
J-STAGEとCiNii Booksを活用した 人文社会系の研究活動の可視化 — 実践と課題 —



横浜国立大学 研究推進機構
久保琢也 特任教員（講師） / URA
kubo-takuya-xv@ynu.ac.jp

2022年3月16日 JINSHA

久保 琢也 博士 (学術)



【経歴】

広島大学 URA (2016.5～)

- ・ 人社系の学術雑誌リストの作成

⇒ 第3回人文・社会科学系研究推進フォーラム (2017年3月3日)



信州大学 URA (2019.4～)

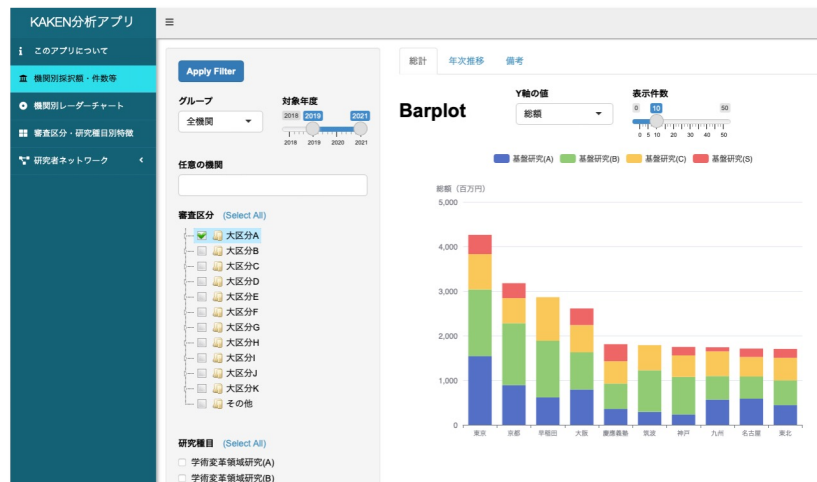
- ・ 文科省による評価のために、日本の学術雑誌に掲載された論文のメタデータを取得・整理

横浜国立大学 URA (2021.8～)

- ・ 大学図書館における人社系の専門書の所蔵状況 (科研費)

C4RAでの活動

- ❖ 研究力分析のためのR言語のコードを作成・共有
- ❖ R言語による分析ツール開発
- ❖ 研究力分析をテーマとしたR言語勉強会の開催
 - 2020年11月～2021年3月（個人的に）
 - 2021年12月～2022年2月（RA協議会テーマ別勉強会）



分析ツール



ホームページ

人社系の研究活動の評価/可視化は久保がURAとなった当初から課題として色んなところで議論されていたが、

① 利用しやすいインフラ（DB、分析ツール）の不在

⇒ 嘆いてばかりもいられない。ないならないうで工夫しなければ前進しない

② 人文社会系はどのような研究文化か？

⇒ 評価以前に人文社会系の研究活動に関する理解は十分か？

今日のお話

- ❖ **J-STAGE**を用いた論文のメタデータの取得と活用
 - J-STAGEからどうやって論文のメタデータを取得したか
 - 論文数について大学間のベンチマーク
 - 人社系の引用の特徴
- ❖ **CiNii Books**を用いた大学図書館における専門書の所蔵状況の調査
 - 所蔵する大学図書館数（所蔵館数）という指標の検討

※お断り※

- 研究結果をお示しするというより、取り組みの紹介をします。
- 手探りしつつ実施しているものですので、必ずしもベスト/ベターなアプローチではありません。

J-STAGEを用いた論文の メタデータの取得と活用

※詳細については主に以下を参照してください。

久保・伊藤（2021）「J-STAGEを活用した日本の学術論文データの整備」, 情報誌『大学評価とIR』, 第12号, 26-35.

伊藤・久保（2021）「多様な強みを把握するための研究力分析：AMEDfindとJ-STAGEを用いた試行」第7回RA協議会年次大会個人発表

J-STAGEとは

❖ J-STAGE (https://www.jstage.jst.go.jp)

- 国立研究開発法人科学技術振興機構 (JST) が運営する電子ジャーナルプラットフォーム
- 国内1500超の機関の、3000誌以上のジャーナルや会議録等の刊行物を収録
- 自然科学から人文・社会科学、さらに学際領域といった幅広い分野の資料を収録

The screenshot shows the J-STAGE homepage. At the top, there is a navigation bar with links for '資料・記事を探す', 'J-STAGEについて', 'ニュース&PR', and 'サポート'. A search bar is prominently displayed with the text 'J-STAGE上の記事を検索'. Below the search bar, statistics are shown: 5,348,870 記事 (with a lock icon and 5,087,765 記事 below it), 3,465 資料 (with a lock icon and 3,013 資料 below it), and 25 専門分野. A 'My J-STAGEにサインイン' button is visible. The '注目トピックス' section features three items: 'メンテナンス予定' (Maintenance Notice), 'セミナー開催のお知らせ' (Seminar Announcement), and 'J-STAGEニュース発行のお知らせ' (J-STAGE News Publication Notice). A '最新情報' section at the bottom indicates a date of 2022年2月10日.

The screenshot shows a specific article page on J-STAGE. The title is '情報の科学と技術'. The page includes a navigation bar with the same links as the homepage. On the right side, there is a 'PDFをダウンロード (1173K)' button. Below the title, there is a 'メタデータをダウンロード' section with options for 'RIS形式', 'BIB TEX形式', and 'テキスト'. The article title is '研究力分析の効率化・高度化に関するCode for Research Administrationの取り組み：URAによる機関を越えた連携'. The author is '平井 克之, 岡崎 麻紀子, 奥津 佐恵子, 久保 琢也, 矢吹 命大, 渡邊 優香'. The keywords are 'リサーチ・アドミニストレーター (URA), 研究力分析, プログラミング, コーディング, シビックテック'. The article is from 'ジャーナル フリー', volume 71, issue 2, page 80-86. The DOI is 'https://doi.org/10.18919/jkg.71.2.80'. There is a '詳細' (Details) link at the bottom.

❖ 大きな背景

- 国際学術論文データベースや分析ツールでは日本の学術雑誌に掲載された論文の多くが捕捉できない
- 文科省による運営費交付金の算定のための評価では人社系の論文数も正確に把握することが求められる

❖ CiNii ArticlesやJ-STAGE

- これらのDBに対応する分析ツールの不在
- 自分で分析しようにもメタデータの取得に制約や限界
 - CiNii Articles：ダウンロード可能だが、得られる情報が限定的
 - J-STAGE：ダウンロード機能なし

- 日本の学術論文のメタデータの取得・整理方法の検討
- 取得したデータを用いて探索的に人社系の研究活動の可視化

メタデータの取得

機械的にJ-STAGEの論文の詳細ページにアクセスして著者名や所属機関等の情報をスクレイピングする

Webブラウザで表示

HTML

The image shows a side-by-side comparison of a web browser view and its underlying HTML source code. The browser view (left) displays the article page for '情報の科学と技術' (Information Science and Technology), volume 71, issue 2, page 80-86. The article title is '研究力分析の効率化・高度化に関するCode for Research Administrationの取り組み：URAによる機関を越えた連携' (Efficiency and High-quality Research Activity Analysis: Collaboration across Institutions Enabled by URA). The author is listed as '久保 琢也' (Tetsuya Kubo). The browser view also shows a 'PDFをダウンロード' button and various download options (RIS, BibTeX, Text). The HTML view (right) shows the document's meta tags, including:

- `<meta name="authors" content="久保 琢也">`
- `<meta name="authors_institutions" content="信州大学学術研究・産学官連携推進機構">`
- `<meta name="authors_orcids" content="0000-0002-6219-3835">`



```
<meta name="authors" content="久保 琢也">
<meta name="authors_institutions" content="信州大学学術研究・産学官連携推進機構">
<meta name="authors_orcids" content="0000-0002-6219-3835">
```

※ 注意 ※

J-STAGEから機械的にデータを取得することは
利用規約上、禁止されている



JSTの担当者に相談



JSTと所属機関とで契約を締結

メタデータ取得の大まかな流れ

- ① 雑誌リストをスクレイピング（オプションル）
 - 現在は刊行されていない雑誌名も含まれる
- ② APIにより、ターゲットの論文群の書誌情報やURLを取得
 - 必要に応じて、雑誌や著者の所属機関を限定する
 - 大学名のバリエーションに注意（例：横浜国立大学、横国大、Yokohama National University等）
- ③ ②で取得した各論文のURLを使ってスクレイピング
 - プログラムは何語でも良い（詳しくは論文参照）
 - 1000件～2000件程度であれば、アクセス間隔を十分にとっても1日、2日でデータの取得は十分に可能
 - サーバーに過度な負荷を与えないように注意

論文によっては文献タイプ（論文、事例紹介、書評、等）があるので、取得しておくが便利

情報の科学と技術

資料トップ 巻号一覧 おすすめ記事 この資料について

J-STAGEトップ / 情報の科学と技術 / 71 巻 (2021) 2 号 / 書誌

事例報告 ← コレ

研究力分析の効率化・高度化に関するCode for Research Administrationの取組み：URAによる機関を越えた連携

平井 克之, 岡崎 麻紀子, 奥津 佐恵子, 久保 琢也, 矢吹 命大, 渡邊 優香

+ 著者情報

キーワード: リサーチ・アドミニストレーター (URA), 研究力分析, プログラミング, コーディング, シビックテック

- <body> … <div class = “global-article-subtitle”>

注：分野分類に関して

J-STAGEでの分野分類

哲学・宗教

文学・言語学・芸術学

人類学・史学・地理学

法学・政治学

経済学・経営学

社会学

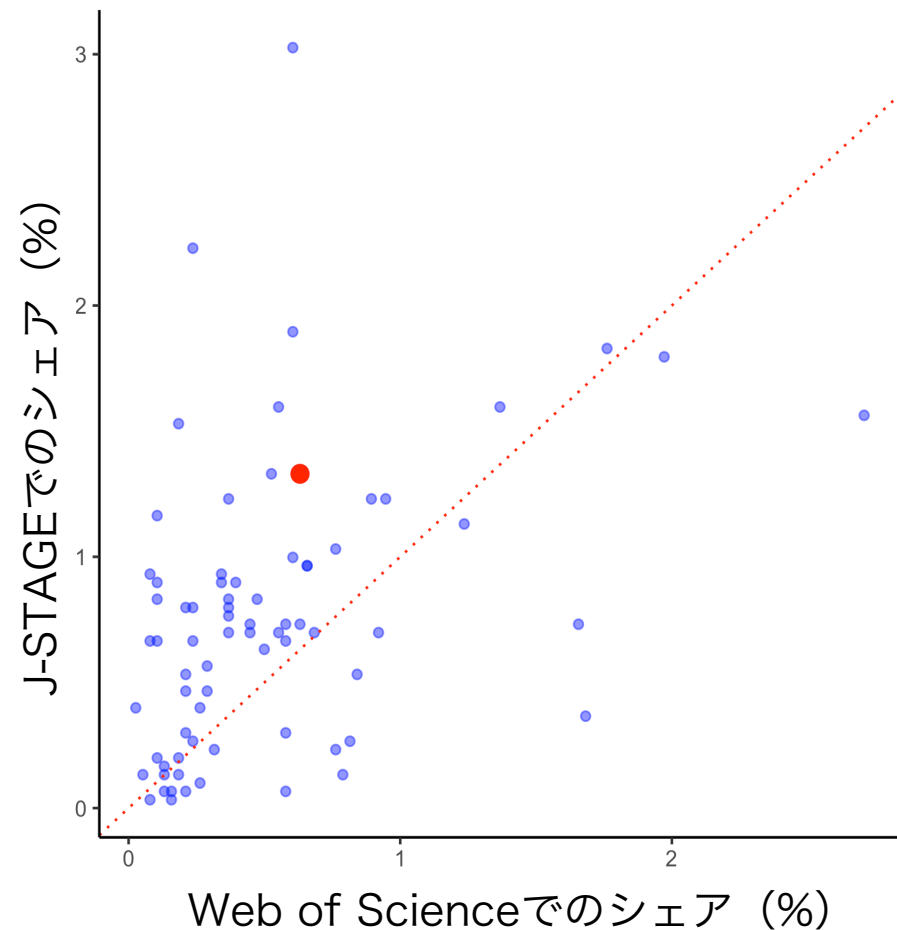
心理学・教育学

- ① 分野分類の括りが大きい
- ② 1つの雑誌が複数の分類にまたがる
- ③ パッと見て人社系が主ではないような雑誌が散見される



活用①大学間ベンチマーク

国立大学の人文社会系の文献数のシェア



J-STAGEでのシェアがWoSでのシェアに比べて優位な大学あり
(その逆もあり)

⇒ 1つのDBだけでは見過ごされる
人社系の研究活動の側面

年 : 2019年、対象 : 国立大学

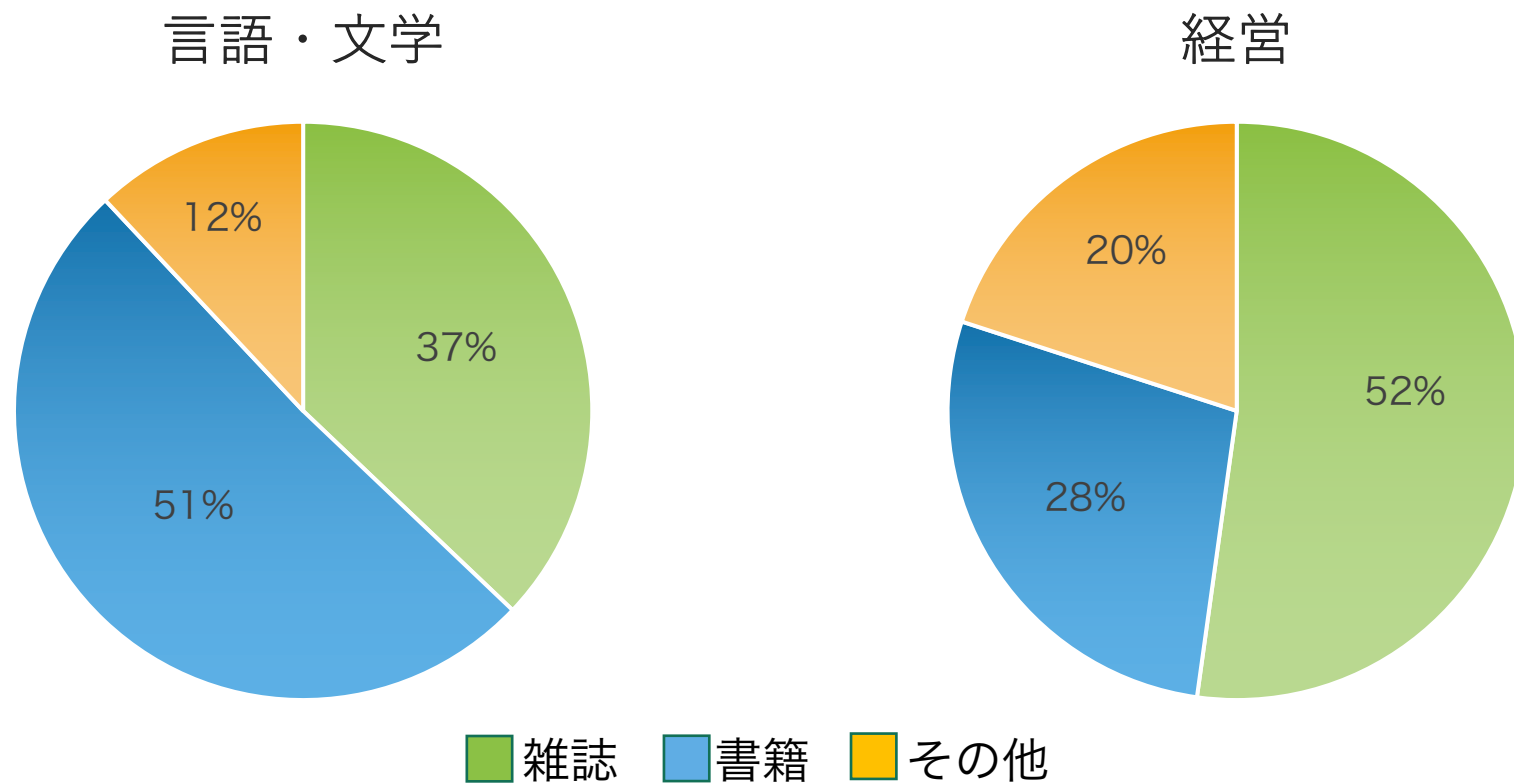
X 軸 : Web of Scienceの文献数 (整数カウント) のシェア
種別 : 全ての文献タイプ
分野 : 5 SOCIAL SCIENCES, 6 HUMANITIES

Y 軸 : J-STAGE上での文献数 (整数カウント) のシェア
※人社系の雑誌のうち、発行機関が学会名鑑上でも人社系に分類されるもの
※全ての文献種別が対象

伊藤・久保 (2021) 「多様な強みを把握するための研究力分析 : AMEDfindとJ-STAGEを用いた試行」 第7回RA協議会年次大会個人発表、より一部改変

※ 発表者が信州大学に在籍時に実施

活用②人社系の引用（暫定値）



- 学会名鑑の分類で、「言語・文学」「経営」に属する学協会の出版物に掲載される文献（2020年）が引用する文献の種別
- ランダムに5000件（言語・文学：3034件、経営：2041件、※重複あり）抽出

- ❖ J-STAGEは人社系の研究活動の可視化/評価において、Web of ScienceやScopusを補完する有用な情報源となり得る
- ❖ が、データ整理が超大変なので気軽にできない

① 文献タイプ

- 様々な文献タイプが混在するため、対象ではない文献は削除する必要がある（巻頭言、編集後記、etc.）
- 機械的に文献タイプを取得できない場合は1つ1つ確認するしかない

② 査読の有無

- 雑誌の査読の有無は機械的に判断することができるが、1つ1つの論文が査読を経たものであるかは、それぞれ確認しないとイケない

CiNii Booksを用いた大学 図書館における専門書の 所蔵状況の調査

※本取り組みは現在、JSPS科研費21K18015の助成を受けて実施しています。

※主には以下の内容のお話をします。

久保琢也（2020）「大学図書館における人文社会系の専門書の所蔵状況 —CiNii BooksとopenBDを用いた予備的調査—」 『第9回大学情報・機関調査研究集会 論文集』

CiNii Booksとは

❖ CiNii Books (<https://ci.nii.ac.jp/books/?l=ja>)

- 全国の大学図書館等が所蔵する本（図書や雑誌等）の情報を検索できるサービス
- 国立情報学研究所が運用する全国の大学図書館等約1300館が所蔵する約1,300万件（のべ1億冊以上）の本の情報を網羅

The screenshot displays the CiNii Books search interface. At the top, the logo 'CiNii Books' is accompanied by the tagline '大学図書館の本をさがす' (Find books in university libraries). Below the logo, there are three tabs: '図書・雑誌検索' (Book/Journal Search), '著者検索' (Author Search), and '内容検索' (Content Search). A search input field contains the text 'フリーワード' (Free word) and a '検索' (Search) button. Below the search field, there are three buttons: 'すべての資料' (All materials), '図書' (Books), and '雑誌' (Journals), along with a '閉じる' (Close) button. The main search area contains several input fields and checkboxes: 'タイトル' (Title) with checkboxes for 'タイトル完全一致' (Title exact match) and '統一タイトルを含む' (Include unified title); '著者名' (Author name) with a checkbox for '別名を含む' (Include aliases); '著者ID' (Author ID), '統一タイトルID' (Unified title ID), and '出版者' (Publisher); 'ISBN', 'ISSN', and 'NCID'; '件名' (Subject), '分類' (Classification), and '注記' (Notes); '資料種別' (Material type) with a dropdown menu set to 'すべての資料種別' (All material types); '言語種別' (Language type) with a dropdown menu set to 'すべての言語' (All languages) and a 'コード' (Code) field; '出版年' (Publication year) with '年' (Year) and '年まで' (Year up to) fields; '図書館ID' (Library ID), '機関ID' (Institution ID), and '地域' (Region) with a dropdown menu set to 'すべての地域' (All regions). At the bottom, there are two checkboxes: '図書館ID・機関ID・地域を記憶する' (Remember library ID, institution ID, region) and 'ローマ字をカナに変換' (Convert Roman letters to Kana).

❖ 図書の評価 / 可視化について

- 人社系において図書は重要であると言われているが、日本において図書に関する取り組みはほとんど行われていない。

❖ 所蔵されている図書館の数（**所蔵館数**）

- 欧米の研究者により、人文社会系の図書に関する評価指標として提案される（Torres-Salinas & Moed, 2009; White et al., 2009）
- Altmetricsとして実用化されている例がある（PlumX; Plum Analytics, 2020）

日本の人文社会系の図書（専門書）の可視化への活用は可能か？

❖ 所蔵館数に関する先行研究の概観

- 所蔵館数の解釈について議論はあるが、1つの解釈としては「**利用可能性**」 (Zuccala & Guns, 2013; Torres-Salinas & Arroyo-Machado, 2020)
- そのインパクトは対象とする図書館群によって変わる (Torres-Salinas & Arroyo-Machado, 2020)
- 所蔵館数と引用数とには一部、正の相関が観察されている (Linmans, 2010; Zuccala & Guns, 2013)

その一方で、所蔵館数の多寡を論じるためには、「所蔵館数」という指標がそもそもどのような特徴を有しているかを理解する必要がある

なぜ専門書か？

- ❖ 第一の関心の対象はいわゆる「**学術図書**」
 - しかし、学術図書を客観的に判断するのは現時点で困難
- ❖ 次善の策として「**専門書**」に対象を狭める
 - Cコードにおける販売対象が「**専門**」である図書

Cコード：書籍出版物の分類基準。販売管理や書店が陳列場所を決める際のガイドラインとして用いられる



調査に用いたデータベース

❖ CiNii Books

- 2015年～2019年の日本語の図書（ISBN 224408件）
- 各図書が所蔵される図書館数（所蔵館数）も併せて取得

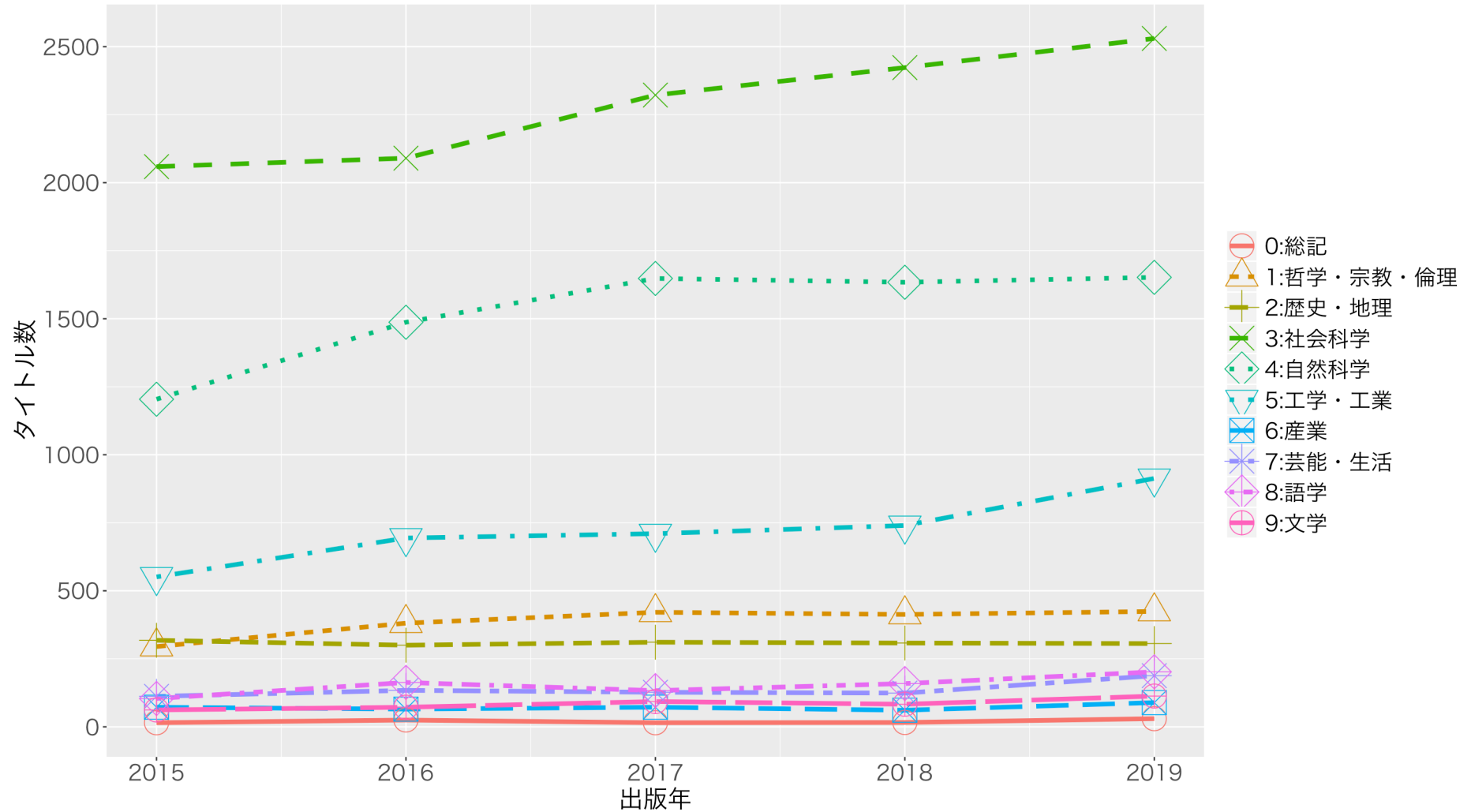


❖ openBD (<https://openbd.jp>)

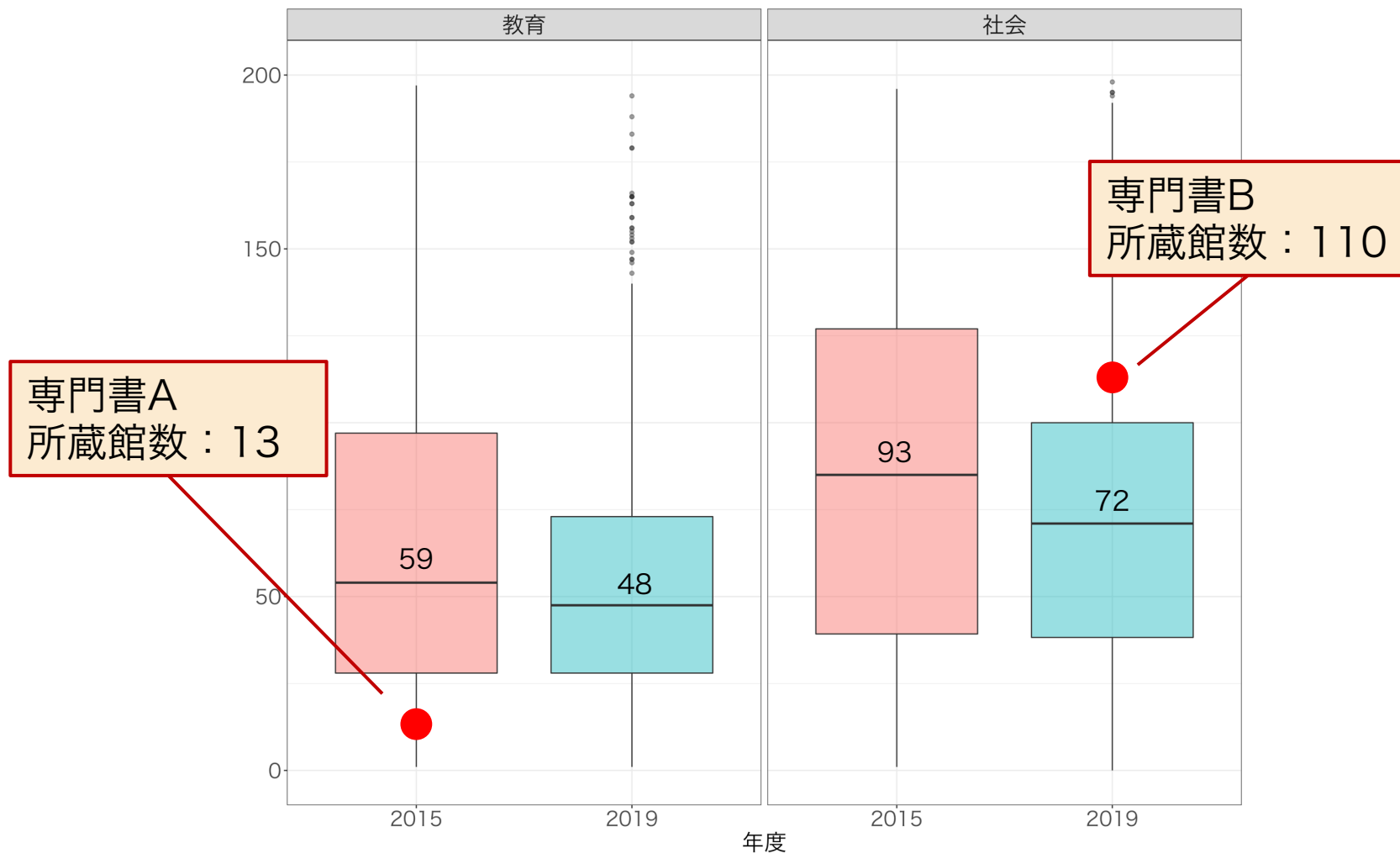
- カーリルと版元ドットコムにより開発されたAPIシステム
- 114543件（ISBN）の書籍に対してCコードを取得
 - ⇒ 専門書は28270件
 - ⇒ 人社系は16653件

※両データベースのAPIを利用するにあたり、管理機関（法人）より許諾を得ている

分野別タイトル数（全体）



所蔵館数の分布（一部：教育 / 社会）



所蔵館数が同じであっても、分野や出版年により位置付けが異なる

- ❖ 所蔵館数という値を解釈するためには少なくとも分野や出版年を考慮する必要があるそう
 - その他に注意すべき観点は何か？
- ❖ 次のステップとしてもう少し「学術」に寄せたい
 - 専門書から学術図書への絞り込み
 - 編著者の所属機関情報が網羅されていれば、少なくとも大学関係者だけを抽出できるのに…
 - 所蔵館数と引用数との関係を見る予定
- ❖ CiNii Booksに登録されている図書とは？
 - 大学によって登録される図書の扱いは異なるらしい
 - 図書館の管理外の図書は補足できない点は留意する必要がある (e.g., 研究者が消耗品として購入)

終わりに

- ❖ 既存のリソースをうまく活用することで、これまでとは違った人文社会系の世界が見えてくるかもしれない
- ❖ その一方、現状では手作業がかなり発生するので、作業効率化が課題
 - チームを組んでやる（作業の分散）
 - 賢く機械的にできないものか（自然言語処理？）
- ❖ 根本的にはやはり、インフラを整えない限り実用化には程遠い

❖ FAQ :

Q. お上がなんとかすれば良いのではないか？

A. 人任せでは何も進展しないので、まずはできることから始めましょう

- ❖ 本取り組みを実施するにあたり、科学技術振興機構、国立情報学研究所の皆様にごこの場を借りて御礼申し上げます。
- ❖ 本取り組み/研究の一部はJSPS科研費21K18015の助成を受けて実施しています。

参考文献

- 久保琢也 (2020) 「大学図書館における人文社会系の専門書の所蔵状況 —CiNii BooksとopenBDを用いた予備的調査—」 『第9回大学情報・機関調査研究集会 論文集』
- 久保・伊藤 (2021) 「J-STAGEを活用した日本の学術論文データの整備」, 情報誌『大学評価とIR』, 第12号, 26-35.
- 伊藤・久保 (2021) 「多様な強みを把握するための研究力分析: AMEDfindとJ-STAGEを用いた試行」 第7回RA協議会年次大会個人発表
- Linmans, A. J. M. (2010). Why with bibliometrics the Humanities does not need to be the weakest link Indicators for research evaluation based on citations, library holdings, and productivity measures. *Scientometrics*, 83(2), 337-354.
- Plum Analytics. (2020). Usage metrics. (<https://plumanalytics.com/learn/about-metrics/usage-metrics/>)
- Torres-Salinas, D., & Arroyo-Machado, W. (2020). Library Catalog Analysis and Library Holdings Counts: origins, methodological issues and application to the field of Informetrics. *Evaluative Informetrics: The Art of Metrics-Based Research Assessment*, pp.287-308.
- Torres-Salinas, D., & Moed, H. F. (2009). Library Catalog Analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in Economics. *Journal of Informetrics*, 3(1), 9-26.
- White, H. D., Boell, S. K., Yu, H., Davis, M., Wilson, C. S., & Cole, F. T. H. (2009). Libcitation: A measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology*, 60(6), 1083-1096.
- Zuccala, A. & Guns, R. (2013). Comparing book citations in humanities journals to library holdings: Scholarly use versus 'perceived cultural benefit. *Proceedings of the 14th International Society for Scientometrics and Informetrics Conference*, 1, 353-360.

ご清聴誠にありがとうございました。

ご質問やコメントは以下のアドレスか
facebookからお知らせください。

kubo-takuya-xv@ynu.ac.jp