

いわゆるハゲタカジャーナルに関する文献調査 [における情報検索の課題]

首都大学東京 学術情報基盤センター 栗山正光

2019年11月2日(土)

情報メディア学会第21回研究会@東北福祉大学

研究の背景, 目的

- 近年, 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: 捕食性, 略奪的
- 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
- 日本でも毎日新聞が取り上げる
- ハゲタカジャーナルに関する文献調査の過程で, 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, 情報検索の事例としても興味深いことが判明
- 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, 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

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
 - どちらも定評があり, 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
- 検索結果をダウンロードしてExcelで読み込み, 比較対照し, 双方の未収録データおよび検索漏れを探索
- 検索漏れについてはその理由を確認
- 作業は2019年9月後半から10月上旬にかけて実施

検索方法と検索結果(1)

- 検索フィールドは、タイトル、抄録、キーワード
 - WoSではトピック(タイトル、抄録、著者キーワード、Keywords Plus)
- predatoryの後に journal[s], publisher[s], publishing, publication[s], conference[s]などが続く場合が多い
- predatory AND (journal* OR publish* OR publication* OR conference*)
 - ヒット件数: 6,051(Scopus), 1,391(WoS) ノイズ多数
 - 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
- "predatory journal*" OR "predatory publish*" OR "predatory publication*" OR "predatory conference*"
 - ヒット件数: 542(Scopus), 485(WoS)
 - ノイズはほとんどないが、検索漏れがある

検索方法と検索結果(2)

- predatory NEAR/1 (journal* OR publish* OR publication* OR conference*)
 - ヒット26, 適合17 (Scopus), ヒット15, 適合12 (WoS)
- predatory NEAR/2 (journal* OR publish* OR publication* OR conference*)
 - ヒット43, 適合29 (Scopus), ヒット24, 適合22 (WoS)
- 近接3語以上は精度が落ちるが, 全くないわけではない
- 近接8語の例も: "publisher boasts the ranking, this paper shows how these predatory magazines ..."
- 抄録に雑誌名や出版社名が入っていてノイズになる場合も
 - (例) IOP Publishing Ltd. The predatory impact of civil construction industry ...

検索方法と検索結果(3)

- "predator* journal*" OR "predator* publish*" OR "predator* publication*" OR "predator* conference*"
 - “predator journals”という言葉が見つかったため
 - ヒット9, 適合3 (Scopus), ヒット5, 適合2 (WoS) #近接演算でヒットしたもの除く
 - ノイズ例: “... period of the **predators. Published** data show ...”
- predator* AND "open access" AND (journal* OR publish* OR publication* OR conference*)
 - ヒット66, 適合17 (Scopus), ヒット40, 適合4 (WoS) #近接演算でヒットしたもの除く
- "fake journal*" OR "fake publish*" OR "fake publication*" OR "fake conference*"
 - ヒット23, 適合19 (Scopus), ヒット20, 適合17 (WoS)
 - questionable, fraudulent, pseudoでも同様にフレーズ検索, いくつかの適合文献

検索結果一覧

	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

ScopusとWoSの件数まとめ

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

ScopusとWoSの比較(1)

- 検索漏れを引き起こす違いがある
- Scopusでついている抄録がWoSではついていない場合がある
- WoSで付与されているpredatory journalsなどのキーワードがScopusでは付与されていない場合がある
 - Scopusで SciVal Topic Prominence のトピック “Periodicals as Topic | Publishing | Predatory publishers” が付与されているのにキーワード検索でヒットしないのは不可解
- 例1: Asadi, A [et al.] “Fake/Bogus Conferences”
 - “fake conference*”というフレーズ検索ではヒットせず, Scopusでは検索漏れ
 - WoS ではfake journals というキーワードが付与されているためヒット

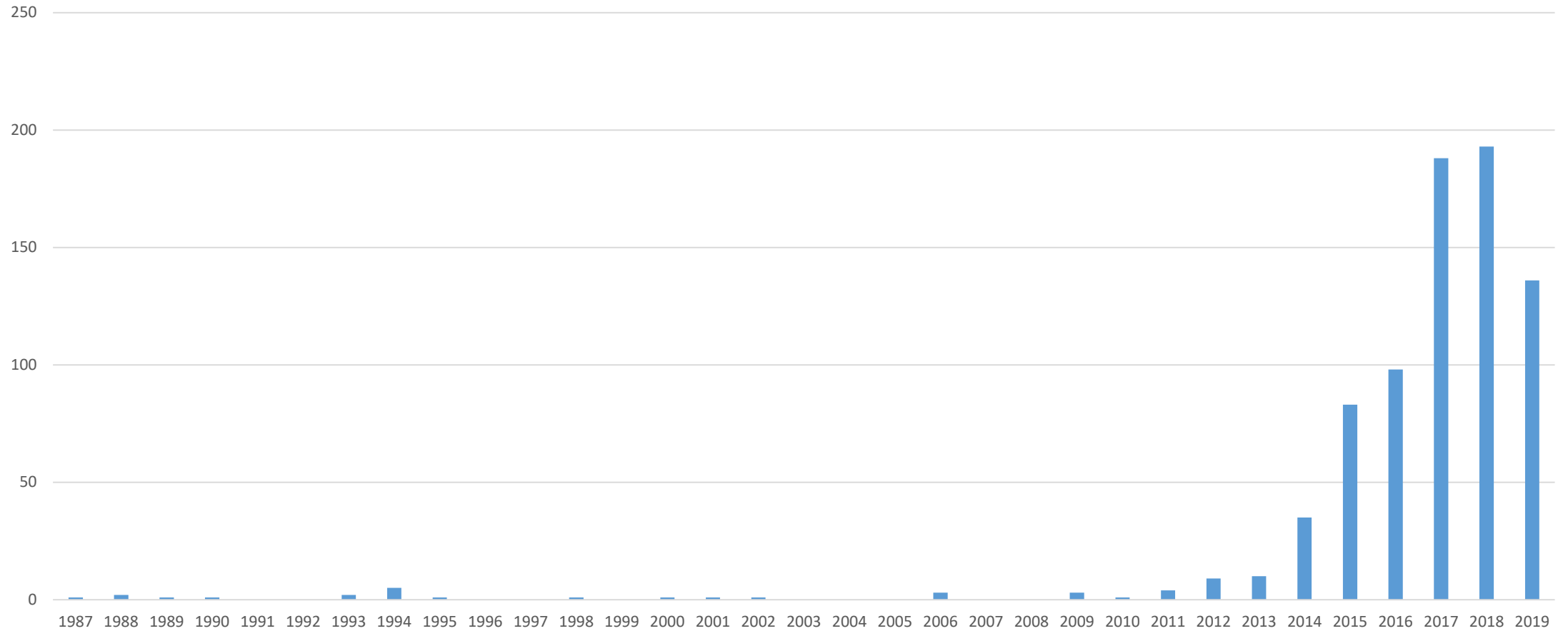
ScopusとWoSの比較(2)

- 例2: “Open Access: Is There a Predator at the Door?”
 - 耳鼻咽喉科関係の医学誌14(?)誌の編集長が連名で各誌に掲載した論説
 - DOIは別だが, タイトル・本文は同じ, 著者は順番が違う場合も
 - 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, predator* AND “open access” AND (journal* OR publish* OR publication* OR conference*)で 12件ヒット, WoSではヒットしない
 - タイトルで検索するとWoSでも14件ヒット
- 例3: Singh Chawla, D. “The undercover academic keeping tabs on ‘predatory’ publishing” (Scopus)
Chawla, DS “Sites warn against 'predatory' journals” (WoS)
 - 『ネイチャー』のCorrespondence の同一記事で, DOIも同じ
 - ScopusはHTML版, WoSはPDF版のタイトルを採用(幸いどちらもヒット)

ScopusとWoSの比較(3)

- 例4: Beall, J. “Predatory journals: Ban predators from the scientific record”(Scopus), “Ban predators from the scientific record”(WoS)
 - 『ネイチャー』のCorrespondence の同一記事, WoSではヒットせず
- 例5: Goodman, S.N. “A quality-control test for predatory journals”(Scopus), “A serious nonsense publishing proposal”(WoS)
 - 同上. WoSではヒットせず, 適合文献かどうかタイトルからは判断不能
- 例6: Richtig, G “Knowledge and influence of predatory journals in dermatology: A pan-Austrian survey”
Shalom, G. “Predatory journals in dermatology: A hidden danger”
 - 掲載誌: Acta Dermato-Venereologica
 - Scopus では2018年, WoS では2019年となっていて, WoS が正しい

適合文献数の推移 (Scopus+WoS)



文献数の推移

- ハゲタカジャーナル(出版社、学会)に関する文献は2014年あたりから増え始め、2017年に急増
- 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
- 1987年から2011年の間はfakeやfraudulentといった語が使われる
 - オープンアクセスに限らない、出版をめぐる研究不正の問題
- インターネット時代ゆえの問題の広がり
- predatoryという言葉のインパクト

まとめ

- ハゲタカジャーナル(predatory journals)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、比喩的に用いられているキーワードゆえにノイズが大量にヒット
- フレーズ検索で精度を上げると検索漏れが生じる
- predatory以外の語を使った関連文献も少数ながら存在
- Scopus, WoS 双方に検索漏れの原因となるデータの不足や誤りがあり、いずれかにしか収録されていない文献も結構ある
- 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