from the scientific record” (Scopus) が “Ban predators from the scientific record” (WoS) となると、短縮でキーワードも付与されていないため、ヒットしなくなってしまう。Goodman, S.N. “A quality-control test for predatory journals” (Scopus) に至っては “A serious nonsense publishing proposal” (WoS) と、ヒットもしないし内容も想像できないようなタイトルに変わっている。

データの食い違いも見つかった。Acta Dermato-Venereologica に掲載の Richtig, G “Knowledge and influence of predatory journals in dermatology: A pan-Austrian survey” と Shalom, G. “Predatory journals in dermatology: A hidden danger” の 2 本は、Scopus では 2018 年、WoS では 2019 年となっていて、これは WoS が正しい。

5. 文献数の推移とまとめ

検索結果をマージし、重複を除いた 780 件の文献について、出版年ごとの数の推移を図 1 に示す。2014 年あたりから増え始め、2017 年に急増していることがわかる。1987 年から 2011 年の間も少数ながら検索されているが、**predatory** という語でヒットするのは、ビールのリストが知られ始めた 2012 年以降である。それ以前は **fake** や **fraudulent** といった言葉が使われており、オープンアクセスとも関係がない。出版と結びついた研究不正は昔からあるが、インターネット時代ゆえの問題の広がり、**predatory** という catchy な言葉の影響の大きさをうかがわせる。

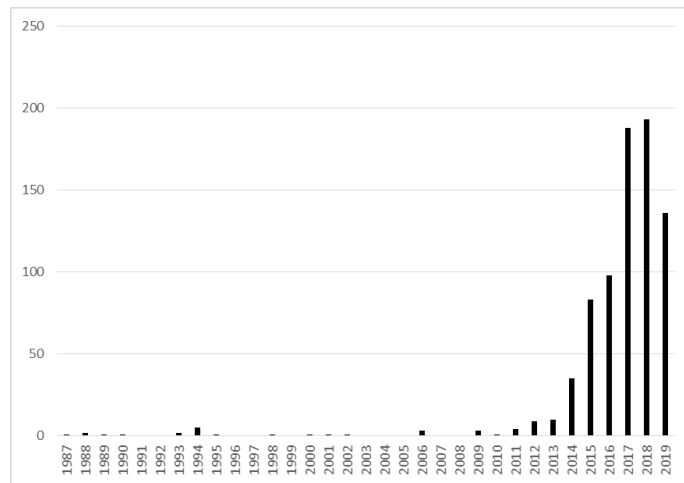


図 1 文献数の推移

本研究ではハゲタカジャーナル（出版社、学会）に関する文献を効率的かつ網羅的に検索する方法を探ったが、上で見たように、比喩的に用いられているキーワードゆえにノイズを防ぐのが困難だった。これを解決するには意味を理解する AI が必要かもしれない。また、Scopus, WoS 双方に検索漏れを起こす要因やデータの誤り・不足があり、いずれかにしか収録されていない文献も少なくないことが判明した。定評ある高価な文献データベースといえども、単独では網羅的な文献収集には不十分な現状が改めて明らかになったと言えよう。