

科学研究費助成事業 研究成果報告書

平成 29 年 8 月 9 日現在

機関番号：82108

研究種目：挑戦的萌芽研究

研究期間：2014～2016

課題番号：26540170

研究課題名(和文) 実験用マテリアル・リソースの引用にもとづく新しい研究機関影響度指標の開発

研究課題名(英文) Development of a novel index for organization impact based on reference of experimental materials

研究代表者

天野 晃 (Amano, Kou)

国立研究開発法人物質・材料研究機構・統合型材料開発・情報基盤部門材料データプラットフォームセンター・NIMSエンジニア

研究者番号：20622012

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：[背景]従来、引用分析における引用とは、研究論文のレファレンスをソースとしている。一方、引用分析の目的は研究主体の影響(関係)を分析することである。この意味で、現状の引用分析では多様な影響関係の一部しかとらえていない。

[目的]引用分析においてより多様な影響関係をとらえるための一歩として、生物実験材料の提供関係の分析を試みる。

[結果]Pubmed Centralの80万を超える論文テキスト検索より、提供機関名、生物実験材料名、被提供機関名の出現を抽出、整理し、3者の関係をデータセットとして整理した。これらの出現においてはpower-lawが見られた。

研究成果の概要(英文)：[Background]Present study of citation analysis is based on expressions of references on articles. But it can not assess the variety of inter-organization relationship of research impact such as use of facilities, donation of materials and other use of research resources. [Objective] We tried to assess the relationship among organizations based on the bioresource distributions as a basic study for the analysis of variety of impacts. [Result] We mined organization (distributor) names, bioresource names and affiliation (target organization) names for creating the "triple relationships" (organization=bioresource=affiliation) from PubMed Central. And we obtained 241 synonyms of organizations, 975 appearances of organization names, 1250 appearances of bioresource names, 3395 appearances of affiliation names and 4394 appearances of triple relationships. The power laws were observed in these appearances.

研究分野：計量書誌学

キーワード：研究機関評価 バイオリソース

1. 研究開始当初の背景

従来、引用分析における引用とは、文献の引用を指し、より具体的には、引用される対象が学術論文または図書である引用のことである。これらの引用は引用された側から引用する側への影響を示唆するもので、この引用-被引用のペアを計測することにより数値が得られる。これにより、雑誌、研究機関、研究者の間の影響関係を計量的に明らかにすることができる。一方で、このような影響関係は文献の引用にとどまらず、施設の供与、技術や材料の提供など多様な形態が考えられるが、現状ではこれらのうち、文献の引用だけが計量の対象となっている。

近年、Altmetrics など、書誌引用関係以外の影響関係を計量することによる評価が開発されつつある。これは、研究成果に対してより多角的な評価を与えようとする試みであるが、現状でも十分とは言えず、とくに上述のような論文本文からくみ取れる影響関係は注目されていない。

このような影響関係を調査するには、まとまった量の論文全文が必要となるが、近年は PubMed Central などの全文データセットが公開され、解析にも十分な量の論文が得られることが期待できる。

2. 研究の目的

すでに述べたように研究(機関)の影響関係を多角的に評価するには前述のような様々なモノや情報の供与・提供を計量する必要がある。当研究はその第一歩として、実験用材料の提供に注目し、これによる影響関係の把握を試みる。より具体的には、ナショナルバイオリソースプロジェクトにより提供される生物実験材料の提供を論文より把握し、各研究機関のアウトカム(書誌の引用にあたる)を計量することを試みる。

ところで、この報告では論文から確認できる提供をアウトカムとしているが実際にはより多くの実験用材料の提供(アウトプット)が行われていることに留意されたい。

3. 研究の方法

(1)対象

データソース

データソースとして 2014 年 7 月時点の PubMed Central の XML ファイルを利用した(843679 論文)。実験材料の利用に関しては[謝辞セクション]および[材料セクション]にほとんどが記述されると予想されるので、対象テキストはこれらのセクションに限った。この結果、[謝辞セクション]484633 件、[材料セクション]377540 件が得られた。すべての論文において、いずれかのセクションが存在した。

対象研究機関

2014 年時点のナショナルバイオリソースプロジェクト参画機関 30 機関を対象とした。

ただし、機関トップレベルを対象としたので 17 機関となる(表 1)。当該の 17 機関に対しては、その下位組織も含めたシノニムのリストを作成した。シノニムの作成には、科学技術・学術政策研究所の「大学・公的機関における研究開発に関するデータ」より SCOPUS における英組織名バリエーション (http://www.nistep.go.jp/wp/wp-content/uploads/name_variations_scopus_ver_2013_1.csv. (accessed: 2015-07-25).)、WoS における英組織名バリエーション (http://www.nistep.go.jp/wp/wp-content/uploads/name_variations_WoS_Ver2014_1.xlsx. (accessed: 2015-07-25).)を利用した。これによりシノニムのリスト 241 件を得た。

表 1: 対象研究機関

Chiba University
Hiroshima University
Kyoto University
Kyushu University
Nagoya University
National Institute for Environmental Studies
National Institute for Basic Biology
National Institute for Physiological Sciences
National Institute of Genetics
Okayama University
Osaka City University
RIKEN
The University of Tokyo
Tokyo Women's Medical University
University of Miyazaki
University of Tsukuba
Yamaguchi University

対象生物実験材料

最終的に 2015 年 7 月時点までに提供実績のあるナショナルバイオリソースプロジェクトの

(<http://rrc.nbrp.jp/referenceListAction.do> (accessed: 2015-07-25)) 生物実験材料 50602 件を対象とした。表 2 は出現(提供)回数が上位のものリストである。当該データはキュレーションが行われていない形で公開されており、複数種のデリミタの存在、シノニム記述法の不一致、タイプミスなどの問題を含む。そこで研究代表者らがマニュアルキュレーションを行った。おもな作業は以下の通りである。

- 生物実験材料名のセパレーション
デリミタが統一されていないため、カンマ、セミコロンをデリミタとみなし、生物実験材料名を分離した。
- 除外
極端に長い文字列など明らかに生物実験材料名でないものを対象外とした。

表 2：生物実験材料名(提供数上位)

材料名	別名	備考
NIES-2170		
MC3T3-E1	RCB1126	別名は請求番号
NA		対象外
RAW 264	RCB0535	別名は請求番号
IFM 40009		
NIES-19		
Wild C. int		種名略称
NIES-1		
Oryzabase		データベース名
NIES-298		
NIES-102		

(2)方法

対象データ(前述の[謝辞セクション]および[材料セクション])のセンテンスごとに、提供者と提供材料が明確なもののみを選択し、影響関係を構築したのちに、提供機関のアウトカムを測定する。

データソースの調整

前述したように対象データソースを全XMLでなく[謝辞セクション]および[材料セクション]とした。[謝辞セクション]については、XML タグが<ack>である要素を対象とした。[材料セクション]については、XML タグ内にターム"material"を含む要素を対象とした。検索対象の単位は要素内のセンテンスとした。

対象研究機関名の検索

機関名のシノニム 241 件を完全一致により検索した。一致箇所には開始・終了タグを挿入した。

生物実験材料名の検索

生物実験材料名 50602 件を完全一致により検索した。一致箇所には開始・終了タグを挿入した。

センテンスの抽出

各セクションにおいて対象となる機関名と実験材料名が同時出現するセンテンスを抽出したのち、前述のタグをもとにハンドサーチにより提供機関と提供材料が明確に確認できるものをさらに抽出した。最終的に 927 センテンスを得た。

著者所属機関の抽出

前述の全センテンスに対して、そのセンテンスが属する論文にある著者所属機関名を抽出した。著者機関名は XML の<aff>エレメントに記述される。

提供機関-実験材料-被提供機関関係の抽出

前述およびより、提供機関-実験材料-被提供機関の関係を得た。

4. 研究成果

(1)出現数

提供機関名、生物実験材料名、被提供機関名の出現数は、それぞれ、延べ 974(異なり:47、表 3)、延べ 1250(異なり:446、表 4)、延べ 3395(異なり:3082、表 5)であった。

表 3: 提供機関名(全機関、名寄せ済み)と提供数

提供機関名	提供数
RIKEN	406
The University of Tokyo	149
Kyoto University	126
National Institute of Genetics	125
Nagoya University	30
Kyushu University	26
Tokyo Women's Medical University	23
Chiba University	18
Hiroshima University	16
Okayama University	15
National Institute for Basic Biology	11
Osaka City University	9
University of Tsukuba	9
University of Miyazaki	4
Yamaguchi University	3
National Institute for Physiological Sciences	2
National Institute for Environmental Studies	2

表 4: 生物実験材料名(名寄せなし)と提供数(上位)

生物実験材料名	提供数
C57BL/6	67
Mouse	53
HeLa	41
E. coli	37
IL	36
Arabidopsis	36
Drosophila	25
BALB/c	23
C. elegans	19
medaka	17
Escherichia coli	17

表 5: 被提供機関名(名寄せなし)と提供数(上位)

被提供機関名	提供数
Baylor College of Medicine, United States of America	8
University of Washington, United States of America	6
University of Cambridge, United Kingdom	6
University of Massachusetts Medical School, United States of America	5

University of California Merced, United States of America	5
Stanford University School of Medicine, United States of America	5
University of Bremen, Germany	4
The Jackson Laboratory, US	4
Osaka University Graduate School of Medicine, Japan	4
Massachusetts General Hospital, United States of America	4
Division of Clinical Neuroscience, Chiba University Center for Forensic Mental Health, Chiba, Japan	4
Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki, Japan	4

(2) 組み合わせ出現数

提供機関名-生物実験材料名(表 6)、提供機関名-被提供機関名(表 7)、生物実験材料名-被提供機関名(表 8)のそれぞれの組み合わせによる出現数を計上した。計上では 1 論文中の組み合わせをすべてカウントした。たとえば、1 論文中に 1 機関から複数の実験材料提供を受けている場合、その実験材料数がカウントされる。また、提供機関名-生物実験材料名-被提供機関名の組み合わせを計上した(表 9)。

表 6: 提供機関名(名寄せ済み)-生物実験材料名(名寄せなし)(上位)

提供機関:生物実験材料名	件数
RIKEN:Arabidopsis	25
RIKEN:HeLa	25
RIKEN:Mouse	25
University of Tokyo:C57BL/6	21
RIKEN:C57BL/6	21
University of Tokyo:IL	20
National Institute of Genetics:Drosophila	18
National Institute of Genetics:E. coli	15
Kyoto University:Mouse	10
National Institute of Genetics:ASKA	9
RIKEN:201B7	9
RIKEN:HepG2	9

表 7: 提供機関名(名寄せ済み)-被提供機関名(名寄せなし)(上位)

提供機関名:被提供機関名	件数
RIKEN:Baylor College of Medicine	5
RIKEN:University of Bremen	4
RIKEN:The Jackson Laboratory	4
RIKEN:Division of Clinical Neuroscience, Chiba University	4
RIKEN:Cell Engineering Division, RIKEN BioResource Center	4

National Institute of Genetics:University of Washington	3
National Institute of Genetics:University of Texas MD Anderson Cancer Center	3
RIKEN:Centre d'Immunologie de Marseille-Luminy, CNRS-Inserm	3
RIKEN:Department of Physiology, Keio University School of Medicine	3
National Institute of Genetics:University of Massachusetts Medical School	3
RIKEN:University of Nevada School of Medicine	3
RIKEN:Department of Biosciences and Informatics, Keio University	3
RIKEN:University of California Merced	3

表 8: 生物実験材料名(名寄せなし)-被提供機関名(名寄せなし)(上位)

生物実験材料名:被提供機関名	件数
C57BL/6:The Jackson Laboratory	4
PC12:Division of Clinical Neuroscience, Chiba University Center for Forensic Mental Health	4

表 9: 提供機関名(名寄せ済み)-生物実験材料名(名寄せなし)-被提供機関名(名寄せなし)(上位)

提供機関:実験材料:被提供機関	件数
RIKEN:C57BL/6: The Jackson Laboratory	4
RIKEN:PC12:Division of Clinical Neuroscience, Chiba University Center for Forensic Mental Health	4

(3) 考察

提供機関ごとの提供数については理化学研究所が最多であり、これは当該機関が生物実験材料配布に特化した研究機構(バイオリソースセンター:BRC)を持つためと考えられる。提供される生物実験材料名においては、"Mouse"や"E. coli"など一般名が多く含まれ、株の特定が困難であることが伺える。提供機関名-生物実験材料名のペアの出現についても、生物実験材料名が一般名であることが多い。

機関の影響評価として用いることができる基本要素は、生物実験材料のレファレンス(数)、引用された年(記述のある論文の出版年)である。その他外部の知識基盤から利用可能なものとして、論文(または掲載誌)の分野、生物実験材料の公開年、提供機関における提供生物実験材料種別(カタログ)がある。はクラリベイトアナリティクス等が提供する文献データベースより、

は各提供機関の生物実験材料カタログより取得可能である。ただし、は一元的に電子媒体にまとめられておらず、入手と整

理は困難を極める。
本研究の成果からは、 が直接利用可能であり、ここから一種のアウトカムが計測できる。一方、アウトプットは にあたるもので、入手困難である。

一方で、検討課題として、インパクトファクターのような雑誌評価に用いられるアウトカム/アウトプット比を実験材料提供機関の評価に用いることが妥当であるかは、議論の余地がある。

本研究成果のまとめとしては、提供機関を明確に限定することにより、当初予定していた「生物実験材料」の定義、 全論文を対象とする調査、 生物実験材料名リスト作成、 提供機関名寄せ、 提供機関名-生物実験材料名ペアのリスト作成、 生物実験材料名の本文出現調査、 提供機関の影響評価、を達成したと言える。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1件)

[1]天野晃, 角田裕之, 石川大介. 実験用生物資源情報に基づく提供機関評価の試み. NISTEP NOTE, No. 19, p79-89. 2016. (査読なし)

〔学会発表〕(計 5件)

[1]天野晃, 石川大介, 角田裕之, 柴田大輔, 水野澄子. 論文マイニングのための生物系実験材料名辞書及び提供機関名辞書の作成. 情報メディア学会第 18 回研究会発表資料, 2016-11-26, 東邦大学大森キャンパス, 東京都大田区.

[2]天野晃, 石川大介, 角田裕之, 柴田大輔, 水野澄子. 実験用マテリアル・リソースの引用にもとづく新しい研究機関影響度指標の開発. 第 17 回図書館総合展, No.16, 2015-11-10-2015-11-12, パシフィコ横浜, 神奈川県横浜市.

[3]天野晃, 角田裕之, 石川大介. 医学生物学実験材料の提供情報による機関影響度の新機軸. 第 32 回医学情報サービス研究大会抄録集, p.19, 2015-07-18, 北海道大学, 北海道札幌市北区.

[4]天野晃, 角田裕之, 石川大介. 実験用生物資源情報に基づく提供機関評価の試み. NISTEP データ・情報基盤ワークショップ, 2015-02-04, 科学技術・学術政策研究所, 東京都千代田区.

[5]天野晃, 角田裕之, 石川大介. インパクト計量のための新しい要素:実験用生物資源. 情報メディア学会第 16 回研究会, 2014-11-22, 日本大学医学部板橋キャンパス, 東京都板橋区.

〔図書〕(計 0件)

〔産業財産権〕

出願状況(計 0件)

取得状況(計 0件)

〔その他〕

<https://github.com/kouamano/KAKEN-26540170>

<https://github.com/kouamano/XML0P>

6. 研究組織

(1)研究代表者

天野晃 (Amano, Kou)
国立研究開発法人物質・材料研究機構・統合型材料開発・情報基盤部門材料データプラットフォームセンター
NIMS エンジニア
研究者番号: 20622012

(2)研究分担者

角田裕之 (Tsunoda, Hiroyuki)
鶴見大学・文学部・教授
研究者番号: 30454961

(3)連携研究者

石川大介 (Ishikawa, Daisuke)
地方独立行政法人神奈川県立病院機構
主任主事
研究者番号: 70554335

(4)研究協力者

柴田大輔 (Shibata, Daisuke)
東邦大学 司書主任

水野澄子 (Mizuno, Sumiko)
国立研究開発法人理化学研究所