

令和 2 年 7 月 9 日現在

機関番号：62615

研究種目：挑戦的萌芽研究

研究期間：2016～2019

課題番号：16K12833

研究課題名（和文）研究文献生産性分析の基盤構築：その動態の解明とシミュレーションモデル

研究課題名（英文）Development of system infrastructure for analyzing research paper productivity

研究代表者

中渡瀬 秀一（nakawatase, hidekazu）

国立情報学研究所・情報社会相関研究系・特任研究員

研究者番号：90599896

交付決定額（研究期間全体）：（直接経費） 1,900,000円

研究成果の概要（和文）：本研究では研究文献の生産効率等を分析する上で必要となる情報基盤の構築方法について継続性やカバレッジの観点から検討を行い、複数の公開情報源を横断的に統合してこれを構築した。これにより豊富な研究者属性を有し広範な研究分野をカバーできる情報基盤を実現した。この方法で構築した基盤を国立大学教員数ベースでカバレッジ評価した結果、約53%の教員情報をカバーすることが判明した。

研究成果の学術的意義や社会的意義

近年、我が国から産出される論文数が減少傾向にあることから令和元年の科学技術白書でもこの状況への危惧が指摘されている。この状況に対する施策立案に資するデータ分析が求められており、その基礎となる情報基盤の構築は重要な課題である。本研究で構築した情報基盤は、文献生産効率の規定要因を探索可能にする点から学術分析において先駆的な意義を持つ。

研究成果の概要（英文）：In this research, we developed a data warehouse as an information infrastructure for productivity analysis of research papers. In developing this information infrastructure that handles a variety of information about researchers, we first investigated public academic online sources. We then implemented this infrastructure by integrating only those persistent sources based on our findings. Comparing to the total number of national university teachers in 2019 to examine the coverage of researchers, we found that this data warehouse contained information for about half (53%) of national university teachers.

研究分野：学術情報基盤

キーワード：文献生産 文献分析 生産動態 情報基盤 研究動態 研究史 学術史 メトリックス

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

1. 研究開始当初の背景

Science and Engineering Indicators(NFS)や近年の科学技術白書の中でも指摘されるように我が国から産出される論文数が減少傾向にある。そして令和元年版の同白書でも科学技術の基盤的な力の現状としてこの状況が危惧されている。

減少の要因としてマクロ的には研究者人口から算出する研究稼働量(文部科学省調査によるフルタイム換算データで補正)や研究費の低下があることは知られているが、現状では研究リソースに対する追加の余地は限られている。これに対して資金配分的な施策(若手支援や重点化など)も試行されているが十分な効果が認められるには至っていない。

以上の状況において今後も多くの施策の策定が期待されるため、それを実現するためにも多種多様な要因を探索可能にする情報基盤の整備は重要な課題である。

2. 研究の目的

これまで論文等成果の生産効率に関する実証的な分析が十分に進んでいなかった理由として、広範な研究者の情報とその著者目録を網羅的に利用できる情報基盤が存在しないことが挙げられる。

従来、このような分析やそのためのデータベース構築に関する研究は特定の分野や組織に限定して可能な方法で試みられており、本研究が目指すようなスケールできる方法では実現していない。

そこで公開されている各種学術情報の統合的な利用に焦点をあて、文献生産効率の規定要因を探索するのに必要な研究者属性等(所属・職名・経歴・専攻分野・研究者間の人的関連等)と著者目録(成果の一覧)の収集と統合に注目する。

本研究の目的は、国内の研究者個人を単位として文献生産効率の規定要因を上記のように多様な研究者属性から実証的に分析可能とする基盤環境を構築することである。その設計にあたっては運用継続性や情報のカバレッジの面にも配慮する。

3. 研究の方法

本研究では、複数の情報源をもとにして研究者の属性(所属・職名・経歴等)と成果(文献・書籍・その他)を横断的に統合し、文献生産効率を実証的に分析可能な情報基盤の構築を以下の順序で進める。

(1) 情報源の検討

この基盤の運用継続性と情報のカバレッジの観点から、収集の対象となる公開情報源を広く検討する。

(2) 基盤構築と評価

検討して定めた情報源からデータを収集し、それらを統合するシステムを作成する。さらに収集・統合されたデータに関しては研究者数ベースのカバレッジ評価を行う。

4. 研究成果

(1) 情報源の検討

本基盤に収集する研究者の属性や著者目録が含まれる公開情報源を 運用継続性、研究者カバレッジの観点から検討した。

研究に関する情報源はその作成単位から次のように分類できる。研究者が単位である研究機関の教員録や研究者総覧、研究課題が単位の助成金データ、出版元や所蔵元が単位の文献情報である。

運用継続性の観点からは、これまで継続的に逐次更新されていること(一時期に一括作成されたものでないこと)、運営元が安定していること(将来的な利用可能性)、集約的情報源であること(収集プロセスの簡略化)を重視した。

カバレッジの観点からは、大規模な情報源であること、分野限定されないことを重視した。

上記の観点から次の公開情報源を付記する理由により選定した。

【研究者】データベース型研究者総覧(リサーチ・マップ): 過去10年以上、政府の研究開発法人によって運営され、個々の研究機関に限定せず研究者をカバーしている。

【助成金】助成金報告書(JSPS): 上記同様

【文献】学術情報ナビゲータ(サイニィ): 上記理由の他、NDL 雑誌記事・学協会の雑誌記事・

多くの機関リポジトリの文献の索引を集約する情報源である。

(2) 基盤構築と評価

上記の各情報源を対象にそれぞれのサービスAPI¹⁾からデータを取得するシステムを作成した。これらの情報源から取得される主なデータ項目を以下に示す。

- ・人名
- ・所属
- ・職名
- ・学位
- ・専攻分野
- ・授業科目(教員の場合)
- ・経歴(学歴・職歴・委員歴など)
- ・表彰
- ・資格
- ・獲得資金とその研究課題
- ・著者目録(論文・図書・講演等)
- ・産業財産権
- ・作品など
- ・メディアメンション(報道など)

これらのデータ項目は複数の情報源に分散しているが、多くは共通識別子等によって統合されている。識別子には研究者識別子(イーラド番号)や文献等の識別子(Digital Object Identifier)がある。識別子に従属しない項目の場合にはメタデータの著者名・所属等から形式的に結合している。

次に研究者カバレッジの評価を行った。研究者の属性を表すデータ項目は多くがリサーチ・マップから得られるため、それを対象に国立大学別の本務教員数を基準とした評価を行った。ただしこのシステムは常時更新が可能であるため通常ある時点でのスナップショット評価は困難である。そこで評価は2020年1月に新システムへの移行に伴う期間中に実施した。以下にこの評価から得られた特徴について挙げる。

評価は全ての国立大学(86大学)の本務教員を対象とした。2020年1月時点の各大学におけるアクティブな登録者でかつ職名が教員である数²⁾と各大学の本務教員数(2019年³⁾)との比較から評価を行った。上記のアクティブとは少なくとも直近2年間(2018年1月1日以降)に情報を更新しているという意味である⁴⁾。大学別のカバレッジ平均は約53%であり、国立大学総本務教員(約63000人)に対する登録者比率は約53%である。

本務教員数の降順に国立大学の本務教員カバレッジと登録教員数の職位別分布を図1と2に示す。このように本務教員数とリサーチ・マップへの登録教員数との間には相関が見られる($R^2=0.75$, $N=86$)。しかし国立大学においては本務教員数(大学の規模)とそのカバレッジの間には殆ど相関が見られない(図3: $R^2=5E-06$,

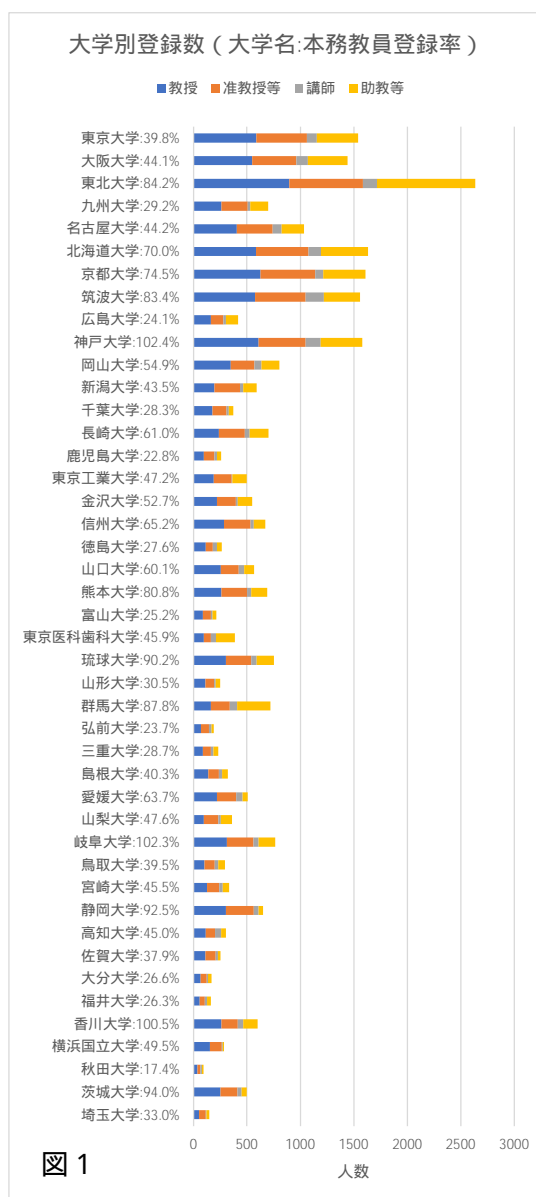


図1

1 一部はクローリング
 2 名誉教授・特任教員・客員教員等を含めていない
 3 公式ウェブサイト等から取得
 4 集計にはこの間の所属移動に伴う誤差がある

N=86)。実際、大学による登録傾向の差は大きい(香川大学の100%超から旭川医科大学の8.4%まで)ことが確認されている。

一方、このカバレッジと高い相関($R^2=0.76$)がある変数として、本務教員数に対する登録者総数の比 R が確認された。 R が高いほどカバレッジは向上する傾向がある。図4にアウトライヤ($R>1.5$)となる大学を除いた分布図を示す。カバレッジは概ね R を上限(登録者が全て本務教員の場合)として、平均的には R の75%程度であった。

以上の結果より本基盤の構築方法によって複数の情報源から研究者情報を統合して文献生産効率の分析を行うための情報基盤を整備することが可能となった。現時点では国立大学本務教員の約53%をカバーする規模の基盤となることが判明している。

なおこのカバレッジ評価は研究者等がシステムへ自主的に登録した内容に基づいている。その登録内容の正確さは第三者による検証を受けていない点には留意する必要がある。

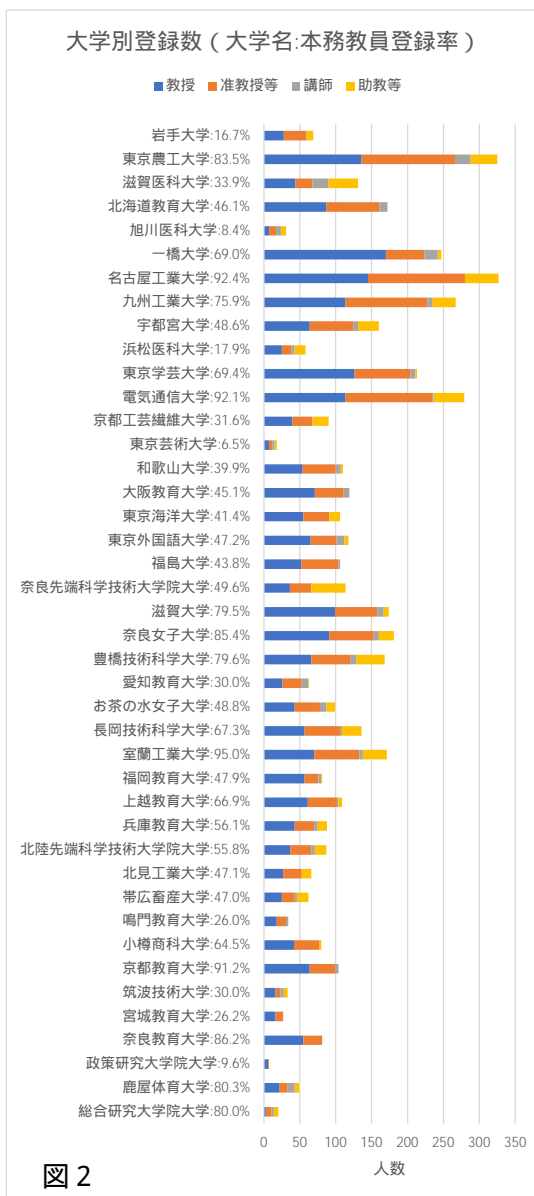


図2

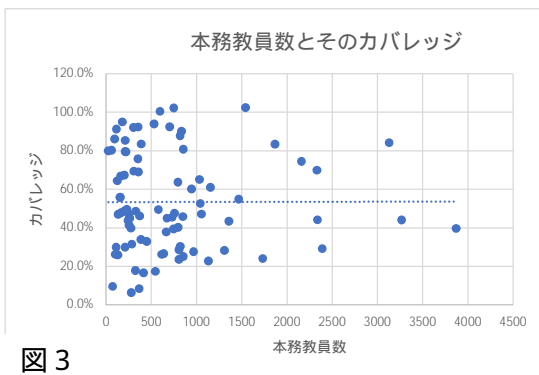


図3

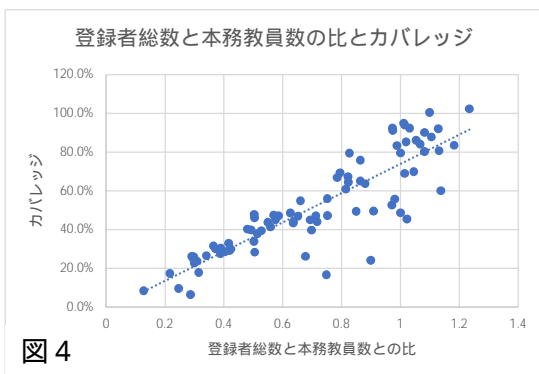


図4

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 中渡瀬秀一	4. 巻 27
2. 論文標題 新興境界領域研究の現況 - 科学研究費助成事業の分類を用いた分析 -	5. 発行年 2017年
3. 雑誌名 情報知識学会誌	6. 最初と最後の頁 1-5
掲載論文のDOI（デジタルオブジェクト識別子） https://doi.org/10.2964/jsik_2017_007	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計8件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 中渡瀬秀一
2. 発表標題 近年における分野間研究活動の分析
3. 学会等名 INFOPRO 2016
4. 発表年 2016年

1. 発表者名 中渡瀬秀一
2. 発表標題 プロジェクト助成成果としてのオープンデータ -成果報告に基づく調査結果-
3. 学会等名 INFOPRO 2017
4. 発表年 2017年

1. 発表者名 中渡瀬秀一
2. 発表標題 助成金プロジェクトから見る国内データ成果の現況
3. 学会等名 情報知識学会第22回情報知識学フォーラム
4. 発表年 2017年

1. 発表者名 中渡瀬秀一, et al.
2. 発表標題 共有データセットを用いた研究成果に関する時系列的分析
3. 学会等名 RA協議会第4回年次大会
4. 発表年 2018年

1. 発表者名 中渡瀬秀一, et al.
2. 発表標題 Yahoo! データセットを用いた研究成果に関する時系列的分析
3. 学会等名 IDRユーザフォーラム2018
4. 発表年 2018年

1. 発表者名 中渡瀬秀一, et al.
2. 発表標題 研究データ共有における研究成果および利用者の時系列的分析
3. 学会等名 第119回人文科学とコンピュータ研究会 (情報処理学会)
4. 発表年 2019年

1. 発表者名 中渡瀬秀一, et al.
2. 発表標題 データ引用による言語資源活用文献の把握の可能性: BCCWJの分析から
3. 学会等名 言語資源活用ワークショップ2019
4. 発表年 2019年

1. 発表者名 中渡瀬秀一
2. 発表標題 学術助成金による成果に対するメディアメンションの計量的分野分布比較
3. 学会等名 研究・イノベーション学会第34回年次学術大会
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

特集:第13回情報プロフェショナルシンポジウム ポスター発表を見て(情報の科学と技術67巻5号)
https://doi.org/10.18919/jkg.67.5_255
 特集: INFOPRO2017 ポスター発表参加記(情報の科学と技術 68巻 5号)
https://doi.org/10.18919/jkg.68.5_234
 新興境界領域研究の調査
<https://researchmap.jp/nakawtse/%E8%B3%87%E6%96%99%E5%85%AC%E9%96%8B>
 共有データセットを用いた研究成果に関する分析資料
<https://researchmap.jp/nakawtse/%E8%B3%87%E6%96%99%E5%85%AC%E9%96%8B>
 学術助成金成果に対するメディアメンション分析(資料)
<https://researchmap.jp/nakawtse/%E8%B3%87%E6%96%99%E5%85%AC%E9%96%8B>
 RA協議会年次大会(資料)
<https://researchmap.jp/nakawtse/%E8%B3%87%E6%96%99%E5%85%AC%E9%96%8B>

6. 研究組織		
氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考