CiNii−機関リポジトリ連携の有効性の検証

佐藤翔（筑波大学大学院図書館情報メディア研究科）　min2fly@slis.tsukuba.ac.jp
逸村裕（筑波大学大学院図書館情報メディア研究科）　hits@slis.tsukuba.ac.jp

抄録

本研究では、学術論文データベース CiNii と機関リポジトリ（IR）の連携の有効性を検証することを目的に、北海道大学、京都大学の IR の 2008-2009 年のアクセスログを分析した。分析の結果、Google から IR を利用する者は論文本文までアクセスする割合が 25%前後であるのに対し、CiNii からの利用者は 75%以上が本文までアクセスしていた。CiNii からの利用者は他のアクセス方法に比べ大学等の所属者の割合も多く、学術論文の探索者に対し CiNii-IR 連携が本文への誘導路を提供していることが確認された。

1. はじめに

大学・研究機関での機関リポジトリ（IR）の設置と収録文献の充実により、その利用状況への注目が高まっている。佐藤らはこれまでに複数の IR のアクセスログから、IR の利用者・利用方法を分析し、利用の多くは大学・研究機関等ではなく自宅等からのアクセスであること、主なアクセス方法はサーチエンジンで、検索結果画面から直接論文本文にアクセスしていること等を明らかにしてきた[1)]。一方でサーチエンジンからのアクセスが多い文献と他のアクセス方法からのアクセスが多い文献は異なる傾向があり[2)3)]、その中で日本語文献については、国立情報学研究所が運営する学術論文データベース CiNii[1]がアクセス経路として大きな役割を果たしている可能性を指摘してきた[1)4)]。

CiNii は書誌情報の検索機能だけではなく、IR や J-Stage 等に収録された文献ともリンクし、論文本文に利用者を誘導する機能を有する。2009 年には月当たりの検索回数が 600 万回、本文アクセス数が 300 万回を超える月もあり、日本の学術情報流通において重要な役割を果たすものとなっている[5)]。CiNii と IR の連携（リンク提供）が始まったのは 2008 年 10 月である。この連携により CiNii は IR に利用者を導く有力なゲートウェイの一つとなったと考えられるが、これまでこのような観点からの研究は行われておらず、CiNii-IR 連携の有効性は明らかにされてはいない。そこで本研究では、北海道大学、京都大学の IR のアクセスログの詳細分析から、IR のゲートウィとしての CiNii の有効性を検証する。

2. 対象・方法

2. 1　分析対象

本研究の分析対象は、北海道大学（HUSCAP）[2]、京都大学（KURENAI）[3]の 2 機関の IR の、2008-2009 年のアクセスログである。

期間中の総アクセス数は HUSCAP 36,632,568 回、KURENAI 42,425,359 回である。ここからバナー画像等の文献以外のファイルへのアクセスや、サーチエンジンの検索ロボットのアクセス等を除いたものが分析対象のアクセスログとなる。

2. 2　分析方法

本研究で主に分析するのは IR 文献へのアクセス方法（どこからリンクをたどり文献を読んだか）と、利用者の所属機関種別（どのような機関からのアクセスか）である。

佐藤らの従来の研究では、文献本文ファイルにアクセスしたログのみを分析対象としてきた[1)2)3)4)]。これは本文の利用状況が主たる興味の対象であったためである。しかし図 1 に示した通り、CiNii からのアクセスは直接本文にアクセスするのではなく、文献のメタデータが記述されたページ（メタデータページ）を経由する。このようなアクセスは本文ファイルへのログの中では「メタデータページからのアクセス」として記録されるため、従来の方法では CiNii からのアクセスの詳細は検証できなかった。

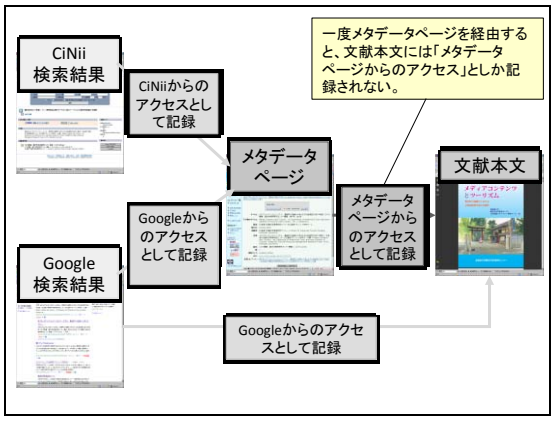

図1. メタデータページ・文献本文とアクセス方法の記録の関係

そこで本研究では文献本文ファイルだけではなく、メタデータページへのアクセスログも分析する。本文ファイルのアクセス方法としてメタデータページが記録されていた場合、当該利用者のメタデータページへのアクセスログに遡ってアクセス方法を確認し、「ある方法でメタデータにアクセスし、さらに本

---

[1] http://ci.nii.ac.jp/
[2] http://eprints.lib.hokudai.ac.jp/dspace/
[3] http://repository.kulib.kyoto-u.ac.jp/dspace/

文までアクセスした回数」（A）を特定する。一方、メタデータページのアクセス数もアクセス方法ごとに集計することで、「ある方法でメタデータにアクセスした回数」（B）が得られる。A を B で割った値が「あるアクセス方法からのリポジトリへのアクセスのうち、本文までアクセスした割合」となる。これを本研究では、そのアクセス方法の「本文到達率」と呼ぶ（図２）。

$$\text{あるアクセス方法の本文到達率} = \frac{\text{そのうち本文ファイルまでアクセスした回数（A）}}{\text{そのアクセス方法でメタデータにアクセスした回数（B）}}$$

図2. 本文到達率の計算方法

この本文到達率を、アクセス数等と併せてアクセス方法ごとに比較することで、それぞれのアクセス方法の IR 文献へのゲートウェイとしての特徴が明らかにできる。例えばメタデータへのアクセス数が同様のアクセス方法であっても、本文到達率が高ければその論文を本当に読みたいと考えている利用者をメタデータページに誘導しているのに対し、本文到達率が低ければ、必ずしも論文が読みたいわけではない利用者をメタデータに「誤誘導してしまっている」可能性もある、と考えられる。

なお、前述のように IR へのアクセス方法で最も多いのはサーチエンジンの検索結果から、メタデータを介さず直接文献本文にアクセスするケースであるが、このような場合には本文到達率は計算できない。また、本稿では同一文献への同一利用者（同一 IP アドレス）からのアクセスについては１回目のみを分析対象とし、２回目以降は分析から除外している（いわゆる「ユニークアクセス」のみ分析対象とする）。これは一つのメタデータに複数ファイルが登録されているケース等の影響を排除するためである。

3. 結果

3. 1 本文ファイルへのアクセスの概況

表１は両 IR の本文ファイルへのアクセス数についてアクセス方法ごとの概況を示したもの、表２は同じく利用者の所属機関種別ごとに示したものである。HUSCAP の 2008-2009 年の総ユニークアクセス数は 747,033 回、KURENAI は 1,163,475 回であった。

表1. 本文ファイルへのアクセス方法

	HUSCAP		KURENAI	
	アクセス数	%	アクセス数	%
サーチエンジン	381,436	51.1%	728,890	62.6%
メタデータページ	248,293	33.2%	239,287	20.6%
その他の IR 内ページ	23,484	3.1%	34,573	3.0%
直接アクセス	73,196	9.8%	128,158	11.0%
その他の外部サイト	20,624	2.8%	32,567	2.8%
合計	747,033	−	1,163,475	−

アクセス方法では、サーチエンジンの検索結果画面から直接本文ファイルにアクセスするケースが両 IR で最も多く（HUSCAP 51.1%／KURENAI 62.6%）、次いでメタデータページからのアクセスが多い（33.2%／20.6%）。

表2. 本文ファイルへのアクセス元利用者所属機関

	HUSCAP		KURENAI	
	アクセス数	%	アクセス数	%
大学等（ac/edu）	114,353	15.3%	143,744	12.4%
企業等（co/com）	61,577	8.2%	101,369	8.7%
自宅等（ne/net）	264,159	35.4%	445,801	38.3%
その他	131,965	17.7%	187,543	16.1%
ドメイン不明	174,979	23.4%	285,018	24.5%
合計	747,033	−	1,163,475	−

アクセス元機関では、「その他」とドメイン不明なケースを除くと、自宅等からのアクセスが最も多く（35.4%／38.3%）、次いで大学等からのアクセスが多い（15.3%／12.4%）。

以下ではこのうち表１のメタデータページから本文へのアクセスについて、メタデータページへのアクセス状況と比較しながら分析を行う。

3. 2 本文到達率

表３は両 IR のメタデータページへのアクセス数（アクセス方法が記録されている場合のみ）、そのうち本文までアクセスした数、本文到達率について、主要なアクセス方法ごとに示したものである。

表3. 各アクセス方法の本文到達率等

HUSCAP	メタデータへのアクセス数	うち本文アクセス数	本文到達率
Google	250,543	55,283	22.1%
IR 内の別ページ	246,863	86,638	35.1%
CiNii	71,517	56,077	78.4%
Yahoo!	25,930	4,479	17.3%
Scientific Commons	6,110	4,283	70.1%
Biblioteca	20,127	11,294	56.1%
全体	716,231	248,293	34.7%
KURENAI			
Google	376,063	94,591	25.2%
IR 内の別ページ	358,392	95,002	26.5%
CiNii	47,241	35,577	75.3%
Yahoo!	27,225	2,231	8.2%
Scientific Commons	7,034	3,691	52.5%
Biblioteca	0	0	−
全体	895,972	239,287	26.7%

全体では、メタデータページへのアクセス数は HUSCAP 716,231 回／KURENAI 895,972 回で、HUSCAP では本文ファイルへのアクセス数（表１参照）に匹敵する程度、KURENAI でも本文へのアクセスの 77%程度に相当する。しかし本文到達率は 34.7%／26.7%にとどまり、メタデータから本文にまでアクセスする利用者は少数派である。

主なアクセス方法別では、メタデータへのアクセスが最も多いのは Google である（HUSCAP

250,543 回／KURENAI 376,063 回)。しかし本文到達率は 22.1%／25.2%で、全体の値よりも低い。次にメタデータへのアクセスが多いのはメタデータページ以外の IR 内のページ（IR 内検索機能やブラウジング画面等）であるが、これも本文到達率は 35.1%／26.5%である。その次に多いのが CiNii からのアクセスであるが、HUSCAP 71,517 回／KURENAI 47,241 回と、メタデータへのアクセス数は Google に比べ１桁落ちる。しかし本文到達率は 78.4%／75.3%で、Google の３～４倍に達し、その結果 HUSCAP では Google からのアクセスよりも CiNii からのアクセスの方が本文到達数が多くなっている。他には Yahoo!（25,930 回／27,225 回）、Scientific Commons[4]（6,110 回／7,034 回）、Biblioteca[5]（HUSCAP のみ 20,127 回）等からメタデータへのアクセス数が多いが、Yahoo!は Google よりも本文到達率が低い（17.3%／8.2%）。Scientific Commons（70.1%／52.5%）、Biblioteca（HUSCAP のみ 56.1%）は本文到達率が高いが、CiNii には及ばない。主要なアクセス方法の中では CiNii の本文到達率が最も高く、論文を読みたいと考えている利用者を確実に IR に導く誘導路になっていると言える。一方、Google は多くの利用者を IR に導いているが、その中には必ずしも論文を読みたいわけではない者も多く含まれていると考えられる。

３．３ 利用者の所属機関種別

表４はメタデータページから本文まで到達した利用者について、アクセス方法ごとに利用者の所属機関（ドメイン）を見たものである。

表4. アクセス方法別の利用者の所属機関種別

	大学等 (ac/edu)		企業等 (co/com)		自宅等 (ne/net)		その他	
HUSCAP	数	%	数	%	数	%	数	%
Google	10,997	24.8%	4,359	9.8%	20,077	45.3%	8,924	20.1%
CiNii	16,832	38.4%	1,377	3.1%	19,176	43.8%	6,398	14.6%
Yahoo!	680	18.4%	280	7.6%	2,028	54.8%	716	19.3%
Scientific Commons	841	25.7%	235	7.2%	1,369	41.8%	832	25.4%
Bibliotheca	1,443	16.8%	1,145	13.4%	3,669	42.8%	2,315	27.0%
KURENAI								
Google	21,334	28.6%	6,404	8.6%	32,859	44.0%	14,008	18.8%
CiNii	9,778	35.1%	1,143	4.1%	12,348	44.4%	4,569	16.4%
Yahoo!	254	13.5%	108	5.7%	1,178	62.5%	344	18.3%
Scientific Commons	576	21.1%	234	8.6%	1,162	42.6%	753	27.6%
Bibliotheca	－	－	－	－	－	－	－	－

全体に、メタデータを見てから本文にアクセスする利用者の方が、直接本文にアクセスする利用者（表２参照）よりも大学等からのアクセス（ac/edu ドメイン）の割合が高い。その中でも CiNii については、他のアクセス方法よりも大学等からのアクセスが多くなっている。CiNii が研究者・学生を IR の文献本文へ誘導する役割を担っていることがわかる。

もっとも、その CiNii にあっても最も多いのは自宅等からのアクセス（ne/net ドメイン）である。CiNii リニューアル等を担当してきた大向は、CiNii を研究者・学生だけではなく、一般ユーザに使われるものにすべく施策を行ってきたとしているが[5] [6]、本稿の結果からも研究者以外の利用が実際に多いことがわかる（ただし、ne/net ドメインの中には研究者・学生等の自宅からのアクセスが一定数含まれている可能性がある）。

３．４ Google/CiNii からのアクセス数の推移

図３、図４は IR ごとに、特にアクセス数の多いアクセス方法である Google と CiNii について、分析期間中のメタデータページへのアクセス数と、そこから本文へアクセスした回数をグラフに示したものである。

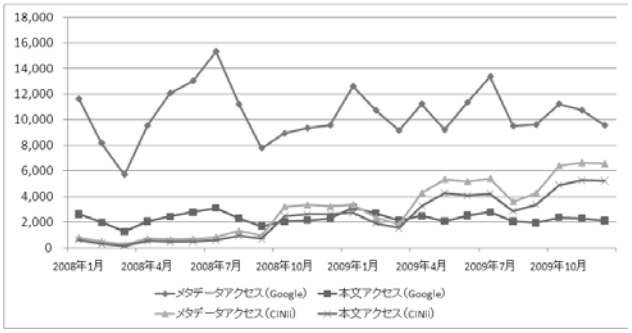

図3. Google・CiNii のアクセス数推移（HUSCAP）

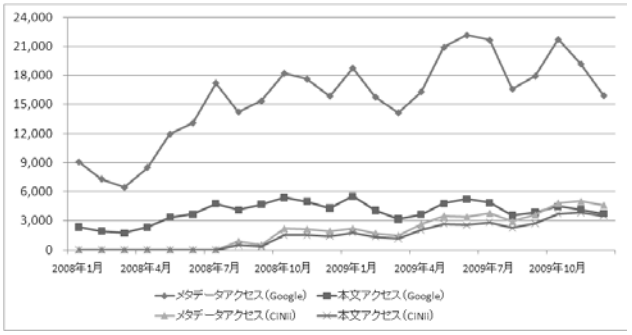

図4. Google・CiNii のアクセス数推移（KURENAI）

HUSCAP、KURENAI とも 2008 年 10 月の正式連携開始から CiNii からのアクセス数が増え始め、その後 2009 年４月の CiNii リニューアル時にも大きく増加している[6]。2009 年 12 月には、CiNii 経由

[4] http://www.scientificcommons.org/
[5] http://biblioteca.universia.net/

[6] HUSCAP では CiNii 側で連携実験を開始する 2008 年８月以前から、CiNii からのアクセスがあるが、これは北海道大学の担当者が一部の紀要等について、連携実験開始以前から手作業により HUSCAP へのリンク情報を CiNii に渡していたためである。

のアクセスが両 IR において 2008 年 12 月の２倍以上になっており、かつ一貫して 75～80%程度が文献本文までアクセスしている。この間、Google からのアクセスはメタデータへのアクセス数、本文到達数とも一定水準で推移している。その結果、2009 年末時点で、HUSCAP では CiNii からの本文到達数は Google の２倍以上に、KURENAI でも Google に匹敵する値になっている。

CiNii 連携により IR へのアクセス数が増え、さらに CiNii 側のリニューアル等の要因でもアクセス数は増加し続けている。一方で本文到達率は高水準を保っており、文献を読みたい人を本文へ誘導する、という点での CiNii 連携の有効性は高まっている。

4. 考察と今後の展望

本研究の分析結果から、CiNii は本文到達率が高く、大学・研究機関等からの利用割合も高いなど、学術文献を求める読者に確実に IR 収録文献を届けるゲートウェイとなっていることが示された。CiNii 経由のアクセス数は増加傾向にあり、CiNii-IR 連携の重要性は今後さらに高まるとも考えられる。

他方、Google は多くの利用者を IR に導いているものの、本文到達率は低い。必ずしも学術文献を求めているわけではない利用者が検索結果からアクセスし、求めている内容と異なったために引き返す場合が多いと考えられる[7]。ただしそれでも 25%前後の利用者は本文まで閲覧しており、結果として本文到達数自体は他のアクセス方法に比べても多い。必ずしも学術文献を求めているわけではない利用者が、IR 収録文献に触れる機会を提供しているとも言える。CiNii-IR 連携が主として最初から学術文献を求めている利用者に対するものであるのに対し、Google は意図せぬ文献との接触機会を提供するものであり、異なる性格を持つゲートウェイとしてそれぞれ重要であると考えられる。

また、これまでの研究から英語文献については日本語文献以上に Google からのアクセスに依存しており、CiNii のようなゲートウェイが存在しないことが問題として指摘されてきたが[1)4)]、本研究の結果からは Biblioteca、Scientific Commons 等の海外サービスからのリンク例が存在し、いずれも高い本文到達率を示していた。他の海外学術サービス等との連携も積極的に模索することで、CiNii 同様に学術文献を求める利用者に対し IR 収録文献を確実に届けることができるようになるだろう。

今後の展望として、本研究では IR 側のアクセスログのみを分析対象としたが、今後は連携元である CiNii についても分析することで、利用者の行動と CiNii、IR が果たしている役割についてより詳細に明らかにすることができると考えられる。

[7] もっとも、利用者がメタデータページで本文へのリンクを発見できなかったり、検索結果画面に戻ってそこから直接本文ファイルへのリンクをクリックした可能性もある。これらの可能性については今後、利用者実験等を通じて確認する必要がある。

謝辞

本研究は「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す。データをご提供いただいた各機関、検索ロボット排除リストをご提供いただいた東北学院大学の佐藤義則教授に感謝いたします。

引用文献

1) 佐藤翔, 逸村裕. “機関リポジトリ収録コンテンツにおける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”. 三田図書館・情報学会 2009 年研究大会. 東京, 2009-09-26, 三田図書館情報学会, 2009, p.9-12. http://hdl.handle.net/2241/103921, (参照 2010-08-19).

2) Sato, Sho; Itsumura, Hiroshi. “RePEc’s impact on the use of working papers in institutional repositories”. DRFIC2009. Tokyo, 2009-12-03/04, DRF, 2009. http://hdl.handle.net/2241/104138, (参照 2010-08-19).

3) Sato, Sho; Nagai, Yuko; Koga, Takashi; Sugita, Shigeki; Saito, Mika; Itsumura, Hiroshi. "ZS Project: Zoological Science meets Institutional Repositories". IFLA Satellite Pre-Conference Open Access to Scientific Information. Chania, Greece, 2010-08-06/08, IFLA’s Science and Technology & Health and Biosciences Libraries Sections, 2010. http://hdl.handle.net/2241/106053, (参照 2010-08-19).

4) 佐藤翔. "リポジトリログ分析による学術情報流通の諸側面". DRF-KanNihonkai 「大学図書館のミッションと機関リポジトリ : 環日本海域からの情報発信」. 金沢, 2010-02-17, 金沢大学附属図書館, 2010. http://hdl.handle.net/2241/104513, (参照 2010-08-19).

5) 大向一輝. “ボーン・デジタルの情報学 第 5 回 : CiNii の挑戦”. artscape. http://artscape.jp/study/born-digital/1213228_2772.html, (参照 2010-08-19).

6) 日詰梨恵, 逸村裕. CiNii 収録率から見たわが国の学術情報電子化の現状：人文学４領域を対象に. 中部図書館情報学会誌. 2010, vol.50, p.19-36. http://hdl.handle.net/2241/105270, (参照 2010-08-19).