

# 機関リポジトリコンテンツの利用数とアクセス元、アクセス方法、コンテンツ属性の関係\*

佐藤翔 (学籍番号 200821656)

研究指導教員：逸村 裕

副研究指導教員：大庭一郎

## 1. 研究背景と目的

機関リポジトリとは学術・研究機関が、構成員の生産したコンテンツを収集・管理し外部に発信するシステム、あるいは一連のサービスである。近年、機関リポジトリの世界的な普及を受け、JISCによるPIRUSプロジェクト<sup>1)</sup>、千葉大学等のROATプロジェクト<sup>2)</sup>等、リポジトリで公開したコンテンツの利用状況に注目したプロジェクトが進められている。

機関リポジトリの利用状況を見る上で注目されるのは「誰が使っているのか」(アクセス元)、「どこからアクセスしているのか」(アクセス方法)、「どのようなコンテンツが利用を集めるのか」(コンテンツ属性)の3点である。これらはリポジトリ運用の参考になるとともに、機関リポジトリが果たしている役割を考える上でも重要である。しかしこれまでこれら3点を個別に見る分析は多く行われてきたが、3点を組み合わせた詳細な分析は行われてこなかった。そこで本研究ではこれらの点を明らかにすることを目的に、機関リポジトリ収録コンテンツのアクセスログ分析を行った。

## 2. 調査方法

分析対象はアジア経済研究所(ARRIDE)、北海道大学(HUSCAP)、京都大学(KURENAI)、筑波大学(Tulips-R)の4つの機関リポジトリに2008年12月31日までに収録されたコンテンツとした。それぞれのコンテンツ数は640件、25,542件、28,356件、7,899件であった。

各コンテンツについて、2008年1年間分の文献本文へのアクセスログから検索ロボットや同一人物による連続アクセス等のノイズを排除するフィルタリングを行った上で、アクセス元ドメイン(国・地域、機関種別)、アク

セス方法(参照元)ごとのアクセス数を集計し、文献タイプ、記述言語、出版年等のコンテンツ属性と合わせて分析を行った。フィルタリングの方法については佐藤義則が提案する、電子レコードの統計標準COUNTER<sup>3)</sup>に基づいた方法を採用した<sup>4)</sup>。また、各コンテンツのファイル形式(テキストデータの付与の有無、セキュリティ設定、オープンデジタルなファイルかOCR処理により生成されたファイルか)のデータも収集し、アクセス数およびコンテンツ属性と合わせて分析した。

## 3. 分析結果

### 3.1 アクセス元

アクセス元機関の種別については、いずれのリポジトリでも最も多いのは民間・プロバイダ(個人の自宅等)からのアクセスで、全体の40~50%を占めた。次いで多いのは大学等また企業からのアクセスであり、あわせて30%前後である。

アクセス元の国を見るとARRIDE以外では日本国内からのアクセスが多く、全体の6~8割を占める。これは収録コンテンツの中に日本語で書かれたものが多いためである。コンテンツの記述言語ごとに、国内/海外からのアクセス数を比べると、いずれのリポジトリでも日本語コンテンツは国内、英語コンテンツは海外からのアクセスが多く、逆は少なかった(表1)。

表1. 国内/海外からの平均アクセス数(記述言語別)

	日本語・国内	日本語・海外	英語・国内	英語・海外
ARRIDE	7.0	1.9	3.2	29.3
HUSCAP	10.3	1.5	2.1	9.3
KURENAI	14.4	2.4	1.9	10.9
Tulips-R	11.5	1.5	2.4	13.1

また、海外からのアクセスの多くは西欧等の高所得国からのものであるが、アジア・中南米等の中・低所得国からのアクセスも相当数ある。

\* "Usage log analysis of the contents of institutional repositories: user domains, types of referrals and content attributes" by Sho SATO

### 3.2 アクセス方法

アクセス方法は ARRIDE では経済学分野の分野別リポジトリ RePEc が多い。他ではサーチエンジンからのアクセスが多く、利用者のほとんどはメタデータの記述されたページを見ることなく直接論文本文にアクセスしている。また、サーチエンジンからのアクセス数はテキストデータ付与の有無で大きく上下し、テキストが付与されていないとアクセス数が顕著に少なくなる。

書かれた言語によってもアクセス方法に違いがあり、日本語文献はメタデータページ経由のアクセスも一定数あるのに対し英語ではサーチエンジンに偏っている。これは日本語は CiNii 等、メタデータページにリンクした外部サービスからのアクセスがあるのに対し、英語にはそのようなサービスが存在しないためである。

### 3.3 文献タイプ

最もアクセス数が多い文献タイプは教材で、次いで図書のアクセス数が多かった。登録コンテンツ数が多いのは雑誌論文、学位論文、紀要論文であるが、これらの文献タイプについても要旨等のみでなく本文があり、フルテキストデータが付与されている場合には多くのアクセスを集めている(表2)。特に学位論文のアクセス数が多いが、その主なアクセス方法はサーチエンジンであり、そのためテキストデータが付与されていないと全くアクセスされないコンテンツにもなっていた。

表2 文献タイプごとの平均アクセス数(HUSCAP・KURENAI、雑誌論文、学位論文、『講究録』以外の紀要論文)

	HUSCAP	KURENAI
雑誌論文	49.1	27.1
学位論文	168.1	90.7
紀要論文	54.9	38.6

### 3.4 出版年

日本国内から、大学等から、メタデータページを経由した、日本語コンテンツへのアクセスは新しい論文ほどよくアクセスされ、海外から、企業から、サーチエンジンを経由した、英語コンテンツへのアクセスは出版年と関わりなくアクセスする傾向があった。

### 3.5 ファイル形式等

前述のようにテキストデータが付与されているコンテンツは付与されていないコンテンツよりアクセス数が多く、またセキュリティ設定はかかっていないコンテンツの方がアクセス数が多い。これはサーチエンジンからアクセスしやすくなるためである。一方、テキストデータ

が付与されていなければ最初からデジタルで作成されたコンテンツでも、紙からスキャンし作成されたコンテンツでもアクセス数に差はなかった。

## 4. 考察・結論

機関リポジトリには研究機関以外にも企業、個人宅など多様な利用者からアクセスがあり、学術コミュニケーションの枠にとどまらない機能を果たしている。一方、機関リポジトリ自身はコンテンツ探索のインターフェースとなっておらず、アクセスのほとんどはサーチエンジンなどの外部サービスを介したものであり、外部サービスから使いやすいものとするのが重要である。また、機関リポジトリにはオープンアクセスの実現や大学の説明責任の確保等、多様な役割が期待されているが、

- ・サーチエンジン以外の発見手段が不十分
- ・記述言語の障壁
- ・テキストデータの不備

という3点がそれらの役割を阻害している。

最後に、機関リポジトリにおけるアクセス数の多寡はコンテンツの中身ではなく、テキストデータの付与等の運営担当者の取り組みの程度により定まる部分が多い。そのため機関リポジトリのアクセス数を研究評価に用いることは不適切であると言える。

## 文献

- [1] PIRUS: Publisher and Institutional Repository Usage Statistics. 2009, 20p., [http://www.jisc.ac.uk/media/documents/programmes/pals3/pirus\\_finalreport.pdf](http://www.jisc.ac.uk/media/documents/programmes/pals3/pirus_finalreport.pdf), (2009-10-10 入手).
- [2] 千葉大学附属図書館. “機関リポジトリ評価のための基盤構築”. <http://www.ll.chiba-u.ac.jp/~joho/CSI/standardization.html>, (2009-10-04 入手).
- [3] Counting Online Usage of Networked Electronic Resources. “The COUNTER Code of Practice. Journals and Databases. Release 3”. 2008, 38p., <http://www.projectcounter.org/r3/Release3D9.pdf>, (2009-10-10 入手).
- [4] 佐藤義則. 動向レビュー:機関リポジトリの利用統計のゆくえ. カレントアウェアネス, 2008, vol.296, p.12-16.