学術情報検索における検索語を用いた論文の特性分析

小林　和央† 　風間　一洋†† 　吉田　光男††† 大向　一輝†††† 佐藤　翔†††††
桂井麻里衣†††††

† 和歌山大学大学院システム工学研究科　〒 640–8510 和歌山県和歌山市栄谷 930
†† 和歌山大学システム工学部　〒 640–8510 和歌山県和歌山市栄谷 930
††† 豊橋技術科学大学情報・知能工学系　〒 441–8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1
†††† 東京大学大学院人文社会系研究科　〒 113–0033 東京都文京区本郷 7-3-1
††††† 同志社大学免許資格課程センター　〒 602–8580 京都府京都市上京区今出川通烏丸東入
†††††† 同志社大学理工学部　〒 610–0394 京都府京田辺市多々羅都谷 1-3
E-mail: †s206099@wakayama-u.ac.jp

あらまし　学術論文は研究者以外の多彩なユーザに利用されるようになり，CiNii Articles などの検索サービスでの閲覧以外に，ソーシャルメディア上では，論文の内容の面白さや実世界の影響を受けて共有されている．本稿では，閲覧した論文を探すために用いた検索語を使用して，閲覧・言及行動の意図や目的の違いやどのような影響を受けて論文が読まれているのかを明らかにする．実際に，CiNii Articles の利用履歴とソーシャルメディア上の論文の言及データを用いて，ユーザの閲覧・言及行動を，何人が閲覧・言及したかで重要度を示す閲覧・言及人数と，ある論文を閲覧・言及したユーザが他にどのくらい論文を閲覧・言及したかで内容の普遍性を示す閲覧・言及普遍度の 4 つの指標を求める．さらに，検索クエリから論文の内容や分野などの分類を表す 2 種類の検索語を抽出する．最後に，論文集合や検索結果を閲覧・言及人数，普遍度で順位付けし，検索語を用いて各指標の特性の違いを分析する．

キーワード　CiNii Articles，検索語，2 部グラフ，特性分析，オルトメトリクス

1 はじめに

研究者にとって，論文を読むことは重要なタスクである．かつては論文や検索手段は紙媒体が一般的であったが，論文とその掲載誌の指数的な増大に伴い，インターネットの普及と論文の電子化が進み，CiNii Articles（以下，本文中では CiNii と呼ぶ）のような学術論文検索サービスが日常的な学習や研究活動に重要な役割を果たすようになった．同時に，研究者だけでなく，学生，開発者などの多彩なユーザに使われるようになったが，学術情報検索では時間順や被引用数順など順位付け方法が限られており，すべてのユーザにとって必ずしも使いやすいとは限らない．

論文等の評価指標として，計量書誌学の分野では，従来では論文の被引用数が用いられてきたが，引用数には評価遅延があり，一般的に論文公開から引用まで 2〜3 年掛かると言われている [1]．また近年は，さらに文献の閲覧数，ブログを含むソーシャルメディアでの言及，マスメディアの報道など，研究者以外の関与を含めた社会的な影響を示す様々な視点を組み入れることで，文献が社会に及ぼした影響度を包括的かつ早期に計測することを目指す指標であるオルトメトリクス（Altmetrics）が提案されている [2]．オルトメトリクスを提供するサービスには Altmetric[1] や Impactstory[2] があり，日本の論文を主対象とするサービスには Ceek.jp Altmetrics [3] がある．ただし，単純に被引用数の代わりにオルトメトリクスを使えばよい問題ではなく，Priem らは，オルトメトリクスと被引用数の相関が弱く，オルトメトリクスは被引用数とは異なるユーザのインパクトを反映していると述べている [2]．また，Mohammadi らは，学術目的で Twitter を利用しているユーザ 1,912 人に対しアンケート調査を行い [4]，そのうち 45%は非学術系の職業のユーザで，59%が社会科学や人文学系のユーザであったと述べている．つまり，オルトメトリクスは従来の被引用数では評価できなかった側面を補完するために，組み合わせて利用することが重要である．そこで，学術情報検索の利用履歴から，学術出版物をどのような意図や目的で利用しているかを明らかにすると共に，その知見に基づいて特定少数の研究者や一般読者のようなユーザの特性に合った指標や特徴量を順位付けに用いる必要がある．

本稿では，学術情報検索サービスのユーザの閲覧行動とソーシャルメディア上での言及行動を，何人が閲覧・言及したかで重要度を示す閲覧・言及人数と，ある論文を閲覧・言及したユーザが他にどのくらい論文を閲覧・言及したかで内容の普遍性を示す閲覧・言及普遍度の各指標と，閲覧した論文を探すために用いた検索語を使用して，各指標の特性の違いや，実世界でどのような影響を受けて論文が読まれているのかを分析する．

1：https://www.altmetric.com
2：https://impactstory.org

2 関 連 研 究

2.1 学術情報検索サービスの分析

学術情報検索サービスのユーザに関する特徴分析として，佐藤らは，国立国会図書館サーチ（NDL サーチ）のアクセスログに基づき，利用者の簡易検索，詳細検索，ファセット検索の利用状況を分析した．NDL サーチのファセット検索では，本や記事・論文などの「資料種別」や「データベース」「所蔵館」「出版年」「分類」「分野」「国・地域」の 7 観点や特徴語による絞り込みが可能で，多くのユーザは簡易検索を行った後，ファセット検索等の絞り込み機能を利用していると述べている [5]．

また，学術情報の評価指標の多様化に関して，Hristakeva らは論文出版数から研究経験を 5 段階（Unpublished，Postgraduate，Postdoc，Lecturer，Professor）に分類して 4 種類の論文推薦システムで評価し，研究経験によって異なる推薦方法が良いことを示した [6]．また，Thelwall は文献利用者を 12 属性（Student，Lecturer， Professor，Librarian など）に分類し，Mendeley のユーザが保存した文献の種類を調査して，属性ごとに異なる傾向を持つことを示した [7]．つまり，ユーザの経歴や現在の仕事などの属性の違いによって，異なる指標を使い分ける必要がある．

2.2 オルトメトリクスの分析

オルトメトリクスを扱った分析には以下のような研究が存在し，被引用数との違いやオルトメトリクスの付与状況などが分析されている．Priem らは，オルトメトリクスと被引用数の相関が弱く，オルトメトリクスは被引用数とは異なるユーザのインパクトを反映していると述べている [2]．吉田らは，日本の学協会が発行する学術雑誌に掲載された論文を対象に，ソーシャルメディア上での言及数の分布や，論文の出版年，記述言語や属する分野等と言及数の関係を分析しており [8]，日本の学協会誌掲載論文約 110 万件に対してソーシャルメディア上での言及が存在する論文は約 1%と少なく，またソーシャルメディアで言及される論文には分野による有意差が部分的に存在し，Ceek.jp Altmetrics と Altmetric には共通して，人文社会系の論文で被言及数が多く，理工系の論文で被言及数が少ないという有意差があると述べている．佐藤らは，Twitter からの言及数が多い論文と 0 回の論文集合について，非専門家にとっての論文タイトルの面白さを 7 段階で得点化し，言及数との相関を分析した [9]．その結果，得点と言及数には有意な相関は見られなかったが，言及数が多い論文集合のほうが 0 回の集合よりも得点が有意に高かったと述べている．しかし，佐藤らの研究は言及数とタイトルの面白さの相関分析にとどまっており，ユーザの実際の意図は検索語などの別のデータを用いて検証する必要がある．

3 論文の評価指標

3.1 ユーザ–論文の 2 部グラフ

論文の重要度を，内容ではなく，オルトメトリクスで用いられるような論文がユーザに与える社会的影響から求めるために，論文 $p_i (i = 1, \ldots, M)$ とユーザ $u_j (j = 1, \ldots, N)$ の **2 部グラフ**（bipartite graph）G としてモデル化し，行が論文，列がユーザを表す M 行 N 列の**接続行列**（incidence matrix）$R_{p\times u}$ で表す．なお，関係があれば $r_{i,j} = 1$，なければ $r_{i,j} = 0$ である．

論文間の関係の有無は，G 上で 2 ホップで p_i から p_j に到達可能かに相当し，接続行列 $R_{p\times u}$ から M 次の**隣接行列**（adjacency matrix）$R_p = R_{p\times u} \times R_{p\times u}^{\top}$ として求める．$R_{p\times u}^{\top}$ は $R_{p\times u}$ の転置行列で，R_p の要素 $r'_{i,j}$ は p_i から p_j に到達可能なら $r'_{i,j} > 0$，不可なら $r'_{i,j} = 0$ である．

同様に，ユーザ間の関係の有無は，接続行列 $R_{p\times u}$ から N 次の隣接行列 $R_u = R_{p\times u}^{\top} \times R_{p\times u}$ で表す．R_u の要素 $r''_{i,j}$ は u_i から u_j に到達可能なら $r''_{i,j} > 0$，不可なら $r''_{i,j} = 0$ である．

3.2 次　　数

論文 $p_i (i = 1, \ldots, M)$ の次数を，接続行列 $R_{p\times u}$ から $deg(p_i) = \sum_{j=1}^{N} r_{i,j}$ で計算する．本稿では，2 部グラフとして，CiNii の利用履歴から求めたユーザと閲覧した論文の関係と，Twitter などのソーシャルメディアにおける論文に関する発言から求めたユーザと言及した論文の関係を用いる．以降では，次数 $deg(p_i)$ をより一般的に，前者では**閲覧人数**，後者では**言及人数**と呼ぶ．

3.3 普　遍　度

論文の閲覧と言及は，それぞれ異なる目的やきっかけで行われる．例えば，論文の閲覧は興味がある研究の検索や論文の引用がきっかけになることが多く，言及は論文が示す知見の拡散を目的とするような違いがある．さらに，論文の閲覧人数が同じでも，必ずしもその内容の特性も同じとは限らない．例えば，最新技術やトップデータを提案する論文以外にも，初学者向けの易しい解説論文や，一般人の興味も惹くような一般的な論文もあり，ユーザによって求める論文の特性は異なるはずである．

そこで，論文の特性は閲覧ユーザ集合から推定できると考えて，ある論文に 2 部グラフ上でエッジで接続しているユーザ群に，さらにエッジで接続している論文の和集合が全体に占める割合を求めることで論文の普遍性を定量化する**普遍度**（general degree）[10] を用いる．論文 p_i の普遍度 $S(p_i)$ は，G 上で 2 ホップで到達可能な（R_p で $r'_{i,j} > 0$）論文数を全論文数 M で割って求める．

例えば，論文が特定の分野のユーザにしか閲覧・言及されない場合は和集合が小さいので，閲覧・言及人数が多くても本指標は小さく，逆に異なる分野や属性のユーザに広く閲覧・言及される場合は和集合は大きくなり，本指標は大きくなる．ただし，この普遍性は，対象とするユーザ集合に大きく影響されることに注意する必要がある．

4 論文の特性の分析手法

4.1 検索語による特性分析手法

学術情報検索サービスとして，CiNii を対象に，ユーザが論文の検索に用いた検索語を使用して分析する．例えば，ユーザ

の検索行動は，初めは 1 語で検索しても検索結果が多いため，2 回目はそれを絞り込むために検索語を追加することが多い．このような性質を利用することで，その論文を検索した意図や目的を判別できると考えられる．

実際に，CiNii のアクセスログから，アクセス先の URL に含まれる NAID（NII 論文 ID）とリファラを手がかりに，論文のトピックを示す**内容指定語**と，論文の分野や分類を示す**分類指定語**を以下の手順で抽出する．

（1） リファラ中の検索エンジンの URL からクエリを抽出する．
（2） 英大文字を英小文字に変換し，mecab-ipadic-NEologd の方式 [3]で正規化する．
（3） MeCab で形態素解析して，2 文字以上の名詞を抽出する．既存の IPA 辞書では認識できない固有表現に対応するため，mecab-ipadic-NEologd を利用する．
（4） Oracle Text で提供されるストップワードリスト [4]に含まれる英単語を除去する．

内容指定語は，文書 $p_i(i = 1, \ldots, N)$ の検索語 t_j の使用頻度 $qf(i,j)$ と使用文書数 $df(t_j)$ から，式 (1) のように $qf\text{-}idf(i,j)$ で重み付けすることで，特定の文書に対して使用される検索語を抽出する．

$$qf\text{-}idf(i,j) = \frac{qf(i,j)}{s(p_i)} \log \frac{N}{df(t_j)} \tag{1}$$

ここで，$s(p_i)$ は文書 p_i の全検索語の使用頻度の和である．

また，分類指定語は式 (2) のように $qf\text{-}df(i,j)$ で重み付けし，多くの文書に対して使用される検索語を抽出する．このような単語は一般的な文書中では利用できないことが多いが，情報検索においては通常，必要な単語のみを入力するため，利用できると考えられる．

$$qf\text{-}df(i,j) = \log(qf(i,j) + 1)df(t_j) \tag{2}$$

4.2 論文のトピックの可視化手法

本稿では，検索結果を閲覧人数や閲覧普遍度で順位付けした結果から，論文のトピックの傾向の違いを比較可能にするために，Mikolov らが提案した word2vec [11] の CBOW モデルを用いて単語の分散表現を作成する．実際に，CiNii の論文の書誌情報 37,420,188 件から，引用情報程度の簡略な情報しかない場合が多い，収録データベース [5]が CJP 引用 [6]のみのデータを除いた 20,934,768 件の各論文のタイトルと概要の組み合わせを 1 文章とし，専門用語コーパスを作成する．この専門用語コーパスを mecab-ipadic-NEologd で形態素解析して単語に分割し，200 次元のベクトルで表される単語の分散表現を作成する．さらに，t-SNE [12] を用いて高次元データを 2 次元に次元圧縮し，可視化可能にする．

3 : https://github.com/neologd/mecab-ipadic-neologd/wiki/Regexp.ja
4 : https://docs.oracle.com/cd/E16338_01/text.112/b61357/astopsup.htm#i634475
5 : https://support.nii.ac.jp/ja/cia/cinii_db
6 : https://www.nii.ac.jp/hrd/HTML/OpenHouse/h16/archive/PDF/701.pdf

4.3 バースト性を持つ検索語の抽出

本稿では，論文の検索語のバースト性・定常性より，実世界の影響を受けて読まれているかを判別する．例えば，学術情報検索においては，定常的に行われている研究目的の論文の検索以外にも，ソーシャルメディアの一般的なユーザの間でも話題になった論文や，実世界のイベントに関係する人物など，情報の拡散性に影響を受けて検索することがあると考えられる．

そこで，検索語の使用頻度の時系列データを，対象期間を一定の間隔に区切り，各期間の全文書数と関連文書数から，Kleinberg の列挙型バースト検知アルゴリズム [13] を用いて検索語のバーストを判定する．

まず文書を，対象単語 w を含む関連文書と含まない非関連文書に分類する．解析期間において m 個の文書集合 $A_1, A_2, ..., A_m$ が離散時間に送られてくる状況を考える．解析期間 $t_1, ..., t_N$ において，t 日目の文書集合 A_t に含まれる文書数を d_t，t 日目の文書集合 A_t に対象単語 w を含む関連文書数を r_t, 解析期間における全文書数を $D = \sum_{t=1}^{N} d_t$，解析期間に単語 w を含む文書数の総数を $R(w) = \sum_{t=1}^{N} r_t$ とする．

Kleinberg の列挙型バースト検知アルゴリズムでは，q_0 を非バースト状態，q_1 をバースト状態とする，2 状態のオートマトンをモデルとして定義している．非バースト状態 q_0 に解析期間全体の単語 w の出現確率 $p_0 = R/D$，バースト状態 q_1 に p_0 にパラメータ s を掛けた値 $p_1 = sp_0$ とする．ただし，s は，$s > 1$ かつ $p_1 \leqq 1$ を満たす値とする．なお，この s の値が小さいほど，文書集合中の関連文書の割合が低い場合でもバーストとみなされやすくなる．

状態遷移は d_t と r_t を入力とすることで決定される．状態系列 $\mathbf{q} = (q_{i_1}, \ldots, q_{i_N})$ と表すことができ，q_{i_t} は t 日目の文書集合によって決定された状態 $q_i(i = 0, 1)$ である．2 状態オートマトンにて，非バースト状態の q_0 とバースト状態の q_1 であることに対するコストを計算する．対象単語 w を含む文書が二項分布 $B(d_t, p_i)$ に従って出現すると仮定すると，状態 $q_i(i = 0, 1)$ であることに対するコスト関数 $\sigma(i, r_t, d_t)$ は次式で定義できる．

$$\begin{aligned}\sigma(i, r_t, d_t) &= -\ln\left[B(d_t, p_i)\right] \\ &= -\ln\left[\begin{pmatrix} d_t \\ r_t \end{pmatrix} p_i^{r_t}(1 - p_i)^{d_t - r_t}\right]\end{aligned} \tag{3}$$

ここで，$\begin{pmatrix} d_t \\ r_t \end{pmatrix} = C(d_t, r_t) = \frac{d_t!}{r_t!(d_t - r_t)!}$ である．つまり，t 日目の文書数 d_t から対象単語 w を含む関連文書数 r_t を選ぶ組み合わせである 2 項係数を表す．この関数は，入力 r_t と d_t，$q_i(i = 0, 1)$ によってコストが決まる．$p_1 > p_0$ であり，t 日目の関連文書の出現確率 r_t/d_t が p_1 に近ければ $\sigma(1, r_t, d_t) < \sigma(0, r_t, d_t)$ となり，バースト状態 q_1 が選ばれる．逆に，r_t/d_t が p_0 に近ければ $\sigma(1, r_t, d_t) > \sigma(0, r_t, d_t)$ となり，非バースト状態 q_0 が選ばれる．

ただし，頻繁にバースト状態と非バースト状態が切り替わらないように，状態 q_i から q_j への状態遷移を妨げる関数 $\tau(i, j)$

を次のように定義する．

$$\tau(i,j) = \begin{cases} (j-i)\gamma & (j > i) \\ 0 & (j \leqq i) \end{cases} \quad (4)$$

τ はパラメータ γ によって調整されるが，特に理由がない場合，デフォルトの値である $\gamma = 1$ とする．

最後に，状態 $q_i(i = 0, 1)$ であることに対するコスト関数 $\sigma(i, r_t, d_t)$ と頻繁にバースト状態と非バースト状態の状態遷移が起こらないようにする $\tau(i,j)$ より，次のコスト関数を最小とする状態系列 $\mathbf{q}$ を求める．

$$c(\mathbf{q}|r_t, d_t) = \sum_{t=0}^{N-1} \tau(i_t, i_{t+1}) + \sum_{t=1}^{N} \sigma(i_t, r_t, d_t) \quad (5)$$

このコスト関数を最小とする状態系列が，対象単語 w におけるバーストの状態を表す状態系列として選択される．

5 データセット

5.1 CiNii Articles の利用ログ

本稿では，CiNii の 2014 年 4 月 1 日〜2016 年 3 月 31 日の 2 年間の Apache Web サーバのアクセスログを用いた．ユーザが閲覧した論文（閲覧論文）の識別子として，アクセス先の URL に含まれる NAID（NII 論文 ID）を用いた．閲覧したユーザの識別子として，IP アドレスとユーザエージェントの組 [14] を用いた．一般的に HTTP Cookie を利用してユーザ識別子を割り振ることが多いが，CiNii のアクセスログには Cookie が記録されていないためである．ただし，この方法には DHCP や NAT により別のユーザにも同じ IP アドレスが割り当てられている可能性や，Web ブラウザの更新でユーザエージェントが変わることで，同一ユーザが別のユーザとして認識される可能性があることに注意しなければならない．

なお，クローラ等の機械的アクセスはユーザの分析に悪影響を与えるので，佐藤の手法 [15] を参考に除去した．この結果，閲覧論文数は 9,106,860 件，ユーザ数は 13,038,381 人となった．

5.2 ソーシャルメディアの言及データ

ソーシャルメディアの言及情報は，収集対象となる学術文献の URL の一部を検索し，その検索結果をもとに学術文献の URL とそれに対する言及テキストを取得している [3]．例えば CiNii に収録されている学術文献は「http://ci.nii.ac.jp/naid/110008898261」のような URL で提供されているが，その URL に常に含まれる「ci.nii.ac.jp」という検索クエリを送信した結果を収集する．また，言及情報は，Facebook，Google+，Twitter，OKWave，Yahoo!知恵袋，CiteULike，Delicious，はてなブックマーク，Wikipedia，レファレンス協同データベースの計 10 サイトを対象とした．

ソーシャルメディアの言及データとして，閲覧論文と同じく 2014 年 4 月 1 日〜2016 年 3 月 31 日の 2 年間のデータを用いた．また，収集した言及データには NAID の他に，NCID（NII 書誌 ID）も含まれているが，このうち NAID を含む言及データのみを抽出した．NCID は CiNii Books の書籍に割り当てられている NACSIS-CAT 書誌 ID である．言及ユーザの識別は，各ソーシャルメディアサービス名とサービスで使われているユーザ名の組として行い，同じユーザ名であっても，異なるサービス間で別のユーザとして扱った．

言及論文でも，閲覧論文と同様に，bot ユーザは普遍度に対して大きな影響を与えてしまうため，Twitter において言及数が上位となる投稿ソースを人手で確認し，一部のソースを bot として除外した．bot 除去後のソーシャルメディアでの言及論文の総数は 31,157 件，総ユーザ数は 29,685 人であった．

6 分　　析

6.1 CiNii Articles の検索結果の特性分析

調査対象のクエリを用いて，使用したアクセスログの期間に合うように出版年を指定し，CiNii の OpenSearch API[7]で取得した論文の検索結果を閲覧人数と閲覧普遍度で再順位付けし，検索語を用いて各指標の特性の違いを分析した．

まず，上位の論文でどのような検索語が使用されているかを分析した．表 1 と表 2 に「twitter」と「推薦システム」をクエリとして，閲覧人数と閲覧普遍度で順位付けした上位 10 件の論文を，閲覧人数，閲覧普遍度の順位，論文の題名および内容・分類指定語の上位 5 語と共に示す．これらの結果を見ると，内容指定語では「観光」や「協調フィルタリング」などその論文の内容を示す単語，分類指定語では「教育」や「心理」,「研究」や「分析」など論文の分野や分類を示す単語が出現していることがわかった．また，CiNii は雑誌記事など学術論文以外の情報も多いため,「論文」などの用語を追加することで学術論文に絞り込んでいるユーザが多いことがわかった．

6.1.1 トピック分布の傾向分析

次に，閲覧人数と閲覧普遍度で論文のトピックの傾向の違いを分析した.「twitter」と「推薦システム」をクエリとして閲覧人数と閲覧普遍度で順位付けしたときの論文の内容指定語上位 5 語を，4.2 節で説明した手法で可視化したグラフを図 1 と図 2 に示す．上位 15 件に出現した検索語のうち，どちらか片方に出現した単語を赤色で，それ以外は青色で表示した．

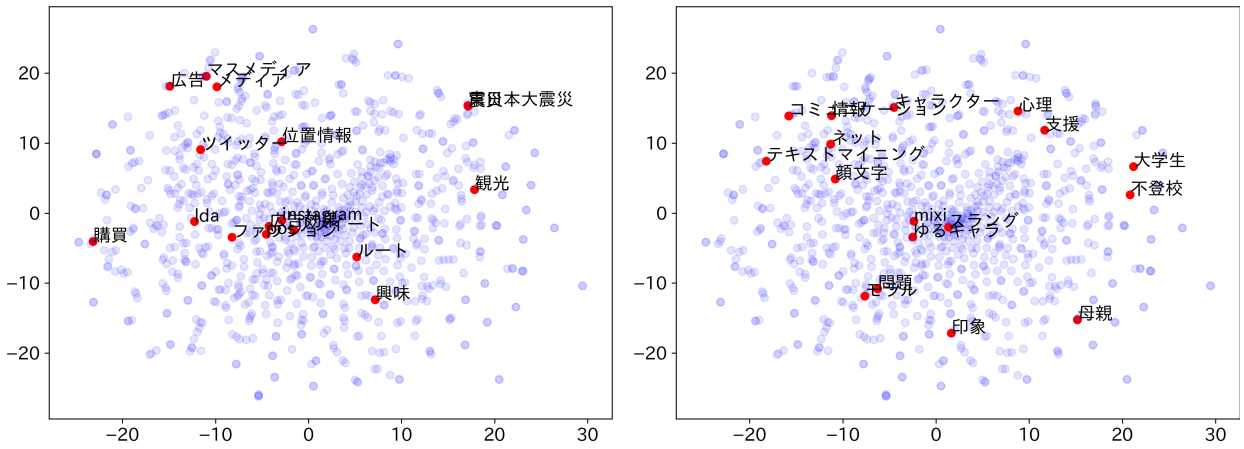
「twitter」の結果では，図 1(a) では「位置情報」や「lda」など研究に関するトピック，図 1(b) では「大学生」や「不登校」，「モラル」など若いユーザ向けのトピックが出現した．また，「推薦システム」の結果でも，図 2(a) では「可視化」や「folksonomy」など専門的なトピックが，図 2(b) では「サッカー」や「ファッション」など一般的なユーザ向けのトピックが出現した．

これらの結果から，閲覧普遍度では一般的なトピックが出現することがわかった．これは，CiNii に大学生や人文系・図書館系のユーザが多いことから，分野横断的な論文の閲覧普遍度が高くなったためだと考えられる．

6.1.2 分類指定語の頻度分析

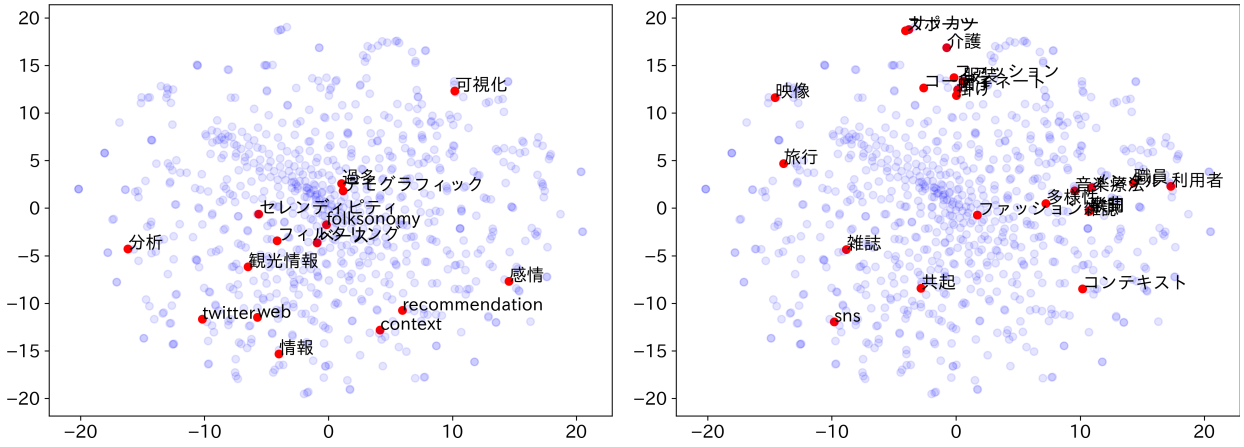
続いて，閲覧人数と閲覧普遍度で検索される論文の分野や分類の違いを分析した.「twitter」と「推薦システム」をクエリと

7 : https://support.nii.ac.jp/ja/cia/api/a_opensearch



(a) 閲覧人数　(b) 普遍度

図 1　twitter の検索語上位



(a) 閲覧人数　(b) 普遍度

図 2　推薦システムの検索語上位

して閲覧人数と閲覧普遍度で順位付けした検索結果上位 15 件の論文の分類指定語上位 5 語の頻度を表 3 に示す．なお，どちらか片方のみに出現した単語を**太字**で示す．

表 3(a) を見ると，閲覧人数では「学習」や「システム」など研究に関する単語が出現し，閲覧普遍度では「社会」や「調査」など社会系の分野や研究の分類を表す語が出現した．表 3(b) を見ると，閲覧人数では「アルゴリズム」などの専門用語が出現し，閲覧普遍度では「音楽」や「スポーツ」など芸術，体育系の分野を示す単語が出現した．これらの結果から，閲覧人数の場合では研究目的で検索されることが多く，閲覧普遍度の場合では専門的な分野とは異なる人文社会系や体育系の分野の論文を検索するユーザが多いことがわかった．また，両者の結果に共通して「論文」が頻度上位に出現し，「twitter」の場合の方がより顕著に現れていた．これは，CiNii は学術論文以外の雑誌記事も多く収録掲載しているため，研究論文に絞り込むために利用されており，さらに「twitter」のような一般的な話題ほど雑誌記事に掲載される場合が多いことが考えられる．

6.2　閲覧・言及論文の特性の比較分析

最後に，閲覧・言及論文のタイトルの面白さや実世界との関連性を，以下の 3 つの基準で各論文にラベル付けして，特性の違いを分析する．

論文の面白さ　佐藤らの先行研究 [9] に則り，論文タイトルの面白さを「非専門家にとっての内容への興味・関心やユーモア性があるか」という基準で，7 段階（値が高いほど面白い）で得点化した．なお，先行研究では人文社会学系の学部生 4 人によって評価されたが，本稿では情報系の大学院生である第 1 著者だけで評価したため，参考程度の指標であることに注意する．

著者の影響　内容指定語の上位 2 語に，著者名に含まれる単語がある論文は 1，そうでないものは 0 とした．

情報のバースト性　内容指定語上位 5 語にバースト検知された単語が含まれる論文に 1，検知されなかったら 0 とした．

バースト性を持つ検索語の抽出のために，2014 年 4 月 1 日から 2017 年 3 月 30 日の CiNii のアクセスログから，4.1 節と同じ手順で検索語の使用頻度の時系列データを作成した．ただし，検索語ごとに集計すると疎になりやすく，バーストが誤検知される可能性が高くなるので，サンプリング期間を 10 日間とした．そして，全検索語 745,403 語のうち，使用頻度が 50 回以上の 79,053 語（全体の 10.6%）をバースト判定した．

CiNii での閲覧とソーシャルメディア上の言及が共に行われていた論文 30,679 件を対象に，閲覧人数（閲人），閲覧普遍度（閲普），言及人数（言人），言及普遍度（言普）の 4 種類の指標で順位付けした時の，上位 10 件の論文を表 4 に示す．左側 4 列に各指標の順位，続いて論文の題名，内容・分類指定語上位 5 語と各ラベルを示す．なお，バースト検知された内容指定語は下線を付けて示す．

まず，情報のバースト性に着目すると，全体的に上位にはバーストフラグが 1 の論文が多かった．これは，ソーシャルメディアや実世界で話題になると顕著に論文へのアクセス数が増えるために当然といえる．

さらに，著者の影響も考慮すると，著者フラグが 1 の論文は，ほとんどがバーストフラグが 1 であった．これは，著者名で検索してくるユーザが多数存在する場合は，論文の内容ではなく，実世界の著者に興味を持っている可能性が高いことを示す．例えば，閲覧人数 7 位や 8 位の論文は，それぞれ Hey! Say! JUMP の「伊野尾慧」，STAP 騒動の「小保方晴子」の実世界における行動やニュースがバーストの原因であったと考えられる．この種の論文が評価が高い理由は，内容や題名を見ただけでは判別できないが，論文のメタ情報（書誌情報）に含まれる著者や所属と照合することができれば，容易に判別できることがわかる．ただし，例外として，言及人数 8 位の「さかなクン」と 9 位の「藤末」の論文のバーストフラグは 0 であった．どちらも，CiNii に論文や紀要，雑誌記事が複数掲載されており，長期に渡る実世界の影響があるからと考えられるが，そのような場合は論文の内容などで識別できると考えられる．

続いて，タイトルの面白さに着目すると，面白さ得点が高い論文にはバーストフラグが 1 の論文が多かったが，バーストフラグが 1 で，かつ著者フラグが 1 のほとんどの論文は面白さ得点が低く，逆は成り立たないことがわかる．これは，先ほど述べたように，バーストする理由が論文の内容ではなく，実世界の著者にあるからと考えられる．ただし，言及普遍度 10 位の「永田」は最大得点だが，永田大輔がオタク文化の研究者であり，論文と実世界の面白さが直結している例外と考えられる．

また，バーストフラグが 0 でも面白さ得点が高い論文も存在した．このような論文は，例えば，閲覧普遍度 5 位は「恋愛」や「コミュニケーション」7 位は「大学生」や「睡眠」など，定常的に使用される単語で検索されていた．つまり，論文の面白さには，長期的なものと，短期的なものがあり，少なくとも前者はソーシャルメディア上で評価されなくても，論文を検索している多くのユーザにとって面白い話題であると推測できる．

7 おわりに

CiNii の閲覧・言及行動を分析するために，検索結果や論文集合を閲覧・言及の人数と普遍度の各指標で再順位付けした結果を，ユーザが検索に用いた検索語を用いて各指標の特性の違いや，論文のタイトルの面白さや著者，バースト性との関連を分析した．その結果，検索語のバースト性から，ソーシャルメディア上の話題性などに反応して論文にアクセスするユーザが多いことや，分類指定語から，CiNii は研究目的で利用するユーザも多いことがわかった．これらの結果から，検索語のバースト性，定常性という評価軸を論文のランキングに適用することは有用であると考えられる．今後の課題として，さらに多くの論文を対象とすることや，論文アクセスのバーストも合わせて分析する予定である．

謝　辞

本研究は JSPS 科研費 19H04421 と国立情報学研究所公募型共同研究「学術情報検索における論文閲覧行動とソーシャルメディアにおける論文言及行動の関連性に関する研究」の助成を受けた．

文　献

[1] 林和弘. 科学技術動向研究 研究論文の影響度を測定する新しい動き: 論文単位で即時かつ多面的な測定を可能とする Altmetrics. 科学技術動向, pp. 20–29, 2013.

[2] Jason Priem, Heather A. Piwowar, and Bradley M. Hemminger. Altmetrics in the wild: Using social media to explore scholarly impact. arXiv:1203.4745 [cs.DL], 2012.

[3] 吉田光男. 計量書誌学の新たな挑戦：国産オルトメトリクス計測サービスの開発 (<特集>計量書誌学を超えて). 情報の科学と技術, Vol. 64, No. 12, pp. 501–507, 2014.

[4] Ehsan Mohammadi, Mike Thelwall, Mary Kwasny, and Kristi L. Holmes. Academic information on Twitter: A user survey. *PLOS ONE*, Vol. 13, No. e0197265, may 2018.

[5] 佐藤翔, 安藤大輝, 川瀬直人, 北島顕正, 塩崎亮, 那珂元, 原田隆史. ディスカバリサービスにおける絞り込みプロセス：国立国会図書館サーチのアクセスログ分析. 図書館界, Vol. 67, No. 4, pp. 244–261, 2015.

[6] Maya Hristakeva, Daniel Kershaw, Marco Rossetti, Petr Knoth, Benjamin Pettit, Saúl Vargas, and Kris Jack. Building recommender systems for scholarly information. In *Proceedings of the 1st Workshop on Scholarly Web Mining (SWM '17)*, pp. 25–32, 2017.

[7] Mike Thelwall. Does female-authored research have more educational impact than male-authored research? evidence from mendeley. *Journal of Altmetrics*, Vol. 1, No. 1, p. 3, 2018.

[8] 佐藤翔, 吉田光男. 日本の学協会誌掲載論文のオルトメトリクス付与状況. 情報知識学会誌, Vol. 27, No. 1, pp. 23–42, 2017.

[9] 佐藤翔, 石橋柚香, 南谷涼香, 奥田麻友, 保志育世, 吉田光男. Twitter からの言及数が多い論文は言及されたことのない論文と比べてタイトルが「面白い」. 情報知識学会誌, 2019.

[10] 小林和央, 風間一洋, 吉田光男, 大向一輝. インターネット上の論文の閲覧行動と言及行動の関係の分析. In *proc. CSSJ2019*, 2019.

[11] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In *proc. ICLR 2013*, 2013.

[12] Laurens Van Der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. Vol. 9, pp. 2579–2605, 2008.

[13] Jon Kleinberg. Bursty and hierarchical structure in streams. *In Proceedings of the 8th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 91–101, 2002.

[14] Liu Bing, editor. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, chapter Web Usage Mining. Springer-Verlag, 2009.

[15] 佐藤翔. コンテンツ入手元として機関リポジトリが果たしている役割. PhD thesis, 筑波大学, 3 2013.

表 1　検索結果の順位付け結果（twitter）

(a) 閲覧人数上位

人数	普遍度	題名	内容指定語	分類指定語
1	1	若者における SNS 利用行動およびリスク認知の検討 : LINE と Twitter を中心に	sns,twitter, 若者,line, リスク	論文, 研究, 行動, 文化, 認知
2	2	ソーシャルメディアにおけるプライバシーリスクの盲点 : リスク遮滅に向けた論点整理	sns,twitter,line, ソーシャルメディア,face-book	論文, 教育, 研究, 情報, 問題
3	6	SNS の投稿を用いた感情記録ライフログシステム ～Emote～	twitter,sns, 感情,facebook, ライフログ	論文, 分析, 学習, システム, 画像
4	76	位置情報付きツイートを利用した観光ルート推薦	instagram, 観光,twitter, 位置情報, ルート	論文, 評価, 行動, システム, 観光
5	4	Twitter における影響力の分析手法	twitter,facebook, マイクロブログ, 影響力, ブログ	論文, 分析, 影響, 問題, 解析
6	30	商品に関する Twitter 上のコミュニケーションと販売実績の関連性分析	twitter, 広告, マーケティング,pos, 広告効果	論文, 分析, 効果, 解析, コミュニケーション
7	10	一般ユーザの観点に基づく Twitter からの人物関係の可視化と事例の考察	twitter,sns, マーケティング, マイクロブログ, トピック	論文, 分析, 研究, 情報, モデル
8	3	朝ドラ『あまちゃん』はどう見られたか : 4 つの調査を通して探る視聴のひろがりと視聴熱	あまちゃん, 朝ドラ, 視聴率, ドラマ,twitter	論文, 日本, 研究, 文化, メディア
9	46	Twitter におけるユーザの興味と話題の時間発展を考慮したオンライン学習可能なトピックモデルの提案	twitter,lda, 購買, トピック, 興味	論文, 研究, 行動, 学習, 教育
10	37	ソーシャルメディアの普及がファッションの学習と情報流通に与えた影響に関する一考察	instagram, ファッション, ソーシャルメディア,twitter, メディア	論文, 影響, 情報, 大学, メディア
11	20	一般ユーザの観点に基づく Twitter からの人物関係の可視化	twitter,sns, ソーシャルメディア, マーケティング, マイクロブログ	論文, 分析, 研究, 関係, 情報
12	5	Twitter 上でのシャイなユーザーの自己開示（思考と言語）	自己開示,twitter, シャイ,cmc, ネス	日本, 研究, 文化, 影響, 日本語

(b) 閲覧普遍度上位

人数	普遍度	題名	内容指定語	分類指定語
1	1	若者における SNS 利用行動およびリスク認知の検討 : LINE と Twitter を中心に	sns,twitter, 若者,line, リスク	論文, 研究, 行動, 文化, 認知
2	2	ソーシャルメディアにおけるプライバシーリスクの盲点 : リスク遮滅に向けた論点整理	sns,twitter,line, ソーシャルメディア,face-book	論文, 教育, 研究, 情報, 問題
8	3	朝ドラ『あまちゃん』はどう見られたか : 4 つの調査を通して探る視聴のひろがりと視聴熱	あまちゃん, 朝ドラ, 視聴率, ドラマ,twitter	論文, 日本, 研究, 文化, メディア
5	4	Twitter における影響力の分析手法	twitter,facebook, マイクロブログ, 影響力, ブログ	論文, 分析, 影響, 問題, 解析
12	5	Twitter 上でのシャイなユーザーの自己開示（思考と言語）	自己開示,twitter, シャイ,cmc, ネス	日本, 研究, 文化, 影響, 日本語
3	6	SNS の投稿を用いた感情記録ライフログシステム ～Emote～	twitter,sns, 感情,facebook, ライフログ	論文, 分析, 学習, システム, 画像
32	7	Twitter からのキャラクター印象表現の抽出 (言語理解とコミュニケーション)	ゆるキャラ,twitter, キャラクター, 印象, テキストマイニング	評価, 調査, コミュニケーション, 言語, 表現
58	8	Twitter による不登校の母親援助 -140 字支援による可能性とその限界-	不登校,twitter, 心理, 支援, 母親	心理, 支援, 研究, 心理学, 社会
46	9	Twitter に見る若者たちのコミュニケーション考 (特集 ネット化社会)	twitter, 若者, コミュニケーション, ネット, 問題	論文, コミュニケーション, 社会, 問題, 若者
7	10	一般ユーザの観点に基づく Twitter からの人物関係の可視化と事例の考察	twitter,sns, マーケティング, マイクロブログ, トピック	論文, 分析, 研究, 情報, モデル

表 2　検索結果の順位付け結果（推薦システム）

(a) 閲覧人数上位

人数	普遍度	題名	内容指定語	分類指定語
1	9	推薦システムのアルゴリズム (1)	推薦, 神嶌, システム, 神嶌敏弘, 敏弘	システム, アルゴリズム, 推薦,system,based
2	22	状況依存型ユーザ嗜好モデリングに基づく Context-Aware 情報推薦システム	推薦, 情報,context, 嗜好,recommendation	情報, システム, 状況, 依存, 手法
3	1	絵本の読み聞かせにおける子どもの好みと絵本の主題との関係性	絵本, 読み聞かせ, 子ども, 好み, 幼児	論文, 教育, 子ども, 分析, 研究
4	36	協調フィルタリングとコンテンツ分析を利用した観光地推薦手法の検討	推薦, 協調フィルタリング, 観光地, 観光, 勇之	論文, 分析, システム, 情報, 観光
5	153	Folksonomy マイニングに基づく Web ページ推薦システム	推薦, 協調フィルタリング,folksonomy, システム,web	論文, システム,web, インターネット, 推薦
6	11	オノマトペロリ : オノマトペを利用した料理推薦システム	オノマトペ, オノマトペロリ, 料理, 味覚, 推薦	論文, 日本, システム, 表現, 日本語
7	3	図書館の貸出履歴と書誌情報を用いた図書推薦システムの有効性	推薦,opac, 貸出, 逸村, 図書館	教育, システム, 評価, 情報, 図書館
8	95	Twitter 感情分析を用いた感情値可視化とユーザ推薦システム	twitter, 感情, 分析, 推薦, 可視化	論文, 分析, 感情, 英語, システム
9	6	図書館の貸出履歴を用いた図書推薦システムの有効性検証	推薦, 協調フィルタリング,opac, 図書, 図書館	システム, 研究, 図書館, 情報, 比較
10	2	読み聞かせ時の反応に着目した絵本に対する子どもの好みの取得方法に関する検討	絵本, 読み聞かせ, 好み, 子ども, 幼児	論文, 教育, 子ども, 日本, 幼児

(b) 閲覧普遍度上位

人数	普遍度	題名	内容指定語	分類指定語
3	1	絵本の読み聞かせにおける子どもの好みと絵本の主題との関係性	絵本, 読み聞かせ, 子ども, 好み, 幼児	論文, 教育, 子ども, 分析, 研究
10	2	読み聞かせ時の反応に着目した絵本に対する子どもの好みの取得方法に関する検討	絵本, 読み聞かせ, 好み, 子ども, 幼児	論文, 教育, 子ども, 日本, 幼児
7	3	図書館の貸出履歴と書誌情報を用いた図書推薦システムの有効性	推薦,opac, 貸出, 逸村, 図書館	教育, システム, 評価, 情報, 図書館
24	4	複数人での旅行における嗜好分析による観光地推薦システムの提案	観光, 推薦, 旅行, 観光地, システム	システム, 評価, 情報, 分析, 観光
87	5	スポーツメンタルトレーニングへの応用を目指した脳波利用の音楽推薦システム	音楽, スポーツ, メンタルトレーニング, メンタル, 脳波	論文, 心理, スポーツ, 音楽, 評価
9	6	図書館の貸出履歴を用いた図書推薦システムの有効性検証	推薦, 協調フィルタリング,opac, 図書, 図書館	システム, 研究, 図書館, 情報, 比較
43	7	映像コンテキストに基づく多視点映像の視点列推薦	サッカー, 映像, スポーツ, 推薦, コンテキスト	スポーツ, 研究, システム, 評価, 画像
15	8	高校生の食行動に関する実態報告 (第 1 報)	脳波, 音楽, 心理, メンタルトレーニング, 推薦	論文, 心理, 音楽, 学習, 影響
1	9	推薦システムのアルゴリズム (1)	推薦, 神嶌, システム, 神嶌敏弘, 敏弘	システム, アルゴリズム, 推薦,system,based
213	10	介護における声かけに着目した介護職員に対する警告・行動推薦システムの提案 (ライフインテリジェンスとオフィス情報システム)	介護, 掛け, 共起, 利用者, 職員	論文, 介護, 文献, 行動, 言語
6	11	オノマトペロリ : オノマトペを利用した料理推薦システム	オノマトペ, オノマトペロリ, 料理, 味覚, 推薦	論文, 日本, システム, 表現, 日本語
28	12	H-021 ファッション雑誌を用いたコーディネート推薦システム (H 分野:画像認識・メディア理解, 一般論文)	ファッション雑誌, ファッション, コーディネート, 雑誌, 推薦	論文, コミュニケーション, 雑誌, メディア, 画像

表 3　出現頻度上位の分類指定語

(a) 「twitter」の場合

閲覧人数上位	頻度	普遍度上位	頻度
論文	12	論文	10
研究	7	研究	7
分析	7	文化	4
情報	5	情報	4
影響	4	分析	4
日本	4	問題	3
行動	3	調査	3
文化	3	コミュニケーション	3
学習	3	日本	2
教育	2	影響	2
問題	2	解析	2
システム	2	社会	2
解析	2	大学生	2
コミュニケーション	2	学生	2
メディア	2	大学	2

(b) 「推薦システム」の場合

閲覧人数上位	頻度	普遍度上位	頻度
システム	12	論文	8
論文	9	システム	8
情報	5	評価	5
推薦	4	音楽	4
分析	4	教育	3
based	3	研究	3
教育	3	情報	3
アルゴリズム	2	心理	3
子ども	2	子ども	2
研究	2	分析	2
観光	2	日本	2
日本	2	図書館	2
評価	2	スポーツ	2
図書館	2	画像	2
system	1	影響	2

表 4　各指標の上位の論文

(a) 閲覧人数の場合

閲人	閲普	言人	言普	題名	内容指定語	分類指定語	面白さ	著者	バースト
1	2	927	10902	修正版グラウンデッド・セオリー・アプローチ (M-GTA) の分析技法	gta, 修正, グラウンデッドセオリー, アプローチ,グラウンデッド・セオリー	論文, 分析, 看護, 研究, 方法	1	0	1
2	850	3	723	ブラジャー着用時と非着用時の運動中の乳房振動特性	ブラジャー, 乳房, 振動, 特性, ブラ	論文, 運動, 研究, 特性, 振動	6	0	1
3	8	8651	20766	看護のための文献検索のポイント : 医中誌 Web を使って	検索, 文献, 看護研究, 看護, 無料	論文, 看護, 研究, 文献, 日本	2	0	0
4	341	39	115	ポスドクからポストポスドクへ	ポスドク, 円城塔, 円城, ポストポスドク, 問題	論文, 問題, 学会, 物理,pdf	3	1	1
5	1672	1	9	コーヒーカップとスプーンの接触音の音程変化	コーヒーカップ,スプーン, インスタントコーヒー, コーヒー, 塚本浩司	教育, 研究, 変化, 科学, 振動	6	0	1
6	1	8651	22844	認知された自己の諸側面の構造	山成, 山本真理子, 松井, 山本, 教育心理学	研究, 教育, 論文, 心理, 日本	2	1	1
7	11	4	5038	7336 津波避難に係る学校施設の整備のあり方 : 津波被害のあった小中学校の避難に係る事例...	伊野尾慧,伊野尾,大学院, 明治大学, 論文	論文, 建築, 研究, 大学, 日本	2	1	1
8	3	2209	3089	海外情報 Harvard Medical School での再生医療教育	小保方晴子,小保方, 晴子,harvard,medical	教育, 情報, 海外,medical, 再生医療	1	1	1
9	26	8651	24544	気管支肺胞洗浄 (BAL)	気管支,bal, 洗浄, 洗浄液,balf	評価, 研究, 方法, 看護, 解析	1	0	0
10	37	1687	21682	事例報告 高機能広汎性発達障害児 (者) と「不登校」「ひきこもり」の臨床的検討 (特集 高機能自閉症とアスペルガー症候群)	不登校, 高機能広汎性発達障害, 雅文, 相澤, 広汎性発達障害	論文, 障害, 機能, 発達障害, 不登校	4	0	0

(b) 閲覧普遍度の場合

閲人	閲普	言人	言普	題名	内容指定語	分類指定語	面白さ	著者	バースト
6	1	8651	22844	認知された自己の諸側面の構造	山成, 山本真理子, 松井, 山本, 教育心理学	研究, 教育, 論文, 心理, 日本	2	1	1
1	2	927	10902	修正版グラウンデッド・セオリー・アプローチ (M-GTA) の分析技法	gta, 修正, グラウンデッドセオリー, アプローチ,グラウンデッド・セオリー	論文, 分析, 看護, 研究, 方法	1	0	1
8	3	2209	3089	海外情報 Harvard Medical School での再生医療教育	小保方晴子,小保方, 晴子,harvard,medical	教育, 情報, 海外,medical, 再生医療	1	1	1
66	4	3059	7743	現代の貧困と子どもの発達・教育	貧困, 教育格差, 学力, 児童虐待, 格差	論文, 教育, 日本, 研究, 心理	4	0	0
27	5	3059	13617	日常的コミュニケーションが恋愛関係に及ぼす影響	恋愛, コミュニケーション, 日常的, 関係, 心理学	論文, 研究, 心理, 心理学, 行動	7	0	0
93	6	4677	15692	ディズニー映画のプリンセス物語に関する考察	ディズニー,ディズニープリンセス, プリンセス, ディズニー映画, 映画	論文, 教育, 研究, 文化, 文献	7	0	1
49	7	4677	24544	大学生における睡眠の質と関連する生活習慣と精神的健康	睡眠, 大学生, 睡眠時間, ストレス, 健康	論文, 運動, 教育, 研究, 評価	6	0	0
3	8	8651	20766	看護のための文献検索のポイント : 医中誌 Web を使って	検索, 文献, 看護研究, 看護, 無料	論文, 看護, 研究, 文献, 日本	2	0	0
40	9	8651	15692	なぜ東京ディズニーランドは人気があるのか。サービス・マーケティングからの分析	ディズニー, ディズニーランド, 東京ディズニーランド, 人気, マーケティング	論文, 分析, 日本, サービス, 東京	7	0	1
15	10	8651	24544	音楽と感情についての心理学的研究	音楽, 感情, 論文, 性格,音楽心理学	論文, 研究, 心理, 心理学, 音楽	5	0	1

(c) 言及人数の場合

閲人	閲普	言人	言普	題名	内容指定語	分類指定語	面白さ	著者	バースト
5	1672	1	9	コーヒーカップとスプーンの接触音の音程変化	コーヒーカップ,スプーン, インスタントコーヒー, コーヒー, 塚本浩司	教育, 研究, 変化, 科学, 振動	6	0	1
12	4385	2	1	日本の ＜ 疑似伝統 ＞	疑似, 伝統, 利麿, 阿満利麿, 阿満	日本, 研究,cinii, 伝統, 論叢	2	0	0
2	850	3	723	ブラジャー着用時と非着用時の運動中の乳房振動特性	ブラジャー, 乳房, 振動, 特性, ブラ	論文, 運動, 研究, 特性, 振動	6	0	1
7	11	4	5038	7336 津波避難に係る学校施設の整備のあり方 : 津波被害のあった小中学校の避難に係る事例...	伊野尾慧,伊野尾,大学院, 明治大学, 論文	論文, 建築, 研究, 大学, 日本	2	1	1
31	65	5	8	部活動への参加が中学生の学校への心理社会的適応に与える影響:—; 部活動のタイプ・積...	部活動, 部活, 文化部, 中学生, 関係	論文, 教育, 研究, 心理, 日本	6	0	0
3780	1454	6	22844	大学生の性役割意識–男女間のギャップを中心に	性役割, 男女, 交際, リーダーシップ, 意識	日本, 意識, 大学生, 学生, 関係	7	0	0
54	3831	7	3387	女子大学生による暑熱環境下におけるパンティ-ストッキングの着用評価	ストッキング, パンティー, パンティ, パンティストッキング,暑熱	論文, 評価, 大学, 環境, 文献	6	0	1
235	13840	8	15	津波被災地の小学校における海の認識に関するテキスト分析	さかなクン, テキスト, 分析,sanriku, 環境リテラシー	教育, 分析, 環境, 臨床, 学会誌	2	1	0
30	2310	9	7	Perfume のダンスはなぜ難しいのか?-多変量ヒルベルトーファン変換によるモーション解析	perfume, ダンス,prfm, モード, ヒップホップ	論文, 解析, 分析, 情報, 大学	7	0	1
16	2541	9	3	神への気付き–宇宙が神に変わるとき	藤末, 医科, 大学	大学, 医科, 藤末	2	1	0

(d) 言及普遍度の場合

閲人	閲普	言人	言普	題名	内容指定語	分類指定語	面白さ	著者	バースト
12	4385	2	1	日本の ＜ 疑似伝統 ＞	疑似, 伝統, 利麿, 阿満利麿, 阿満	日本, 研究,cinii, 伝統, 論叢	2	0	0
551	9959	14	2	カール禿頭王は本当に禿げていたか	禿頭, カール, ハゲ, 男性,tonsure	男性, 俊一, カール, ハゲ, 赤阪	6	0	0
16	2541	9	3	神への気付き–宇宙が神に変わるとき	藤末, 医科, 大学	大学, 医科, 藤末	2	1	0
8941	21091	156	4	漢字圏の文学における西方占星術の要素 : 東西文化交流における仏教の役割	占星術,120005703462,kotyk, 東西, コテック	文化, 仏教, 交流, 物語, 漢字	2	0	0
1003	7769	47	5	＜ 翻訳 ＞ アリストテレス『政治学』	アリストテレス, 政治学, 翻訳, 岡大, 全文	論文, 翻訳, アリストテレス, 政治学, 荒木	3	0	0
1248	7914	62	6	大学、短期大学、高等専門学校図書館等におけるライトノベルの所蔵状況	ライトノベル,佐藤翔,所蔵,図書, 状況	大学, 状況, 図書, 所蔵, ライトノベル	7	0	1
30	2310	9	7	Perfume のダンスはなぜ難しいのか?-多変量ヒルベルトーファン変換によるモーション解析	perfume, ダンス,prfm, モード, ヒップホップ	論文, 解析, 分析, 情報, 大学	7	0	1
31	65	5	8	部活動への参加が中学生の学校への心理社会的適応に与える影響:—; 部活動のタイプ・積...	部活動, 部活, 文化部, 中学生, 関係	論文, 教育, 研究, 心理, 日本	6	0	0
5	1672	1	9	コーヒーカップとスプーンの接触音の音程変化	コーヒーカップ,スプーン, インスタントコーヒー, コーヒー, 塚本浩司	教育, 研究, 変化, 科学, 振動	6	0	1
2122	2436	401	10	コンテンツ消費における「オタク文化の独自性」の形成過程 : 一九八〇年代におけるビデオテープ...	オタク,永田, 大輔, オタク文化, 消費	文化, 消費, 社会学, 大輔, 永田	7	1	1