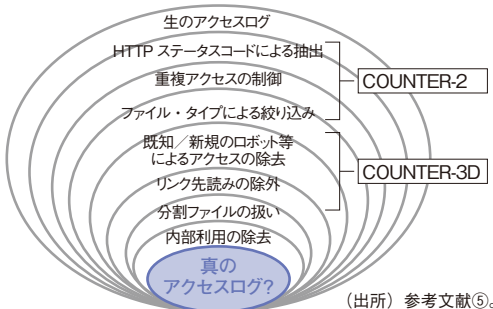


表1 フィルタリング結果

	ログ数	生のアクセスログに対する割合
生のアクセスログ	1,812,406	100.0%
HTTPステータスコードによる抽出	1,710,418	94.4%
検索ロボットのアクセスを排除	207,764	11.5%
PDFファイルに限定	13,077	0.7%
重複アクセスの制御	9,958	0.5%
ドメイン名の取得成功	5,459	0.3%

図1 アクセスログのフィルタリング



(出所) 参考文献⑤。



アクセスログから見るARRIDEの利用状況

佐藤翔・逸村裕

はじめに

電子ジャーナルをはじめとする電子的資料の利用が一般化するにつれ、近年アクセスログに基づく学術文献の利用 (usage) 分析が注目を集めている。二〇〇八年、電子ジャーナル等の利用データ記録に関する指針であるCOUNTER第三版が公開され (参考文献①)、また二〇〇六〜二〇〇八年にかけて利用データに基づき学術インパクトを評価するMESURプロジェクト (参考文献②) も実施されるなどデータや手法の標準化が盛んに試みられている。

機関リポジトリにおいてもこの傾向は同様である。英国では二〇〇五〜二〇〇七年にかけてリポジトリの利用統計に関するプロジェクトであるIRSPプロジェクトが実施され (参考文献③)、二〇〇八年にはJISCがリポジトリの利用統計に関するレビューを発表した (参考文献④)。実際に誰が、何を、どのように利用しているのかを明らかにすることは、機関リポジトリが果たしている役割を知る上で不可欠である。機関リポジトリは誰でも自由に利用可能

であるために検索ロボット等の機械的ダウンロードも多く、アクセスログのフィルタリング (検索ロボット等のノイズを排除すること) が重要となる。この点について佐藤義則は前述のCOUNTERに基づくフィルタリングについて提案を行った (参考文献⑤)。また、筆者らは北海道大学、京都大学の二つのリポジトリについて、フィルタリングを行ったログに基づいた分析を行い、サーチエンジン経由で直接本文を利用する利用者が多いこと、利用者の所属国・機関の種類により利用される文献の種類に違いがあること等を明らかにした (参考文献⑥)。

本稿では同様のフィルタリング手法を用い、アジア経済研究所学術研究リポジトリ (ARRIDE) (参考文献⑦) のアクセスログを分析した。その結果からARRIDEの利用状況を明らかにすることで、ARRIDEが国際的な研究活動の中で果たす役割について考察する。

分析手法

分析対象は二〇〇八年三月〜九月の七カ月

分のARRIDEへのアクセスログ一八万二四〇六件である。このログを前述の手法に従い、COUNTERの基準に則ってフィルタリングした (図1参照)。

ただし検索ロボット等からのアクセスとして分析から除去する対象のリストについては現在のCOUNTERでは不十分な部分が多い。そこで本稿では筆者らが過去の分析に用いた除外リストをもとに、実際のARRIDEのログを確認しながら新たに作成した除外リストを用いた。これによりノイズを排除し、人が論文本文を閲覧したアクセスログのみに特定した分析を行うことができると考えられる。

さらにアクセスしたユーザの属性 (アクセス元の国や所属機関等) を分析するために、IPアドレスからアクセス元のドメイン名の取得を試みた。

分析結果

① フィルタリングの結果

分析対象のログに対するフィルタリング結果およびドメイン名の取得成功件数は表1の通りである。

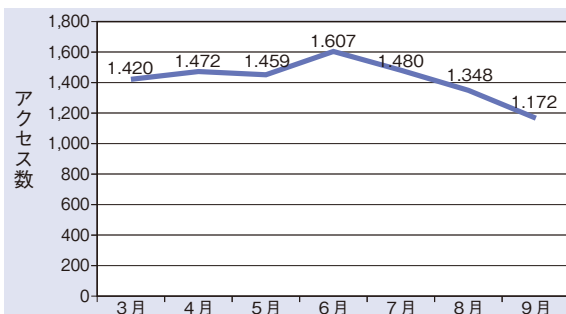
表2 アクセス元の国内／海外比

	アクセス数	割合
国内 (jp)	1,531	28.0%
海外 (jp以外)	3,928	72.0%
合計	5,459	

表3 アクセスの参照元

	ARRIDE		(参照)北海道大学		(参照)京都大学	
参照元なし(記録なし)	4,028	40.4%	28,005	22.6%	105,453	21.7%
リポジトリ内部	1,067	10.7%	33,930	27.4%	111,985	23.0%
サーチエンジン経由	1,317	13.2%	60,726	49.1%	265,784	54.7%
うちGoogle	1,098	11.0%	54,629	44.1%	235,599	48.5%
その他の参照元	3,546	35.6%	1,140	0.9%	2,820	0.6%
うちRePEc	3,278	32.9%				
うちwebメール	150	1.5%				
合計	9,958		123,801		486,042	

図2 月ごとのアクセス数の推移



生のアクセスログが一八〇万件以上あったのに対し、重複アクセス等の制御を行った結果、人が文献を閲覧したと考えられるログは九九五八件であり、全体の〇・五%であった。検索ロボット等の機械的アクセスが多く、ログ全体の八〇%以上にのぼる。以下の分析では、ドメイン名が必要な場合はその取得に成功したログ(五四五九件)を、それ以外は重複アクセスの制御までを行ったログ(九九五八件)を用いる。

②月ごとのアクセス数の推移

図2は分析期間内の月ごとのアクセス数の推移を示したものである。期間中、月ごとのアクセス数は安定している。ARRIDEが定常的に利用されている様子が窺える。

③アクセス元のドメイン

ドメイン名の取得に成功した五四五九件のログについて、アクセス元の国等を示すトップレベルドメインについての分析を行った。

トップレベルドメインについては一一八の異なるドメインからのアクセスがあり、日本からのアクセスと"com"、"net"などのアクセス元の国を示さないドメインを除くと、イギリス(一六九件)、オーストラリア(一六八件)、インド(二五二件)、インドネシア(二〇六件)、マレーシア(九七件)からの利用が多い。

表2はアクセス元の国内／海外比を示したものである。筆者らが過去に分析した北

海道大学、京都大学のリポジトリではないずれも国内からのアクセスが全体の過半数を占めたのに対し、ARRIDEへのアクセスのうち国内(JPドメイン)からは二八%にとどまった。ARRIDEは海外ユーザーからよく利用されていると考えられる。

④参照元

表3はARRIDEへのアクセスの参照元(当該ページにたどり着く直前に経由したwebページ)について示したものである。比較のため、北海道大学、京都大学のリポジトリでの分析結果についても併せて示してある(参考文献⑥)。

参照元が記録されていない(直接URLを入力、あるいはPDF文書等webページ以外のリンクからアクセス)件数を除くと、多いのは経済学分野のプレプリントサーバであるRePEc(参考文献⑧)からARRIDEにきた場合であり、アクセス全体の約三割を占める。分析対象期間が異なるため比較は難しいが、他の二リポジトリではサーチエンジン経由のアクセスが半数前後を占め、次いでリポジトリ内部からのアクセスが多いことはARRIDE利用状況の特徴であると言える。

また、ARRIDEではwebメール経由のアクセスが一・五%存在し、これは他のリポジトリにはない特徴である。RePEcのメールアラートサービス等からアクセスしていることが考えられる。

⑤アクセス先コンテンツ

表4はアクセスの多かったコンテンツ上位一〇件について、掲載誌とコンテンツ種別を示したものである。"IDE Discussion Paper"が八件を占め、この傾向は上位五〇件まで見ても同様で、五〇件中三二件が"IDE Discussion Paper"掲載論文であった。

さらに詳細を見るため、アクセス元ドメインの国内(JP)・海外(非JP)別にアクセスの多かったコンテンツ上位一〇件を見たと表5、6である。国内・海外双方で上位一〇件に入っている論文はない。

これは上位二〇件まで見ても同様である。国内からのアクセスと海外からのそれとは利用傾向が異なっている。海外からのアクセスは"IDE Discussion Paper"に集中している。同紙はアジア経済研究所が発行する論文草稿で、全号がRePEcに

表4 アクセス数上位10件の資料タイプと掲載誌

順位	アクセス数	タイプ	掲載誌等
1	251	ワーキングペーパー	IDE Discussion Paper. No. 101. 2007.4
2	174	ワーキングペーパー	IDE Discussion Paper. No. 103. 2007.4
3	163	ワーキングペーパー	IDE Discussion Paper. No. 86. 2007.1
4	151	雑誌論文	社会科学研究 55.5-6 (2004.3): 101-129
5	150	ワーキングペーパー	IDE Discussion Paper. No. 128. 2007.10
6	140	ワーキングペーパー	IDE Discussion Paper. No. 53. 2006.3
7	132	その他	The Developing Economies 43.4 (2005.12): 519-522
8	130	ワーキングペーパー	IDE Discussion Paper. No. 120. 2007.9
9	126	ワーキングペーパー	IDE Discussion Paper. No. 137. 2008.3
10	116	ワーキングペーパー	IDE Discussion Paper. No. 116. 2007.8

表6 海外(jp以外のドメイン)からのアクセス数上位10位

順位	アクセス数	タイプ	掲載誌等
1	93	ワーキングペーパー	IDE Discussion Paper. No. 86. 2007.1
2	91	ワーキングペーパー	IDE Discussion Paper. No. 103. 2007.4
3	81	ワーキングペーパー	IDE Discussion Paper. No. 128. 2007.10
4	73	その他	The Developing Economies 43.4 (2005.12): 519-522
5	69	ワーキングペーパー	IDE Discussion Paper. No. 137. 2008.3
6	67	ワーキングペーパー	IDE Discussion Paper. No. 70. 2006.8
7	62	ワーキングペーパー	IDE Discussion Paper. No. 154. 2008.5
8	56	ワーキングペーパー	IDE Discussion Paper. No. 120. 2007.9
9	54	ワーキングペーパー	IDE Discussion Paper. No. 100. 2007.4
10	53	ワーキングペーパー	IDE Discussion Paper. No. 109. 2007.6

表5 国内(jpドメイン)からのアクセス数上位10件

順位	アクセス数	タイプ	掲載誌等
1	71	雑誌論文	社会科学研究 55.5-6 (2004.3): 101-129
2	45	雑誌論文	アフリカ研究 58 (2001.3): 41-58
3	44	雑誌論文	東亜 第463号 (2006.1): 68-79
4	41	雑誌論文	中国研究月報 631 (2000.9): 1-17
5	37	図書	『貧困削減と人間の安全保障』東京、国際協力機構国際協力総合研修所、2005年11月、121-137ページ
6	31	雑誌論文	SPFオンデマンドレポート集 Voices from the World 東京、笹川平和財団、2006年
7	30	雑誌論文	イスラム世界 66 (2006.3): 74-96
7	30	図書	『シリーズ国際開発第1巻 貧困と開発』絵所秀紀・穂坂光彦・野上裕生編 東京、日本評論社、2004年、21-36ページ
9	26	図書	『民族の二〇世紀』端信行編 東京、ドメス出版、2004年、22-39ページ
10	25	雑誌論文	国民経済雑誌 193.1 (2006.1): 1-16

登録され、RePEcから直接ARRIDE掲載コンテンツがダウンロードできるようなっているため、RePEcからのアクセスが多いものと考えられる(海外からのアクセス四位に入っている「The Developing Economies」も同じくRePEc登録コンテンツ)。

一方、国内からのアクセスでは上位一〇件は全て日本語文献であり、七件が雑誌論文、図書(の一部)が三件である。

●考察と今後の課題

フィルタリングを行った後のアクセスログ分析の結果から、ARRIDEへのアクセスではRePEcを経由したアクセスが多く、利用されるコンテンツも「IDE Discussion Paper」などRePEcに登録された英語コンテンツであることがわかった。

ARRIDE登録文献をRePEcにも併せて登録したことで、海外から多くの利用を得ており、国際的な研究活動に貢献しているのではないかと考えられる。

一方で、量はやや少ないが国内からは日本語の雑誌論文や図書に対する利用もある。今後はドメイン名の取得方法の改善やアクセス先のコンテンツについてのより詳細な(利用の多いものに限らず全件を対象とするような)分析を行うこと、被引用数などアクセス数以外の指標との比較を行うこと等が必要であろう。

また、他の機関リポジトリとの比較・分析を行うことで、機関リポジトリによるオープンアクセスコンテンツが果たしている役割についても明らかにできると考えられ、同様の分析が拡大深化していくことが望まれる。

(おとう) しょう／筑波大学大学院図書館情報メディア研究科、いづむらひろし／同教授)

〔付記〕本報告は「科学研究費補助金(基盤研究(C) 機関リポジトリへの登録が学術文献流通に及ぼす効果についての定量的分析)の支援により行われたものである。

《参考文献》

- ① 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>. (2008-12-13 入手).
- ② MESUR, "MESUR: Metrics from Scholarly Usage of Resources". <http://www.mesur.org/MESUR.html>. (2008-12-13 入手).

org/MESUR.html. (2008-12-13 入手).

- ③ Interoperable Repository Statistics, "About the Project". <http://irs.eprints.org/about.html>. (2008-12-13 入手).
- ④ Merk, Christine, "Final Report JISC Usage Statistics Review". JISC. (<http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/usagestatisticsreviewreport.pdf>). (2008-12-13 入手).
- ⑤ 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ん」(『カレントアウェアネス』二九六号、二〇〇八年)。
- ⑥ 佐藤翔「誰が、何を読んでいるのか—アクセスログに基づく機関リポジトリの利用実態」SPARC-Japan セミナー二〇〇八年「日本における最適なオープンアクセスとは何か。」東京、二〇〇八年一〇月一四日 SPARC-Japan. <http://www.ni.ac.jp/sparc/event/2008/20081014.html>. (2008-12-13 入手).
- ⑦ 「アジア経済研究所学術研究リポジトリ」<https://ir.ide.go.jp/dspace/index.html>. (2008-12-13 入手).
- ⑧ "RePEc". <http://repec.org/>. (2008-12-13 入手). RePEcを利用できるコンテンツは複数あるが、ARRIDEではIDEAS (<http://ideas.repec.org/>) からアクセスが主であり、次ぐEconPapers (<http://econpapers.repec.org/>) からアクセスが多かった。