

論文 / 著書情報
Article / Book Information

論題	オープンデータによる大学IR情報システムの可能性
著者	森 雅生, 船守 美穂
出典	日本教育情報学会年会論文集32, pp. 46-49
発行日	2016, 8

オープンデータによる大学IR情報システムの可能性

森 雅生*1 船守 美穂*2

<概要> 大学IRは、根拠データに基づく教育改善や大学経営の支援を意味する。特に、同じ系統の大学と自大学の比較によって、自身の位置付けを把握するには、大学内のデータのみならず、オープンデータを利用する必要がある。本研究では、オープンデータを用いた大学の比較可能性について議論する。

<キーワード> IR, オープンデータ, 大学マネジメント

1. はじめに

近年の我が国の大学、特に国立大学法人においては、ガバナンス改革が強く求められている（中教審・平成26年）。特に、データに基づく大学経営の高度化は喫緊の課題とされている。そうした課題を解決するため、Institutional Research（以後IRと略）は急速に注目を集めつつある。

IRは教学改善に活用する機能として、2000年初頭に米国の事例が紹介された（Swing 2005）。これを皮切りに、我が国において特に教学面でのIRの開発が進み、例えばFDと組み合わせて授業改善に活用されている。

また、機関別認証評価や国立大学法人評価、自主的に行われる自己点検・評価において、教育研究などの諸活動の情報を収集し、分析する必要が認識されている。特に、内部質保証の観点からは、機関が内生的な大学情報を定期的・一元的に収集し、機関の活動の目的と照らし合わせた分析のためのデータ管理を行うことが重要であるが、すべての大学でそのインフラが整備されているわけではないようである（高田ほか 2013）。

一方、機関の研究活動に関するパフォーマンス計測も重要なIRの仕事であると認識されてきている。例えば、近年では学際的領域での研究成果として論文に関する指標の分析が盛んである（Gautam 2016）。

研究活動の分析については、早くから国内外の大学や研究機関とのベンチマークやランキングについて議論がなされているが、教育に関しては内生的であり、教育目的は大学によって大きく異なるがゆえに、教育ベンチマークや単

純な比較は、管見の限り見ることが少ない。

では、教育や研究のベンチマークを行うと仮定して、どのようなデータが利用できるのだろうか。内部質保証の部分と教育の例でもあげたように、大学情報は活動の目的を明確にした上で収集するが、他の機関との比較では機関の活動目的や、業務運営での都合により、項目が整合しない場合が多い。よって、なんらかの標準的なスキームが必要となる。そこで本稿が注目したのは、大学情報に関するオープンデータ、またはそれに準じる情報である。

米国では Integrated Postsecondary Education Data System という、日本の学校基本調査（大学対象）に似た調査が毎年行われており、そのデータセットは構造化され分析に利用可能な形であり、米国 National Center for Education Statistics のホームページから一般に入手できる。日本には IPEDS のようなデータセットは未だ無いが、平成24年度から27年度の学校基本調査データで国公立大学のものが試行的にExcelのデータ形式で公表されており（大学基本情報¹）、継続してデータの公表が期待されているところである。

以下の節では、筆者らがプロトタイプとして構築した大学IRに活用する情報システムを紹介する。第2節では、オープンデータに準じるものとしてのWeb of Science®とInCites®のデータを用いた、研究力ベンチマークを行うシステムを紹介する。第3節では、上述した平成24年度の学校基本調査データを活用した国公立大学の比較を行うシステムを示す。第4節では、Twitterに投稿されたつぶやきを大学ごとにある一定数集め、そこに現れる語の頻度分析

*1 Masao Mori：東京工業大学 e-mail=mori@irds.titech.ac.jp

*2 Miho Funamori：国立情報学研究所 e-mail=funamori@nii.ac.jp

¹ <http://portal.niad.ac.jp/ptrt/table.html>

に基づいた特徴分析の試みを示す。

2. 論文関連指標による大学比較分析ツール

Web of Science®に採録されている論文书誌情報とThomson Reuterが収集している引用情報に基づく研究パフォーマンス解析ツールInCites®では、大学ごと、各年度ごとに次の指標のデータが入手で可能である。

- 論文数
- インパクト (平均引用数)
- 相対インパクト
- 世界シェア
- 被引用率
- 被引用論文の対世界平均比
- 上記を国際共著に限定した場合の数値

本ツール (図1) では、OECDの研究分野分類に基づき、選択した大学について、分野ごとにバブルチャートを表示し、3つの指標を一つのチャートで示せるようにしている。また、設定した3指標が、経年でどのように変化していくかをアニメーションで示す機能が配備されている。

このツールの特徴は、アニメーション機能であり、指定した指標と大学の組み合わせで10年間の変化を視覚的・動的に捉えられることである。

3. 学校基本調査に基づく比較ツール

平成28年7月現在、学校基本調査のデータは24年度から27年度のもものが公表されているが、本ツール (図2) では24年度のデータを用いて、次の指標について、大学を選択し比較することができる。

- 志願倍率
- 女性教員比率
- 外国籍教員比率
- 教員一人当たりの指導学生数
- 教員職位比率

このほかにも幾つかの指標が挙げられるが、最も重要なデータは、職業別就職状況や産業別就職状況、留学生の分布などが学科専攻別に集計されていること、出身高校の県庁所在地などが得られることである。

4. Twitterに顕れる大学の特徴

一般に閲覧できるツールとしては未完成であるが、オープンデータとして大学名を含むつぶやきを1000件収集し、それを一つの文書 (大学自身とみなす) として、複数の大学すなわち文書群を対象にTF-IDF法による特徴語解析を行った。TF-IDF法とは、次のように定義される。まず、ある文書 d に含まれる語 w の頻度を $tf_{w,d}$ と表し、逆に語 w に関する文書頻度を

$$df_w = \#\{d \in D \mid w \text{ appears in } d\}.$$

で表す。TF-IDF値は

$$tf_idf_{w,d} = tf_{w,d} \times \log\left(\frac{\#D}{df_w}\right)$$

によって与えられる。ただし、 $\#D$ は文書の数。

この指標値を正規化し、文書 (大学) ごとに特徴語のランクを示したものが図3である。

5. むすび

本稿では筆者らが開発したオープンデータ、またはそれに準じる情報を活用したIR情報システムを紹介した。これらはまだ開発途中であり、広く意見を求め、さらなる改良が必要である。また、これらのシステムは大学情報・機関調査研究会²の参加主要メンバーにより、共同で開発されており、希望者は開発に加わることが可能である。

参考文献

- 中央教育審議会・大学分科会「大学のガバナンス改革の推進について」(審議まとめ) 平成26年2月12日。
- Randy L. Swing著, 山田礼子 訳「米国の高等教育における IR の射程, 発展, 文脈」大学評価・学位研究, 第3号, 2005。
- Pitambar Gautam, “Comparative Analysis of Scientific Publications of Research Entities Using Multiple Disciplinary Classifications” *Proc. of Advanced Applied Informatics*, DSIR-2016, 2016。
- 高田 英一, 森 雅生, 高森智嗣, 桑野典子, 「国立大学法人におけるIRの機能・データベース・組織のあり方について—IR担当理事に対するアンケート調査結果を中心に—」 大学基準協会, 2012。

² <http://mjir.info/>

トップ10大学・工科大学 論文比較分析 OECD大分類

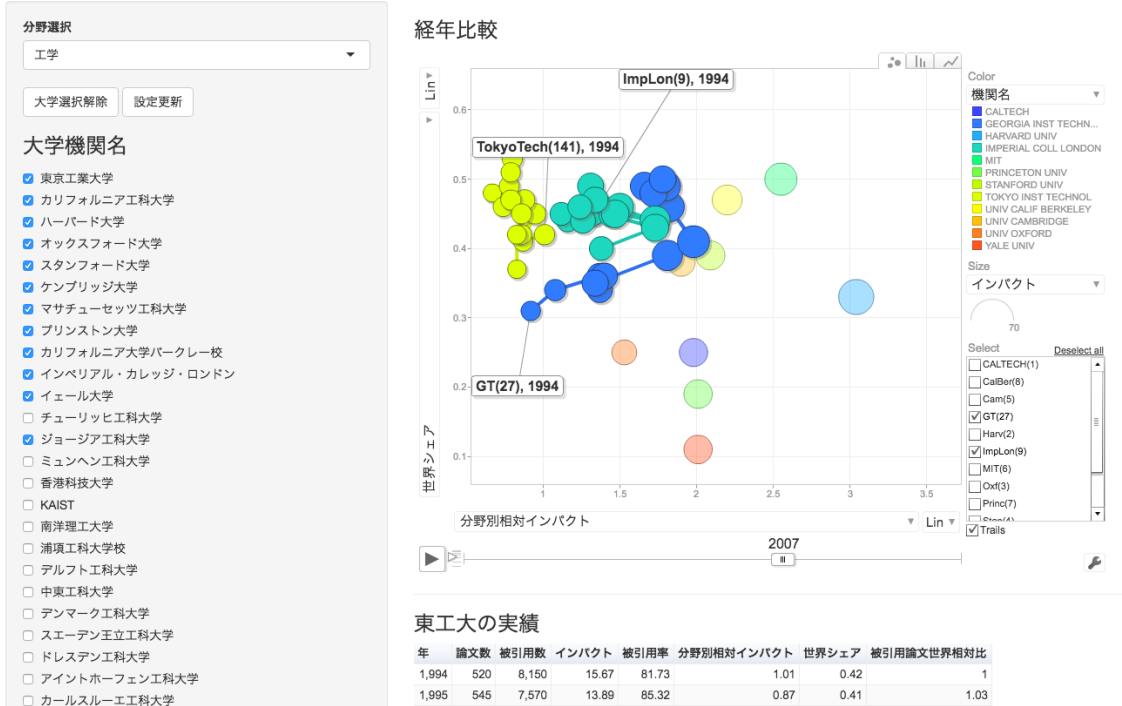


図 1 理工系大学の論文関連指標分析

平成24年度 国公立大学 学校基本調査



図 2 学校基本調査に基づく分析

