

Bibliometrics を利用した高等教育における文献探索のシステム化について

吉永 敦征

山口県立大学情報化推進室

概要

本論文では Bibliometrics の手法を利用しテーマに即して収集した文献群から基礎的な文献を自動的に発見するシステムの概説を行う。また、本システムを利用することで効率的な研究の促進が高等教育において可能かなのかどうかの評価を行い、学生からの返答を元に学生にとっては有用でないとの暫定的な結論を得た。

On a systemization of bibliographic search in higher education by means of Bibliometrics

Nobuyuki YOSHINAGA

Yamaguchi Prefectural University

概要

In this paper I give an overview of a system that automatically finds the most important papers from literature on certain subject by means of Bibliometrics. Furthermore, I sought to verify this system to see whether it could be used to help students in higher education do research more effectively and efficiently. As a result, I drew a conclusion that this system is not currently useful for that purpose on the ground of responses from students.

1 本システムの目的

本システムの目的は主題に即した文献探索を Bibliometrics の手法を利用したウェブベースのアプリケーションを通じてより簡便に学生に利用させることにある。学生は本システムを利用することで無意識的にアカデミックスキルの習得が可能であり、いわゆる文献探索法を身に付けさせることができる。学生が自ら文献を収集することで基本論文や基本文献に容易に辿り着くためのプロセスを自動化することを目指している。この意味で、本システムはたとえば武田など [1] が行っているような、大規模なデータベースを利用した科学者コミュニティの分析や知識生成の足跡を辿ることは目的ではない。

Bibliometrics という言葉を本システムが目指しているものに適用するかに問題を感じるかもしれないが、Bibliometrics という単語自体が持つ意味の幅¹や、Pritchard による Bibliometrics の定義「記録されたコミュニケーションのさまざまな側面を計量・分析し、記録されたコミュニケーションの過程およびひとつの分野の発展の特性と過程に光を当てること」[4]における計量・分析するという手法を計算機上で処理するという点で本システムは Bibliometrics の範疇内と考えることが可能だろう。

2 システムの概要

2.1 使用しているソフトウェア

本システムは Linux をベースとし、Apache, MySQL, PHP で構築している。文献情報の格納には SQL を利用し、各文献間の参照関係図を作成するためのソフトウェアとして AT&T が開発・配布している Graphviz²を利用している。本プログラムは javascript と一つの PHP のスクリプトから構成しているウェブベースのプログ

¹Bibliometrics, scientometrics, informetrics などの用語も流通しており、それぞれの用語の意味や使用法などがいまだ定まっておらず、使用者によって異なっているという指摘がある [2]。

²Graph Visualization Software, <http://www.graphviz.org/>

ラムであり、同様の環境を持つ他のシステムで容易に動作させることができ、個人ベースで資料の整理に利用したり研究会などでの文献を蓄積するための利用も可能である。³。また、本プログラムは GPL に準拠しウェブ上で公開している⁴。

2.2 システムの概要

本システムの基本は三つのテーブルで構成している。一つは文献情報を登録するテーブル (A) でありもう一つは一つは文献間の関係を表現するテーブル (B)、「テーマ」を格納するテーブル (C) である。(A) には文献情報を一つのレコードとしてデータベースに格納し、(B) には「論文 (a)」と「論文 (a) が引用している論文 (b)」の ID を格納している。(C) には収集した文献群をまとめるための主題を格納し、入力した文献情報がどのような主題の元に収集されているのかを判別可能にしている。テーマは十進分類法に基づき管理している。似たような主題を持つ文献情報を互いに利用できるようにするための処置である。

たとえば、図 1 のようにある主題の元で収集した三つの文献が引用関係を持っているとする。a(1) は b(2) と c(3) を引用し、b(2) が c(3) を引用するという関係である。この関係はデータベース上では図 2 のように表現している。入力された文献は「テーマ」によってまとめられ、一つの群を構成する。

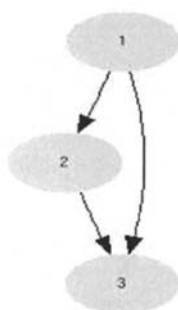


図 1: 文献の関係図

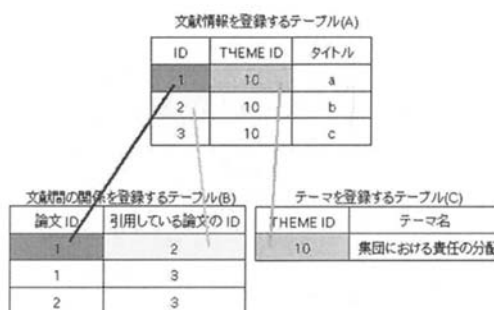


図 2: 各テーブルの関係

テーブル (A) には「著者」「雑誌名」「発行年」「ページ数」「DOI」などのカラムを用意し、書誌情報の分類を行っている。これらの項目では文献の管理には十分とは言えないかもしれないが、テーマに沿って入力された文献群から重要な文献を探し出す程度には文献を個別化できる点で有効であろうと思われる。

重要な文献の発見にはテーブル (B) の引用関係を利用する。つまりテーブル (B) に格納されたそれぞれの論文についての引用数を数え上げることで、もっとも引用数が多い文献を自動的に算出させることが可能であり、その結果を利用して文献間の参照関係図の作成を自動的に行っている。

2.3 動作画面

利用者はすべての操作をウェブブラウザ上で行う。作業内容の基本は下記の三つのステップである。

1. 検索した論文の書誌情報を入力
2. 引用文献の書誌情報を入力
3. 文献の参照関係を入力

³現在稼働させているシステムには LDAP を用いた認証を組み込んでいるが改変は容易である。

⁴<http://plato.ypu.jp/nyoshi/>

書誌情報の入力には図3にあるようにPHPで作成したインターフェースを利用し、記入したデータをテーブル(A)に格納する。

図 3: 書誌情報の登録

図 4: 引用関係の登録

入力したデータに引用文献を登録するインターフェースが図4である。引用関係の設定は、入力済みの文献に対してチェックをつけることで、テーブル(B)へと文献の関係を登録する。上記のステップに従い文献を入力し、引用文献にチェックをするという作業を繰り返すことで、自動的に核となる文献が探し出されることとなる⁵。

入力した文献群から Graphviz での描写に必要となる形式へと文献情報や参照関係のデータを加工し画像ファイルを作成する。作成するのは「時系列」「ネスト状」「著者別」の三つの図であり、これらの図はデータベースに新規に情報が追加された時点で再描写し、核となる文献もしくはゲートキーパーとなる人物を中心とした図が再び自動的に作成され文献群の関係を直観的に把握することが可能となる。

時系列の図では、文献の発表年に合わせて入力された文献情報を並べている。この図により、自らが設定した主題においてもっとも古く且つ重要と思われる文献を発見することができる。

次に図6にあるネスト状の図で表現したい内容は次の通りである。すなわちもっとも古い文献が重要な文献とは限らないことを考慮し、重要な文献を中心として文献を配置しており、この図によって主題が似ているが関連が薄い文献を洗い出すことにも有用である。

図7は主題におけるゲートキーパーを見つけ出すために作成している。ゲートキーパーを発見した後には、当該の人物を手がかりにしてさらなる文献探索が可能となる。

2.4 ユーザの管理

本システムは学生全員の利用を見込んでおり、このようなユーザ参加型のシステムでは個々の利用者が正しく情報を入力・更新することで、全体としての利益が上がるが見込まれる。しかしながら、学生が自らのテーマを設定し文献探索を行うことが本システムの主眼でありデータベース総体としての発展は念頭に入れていないため下記の条件をつけてユーザ管理を行うこととした。

1. 文献情報の入力・編集を行えるのは自分が設定したテーマのみ

⁵ この手順によって算出された文献が、ある領域における重要な文献であるとは限らない。しかしながら自らが設定したテーマに即して収集した文献群の中では当該の文献がコアとなるということまでは言えるだろう。

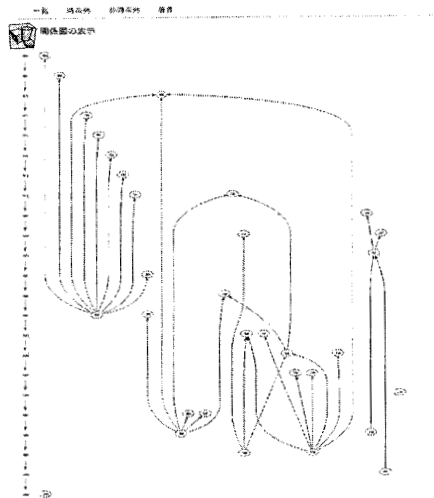


図 5: 時系列

2. 他者の設定したテーマにおける文献情報は (システム上の許可があれば) 自由に利用できる

この条件により、悪意のある改変を防ぎデータベース内の文献情報の整合性を求めている。当初、ユーザ管理は SQL 自身で行っていたが、全学的な情報サービスの認証を LDAP へ変更したことに合わせて本システムも LDAP での認証に切替えている。

3 実際の運用

Bibliometrics については、特に引用を中心とした分析には本質的な問題が存在していることが明らかだと思われるが、学部教育での有効性を検証するために実際に授業で本システムの運用を行った。本学で開講している司書・司書教諭資格用の科目である「情報検索演習」を履修している 25 人の学生に本システムを利用させ、基本論文の収集を課題として、テーマに応じた情報検索の方法を実践する道具として有効であるかどうかを学生からの感想を元に評価を行った。対象となった学生は大学院生が 1 人、科目等履修生が 2 人、学部 3 年生が 2 人、学部 2 年生が 20 人であり、文献の収集を行った事無い学生が 75% を占めている。全部で二コマ分の授業時間を費し、システムの利用方法や文献探索法、引用文献の集めかたなどを教授しつつ学生には主題に応じた文献の収集を行わせた。

本システムを利用させた当初の推測では学部での利用には本質的な問題は問題とならず、むしろ教育的効果が期待できるはずであったが、実際には教育的効果についても疑わしいという結果になった。

3.1 本質的な問題

本質的な問題は二点存在する。一点は Bibliometrics が適用できる分野の正当性であり、もう一点は引用行為自体に内在する困難さである。

Bibliometrics はその概念が提出された当初とはその利用方法が異なり、現在では科学知識の生成過程や、科学者コミュニティと科学知識の関係などの科学社会学の側面が強いように見受けられる [6, 倉田,1999]。また

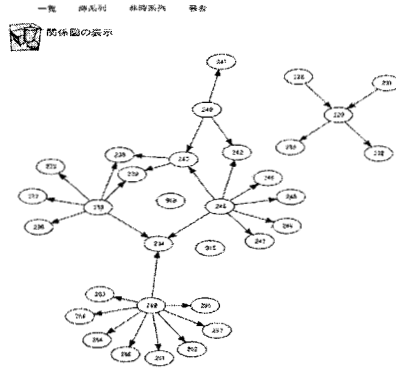


図 6: ネスト状

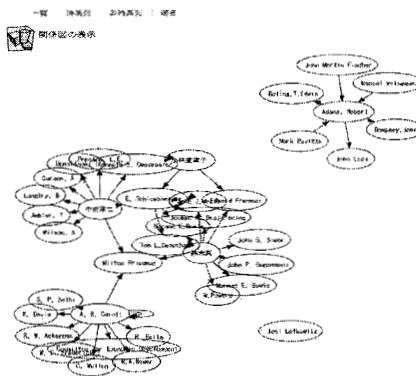


図 7: 著者別

業績評価のために IF を利用するという側面も批判されつつ存在するようであるが⁶、基本的には自然科学を対象とした知識の生成・蓄積に関する数学的・統計的分析となっている。

本学には 3 つの学部「国際文化学部」「社会福祉学部」「看護栄養学部」が存在し、そのなかで理科系の学科は栄養学科と看護学科の二つである。栄養学科や看護学科に所属する学生が主題に即した文献探索を行う場合には、まさに科学的方法論を適用することで知識の獲得が可能であり、その恩恵に与れることは確かだろう。仮りに人文科学の学問分野も自然科学と同様の方法での知識生成が行われており、人文科学の方法論が Bibliometrics に還元可能であれば、文科系の学生にも本システムを利用することで、当該分野の知識のマッピングを可能とする手段を提供できることになる。しかしながら他の学科の学生にとって本システムが本質的に有効であるかどうかは、人文科学における Bibliometrics の有効性が明らかにされた後に判断しなければならない。

引用行為そのものに対する批判的な指摘は MacRoberts[9] により行われている⁷。

- 引用されない文献の存在: 実際には影響を受けているにもかかわらず引用している文献は 30%にすぎない。
- 偏った引用が行われる傾向がある: 調査の結果特定の文献は引用される一方で、まったく引用されない文献がある。

⁶小野寺 [7] や根岸 [8] など

⁷MacRoberts の論文では引用に関する伝統的な問題点のまとめと、社会構成主義的観点からの引用に対する批判という二つの論点が描かれている。本論文で引用したのは引用に関する伝統的な問題点のみである。

- 二次文献の引用: 38%程度の引用が二次文献の引用となっている。
- 文献によらない影響: インフォーマルコミュニケーションを拾いあげられていない
- 引用者の動機: 文献による影響のみではなく、心理学的要素によっても引用が行われている
- 引用パターンが分野や様々な違いを反映して異なる:
- 伝統的に引用を行わない分野の存在:
- データベースの恣意性: SCIなどが測定している雑誌の恣意性

これらの問題が解決されない限りは、真の意味で有効性を主張することはできないことは明らかである。

3.2 学生からの声

本システムが真に有効であるかどうかについては棚上げし、ここでは実用に耐え得るかどうかという側面から学生の反応を述べる。学生からの返答はすべて本学で実施している授業評価システム [10] で収集した。

3.2.1 第一回目の授業

第一回目の授業では、いわゆるアカデミックスキルとしての文献の収集の方法を説明しつつ、システムの操作方法を学んでもらった。学生への課題は「主題に即して収集した文献が、その文献で引用している文献を収集する」というものである。授業の内容に対する学生からの返答率は52%であり、返答内容は好意的であるといえる。感想は下記の通りであるが、主として文献探索法に対する評価であり、本システムに言及している返答は無かった。

- 参考文献からさらに参考文献を探すという作業が意外に大変だということがわかりました。
- 最も読むべき文献の探し方を、初めて知りました。この作業は資料集めとその整理で大変だと思いますが、今こういう経験をしておくと卒論の時にすごく役立つと思います。課題、頑張ります。
- 課題は大変ですが、卒論の資料集めの練習として頑張りたいと思います。
- レポートや論文を書く上で今回のような作業は、ぜひ活用していきたいです。宿題頑張りたいです。
- 同じようなテーマの論文を探すのが難しかったです。
- 今からこうした文献の整理法を知っておくと、役に立つと思うので、しっかり学びたいと思います。
- 参考文献がうまくリンクしてくれるか、また、期限内に多くの参考文献を探して図式化できるか少し不安ですが、できるだけ頑張ってみようと思います。
- 思ったよりたくさんの論文が集まり、私が大学生だったころと比べて、格段に情報を集めやすくなっていることに驚きました。

3.2.2 第二回目の授業

第二回目の授業では、引用関係を入力させることにした。学生からの返答率は20%であり、返答した学生の感想は好意的であったが、返答をしていない学生が80%となっている。

- 論文を入力しただけで、図が自動的に作られることに感動しました。とても便利ですね。
- 先週は、大量の文献をどう図にすればいいのか迷いましたが、あのプログラムのおかげでスムーズにできそうな気がします。
- 図が自動的にできるのは、便利で面白いです。
- 手書きでやっていくより全然簡単なので、卒論の練習としても頑張ろうと思います。
- 情報を処理する方法というのは合理的で必要な情報も形式的に排除してしまったり、弊害の多いもののような気がしていましたが、今回の参考文献は関係の薄い文献などが視覚的にはっきりわかり、きちんとした成果がでそうです。

3.3 実用性という意味での評価

返答をしなかった学生の意見を否定的なもの捉えるなら、本システムは文献探索法には不向きなシステムということになり、実用性は無いといえる。また、返答をしなかった学生が本システムに対し興味関心を持たなかったために返答をしていないのだと考えたとしても、それは学生に対し有用性を説明し切れなかったということであり彼らにとっての実用性は無かったということになるだろう。

4 今後の課題

昨年度は25人を対象とする授業であり、学生自らが主題を設定し文献探索を行わせた結果、彼らにとっては実用的では無かったようである。これは、本システムの利用が授業という場であり課題として行うという消極的な立場での評価の結果であったとみなすこともできるだろう。学生にとって切実な研究活動である卒業論文時にこのシステムが有効であるかどうかを検証する余地があると考えられる。そのため実際に卒業論文の執筆する際に本システムが実用的かどうかの検証を行いたい。

また、現在のシステムでは共引用や共著者、書誌結合といった概念を取り込んでいないため、自らが設定した主題の全体像を把握すること、すなわち分野の全体を俯瞰し知識の生成過程を調査することができない状態である。これらの概念をシステム化することができれば、当該分野における全体像の把握がより進むはずである。

参考文献

- [1] R. Ichise, H. Takeda and K. Ueyama: Exploration of Researchers' Social Network for Discovering Communities, in T. Washio, A. Sakurai, K. Nakajima, H. Takeda, S. Tojo and M. Yokoo eds., *New Frontiers in Artificial Intelligence: Joint JSAI 2005 Workshop Post-Proceedings*, Vol. 4012 of *Lecture Notes in Computer Science*, pp. 458 – 469, Springer, Berlin, Heidelberg (2006)
- [2] W.W.Hood;C.S.Willson:”The lierature of bilbliometrics, scientometrics, informetrics,”*Scientometrics* Vol.52, pp.291-314, 2001.
- [3] R.N, Broadus:Toward a definition of “BIBLIOMETRICS,”*Scientometrics* Vol.12,pp.373-379, 1987.
- [4] Pritchard, Alan: *Statistical Bibliography or Bibliometrics*, *Journal of Documentation*, 25(4), pp.348-349,1969.
- [5] 仲本 秀四郎「計量情報学-ビブリオメトリックス-」, *ドクメンテーション研究* 31(4), pp.161-168, 1981.
- [6] 倉田敬子「情報の生産、伝達、利用からみた科学研究活動」『*科学を考える*』, 北大路書房,1999.
- [7] 小野寺夏生「引用文献数で研究を評価できるか?」, *学会会報*, Vol.838, 2003.
- [8] 根岸正光「研究評価とビブリオメトリックス」, *情報の科学と技術*, Vol.49, No.11, 1999.
- [9] MacRoberts,M.;MacRoberts,B: Problems. of citation analysis. *Scientometrics*, Vol.36, p.435-444, 1996.
- [10] 永崎研宣「*全学毎回授業評価システムの開発と運用*」『*平成17年情報 処理教育研究集会講演論文集*』,2005, pp. 109-112.