アクセスログに基づくDOIリンクの参照状況の分析: JaLC DOIを対象に

吉川次郎 (筑波大学大学院) jiro@slis.tsukuba.ac.jp
高久雅生 (筑波大学図書館情報メディア系) masao@slis.tsukuba.ac.jp
武田英明 (国立情報学研究所情報学プリンシプル研究系) takeda@nii.ac.jp
逸村裕 (筑波大学図書館情報メディア系) hits@slis.tsukuba.ac.jp

1 はじめに

学術情報流通の電子化に伴い，誰もがウェブを通じて学術情報を容易に即時入手可能な環境が提供されている。このような環境において，電子的資源の同定識別に不可欠な存在がデジタル識別子であり，その1つが，解決可能，持続可能，相互運用可能なリンクを提供するための仕組みである「DOI(Digital Object Identifier，デジタルオブジェクト識別子)」である。

DOIは,「10.」で始まるPrefix,「/」(スラッシュ), Suffix，で構成されており,「http://dx.doi.org/」または「http://doi.org/」の後方に加えることでURLとして機能し，当該コンテンツのURLへのリダイレクトが行われる。本研究では，このURLを通じたハイパーリンクを「DOIリンク」と定義する。

DOIの運用は，IDF(The International DOI Foundation，国際DOI財団) [1]，RA(Registration Agency，DOI登録機関), Registrant(DOI付与組織), の3層構造で行われている。2015年10月時点でRAは9機関であり，世界最大規模のRAであるCrossRef [2]，2012年に発足した日本国内唯一のRAである Japan Link Center [3](以下，JaLC) などがある。それぞれのRAによって直接登録されたDOIをCrossRef DOI，JaLC DOIと呼ぶ。RAはRegistrantにPrefixを割り与え，RegistrantはDOI付与対象のSuffixを設定し，RAを通じてDOIの登録を行なう。2015年10月時点でDOIの総登録件数は約1億件であり，そのうち，CrossRef DOIは約7,500万件，JaLC DOIは約140万件である。

DOIリンクは学術論文での引用に限らず，ウェブ上のあらゆる場所から参照可能である。ウェブ上でのDOIリンクの参照について，CrossRefのブログでの報告 [4] によると，CrossRef DOIの参照元のうち8番目に大きなウェブサイトがフリー百科事典のWikipediaであり，Wikipediaの利用者はDOIリンクをクリックして閲覧している。

日本語版Wikipediaにおける DOIリンクの分析事例としては，吉川らが挙げられる。吉川ら [5] は2015年3月時点の日本語版Wikipediaの標準名前空間ページに含まれる27,201件のDOIリンクを分析し，CrossRef DOIが97%，JaLC DOIが2%があること，日本国外の大手出版社が多く，雑誌タイトルレベルではNature，Science，PNASなどの自然科学分野の有力誌が多く含まれていることを明らかにした。吉川ら [6] は日本語版と英語版Wikipediaにおける DOIリンクの重複状況の分析から，DOIリンクが記述されている日本語版の項目の約94%に英語版項目への言語間リンクが設定されており，それらの項目間での共通のDOIリンクは，英語版の翻訳を通じて日本語版に記述されたものが大部分を占めることを示唆する結果を示した。

以上から，CrossRef DOIリンクの参照状況や日本語版Wikipediaにおける DOIリンクの分析事例がある。しかし，日本国内の学術情報であるJaLC DOIリンクの参照状況に関する分析は行なわれていない。そこで本研究はアクセスログの分析を通じて，JaLC DOIリンクについて，(1) どのようなコンテンツが，(2) どのような場所から参照されているか，について明らかにする。

2 対象・方法

2.1 分析対象

DOIリンクのアクセスログは,「http://dx.doi.org/」または「http://doi.org/」を通じてコンテンツへのリダイレクト処理が行われる際に，いつ，誰が，どのコンテンツにリクエストを行ったか，リクエスト結果の成否，などを記録したデータ群である。

本研究では2014年4月から2015年9月までの期間における JaLC DOIリンクのアクセスログの分析を行なう。アクセスログに記録されているデータは

3,096,959 件であり，そのうち，人間によるアクセス (以下，実アクセス)1,387,321 件を分析対象とする。

JaLC DOI の総登録件数は，2015 年 10 月 19 日の時点で 1,401,149 件である。登録件数が 10,000 件以上のプラットフォームは，J-STAGE(743,648 件)，MedicalFinder(398,892 件)，国立国会図書館デジタルコレクション (237,691 件) の 3 つである。国立国会図書館デジタルコレクションの DOI 登録コンテンツは，国立国会図書館がデジタル化した学位論文 (約 14 万件) と古典籍，貴重書等 (約 9 万件) である。Prefix 単位での登録件数の上位 3 件の Registrant は，「医学書院」(398,892 件)，「国立国会図書館」(238,537 件)，「日本理学療法士協会」(15,370 件) である。

2.2 分析方法

アクセスログの要素のうち，「利用者 IP アドレス (Source IP Address)」，「日時 (Date)」，「アクセスリクエストが行なわれた DOI リンク (Requested Handle)」，「リファラ (Referer URL)」を分析に使用した。

アクセスログに記録されているアクセスのうち，サーチエンジンのロボットなどの人間以外によるアクセスの除外を行い，実アクセスの特定を行った。この特定には User Agent を用いる場合があるが，アクセスログに User Agent が含まれていないため，アクセス元 IP アドレスをもとに DNS の逆引きでホスト名を取得し，ロボット等によるアクセスの特定，除外を行った。除外したデータは，「google-bot.com」，「crawl.baidu.com」，「yse.yahoo.net」，「crawl.yahoo.net」，「search.msn.com」，「twttr.com」のいずれかをホスト名に含むものと JaLC 内部での確認のためのアクセスである。JaLC DOI リンクのアクセス件数および実アクセス件数を表 1 に示す。

表 1: JaLC DOI リンクのアクセス件数 (n=3,096,959)

時期/条件	実アクセス	ロボット等
2014 年 4~6 月	32,925	3,224
2014 年 7~9 月	34,699	8,795
2014 年 10~12 月	62,109	25,750
2015 年 1~3 月	76,975	135,040
2015 年 4~6 月	571,801	680,389
2015 年 7~9 月	608,812	856,440
合計	**1,387,321**	**1,709,638**

実アクセス 1,387,321 件について，アクセスされているコンテンツおよび参照元の分析を行った。

アクセスされているコンテンツについては，Prefix 単位，DOI 単位での集計を行った。Prefix 単位の集計から，どの Registrant のコンテンツにアクセスが多いのか，アクセスが多く行われているコンテンツおよび提供元プラットフォームについて分析を行った。

参照元については，リファラをもとに，URL 単位，完全修飾ドメイン名単位での集計，分析を行った。

3 分析結果と考察

3.1 アクセス数の多いコンテンツと参照元

アクセスが多いコンテンツについて，Prefix 単位での集計結果の上位 15 件を表 2 に示す。上位の Registrant は，いずれも医学分野関連の組織である。JaLC DOI の Prefix ごとの登録件数と比較すると，1 位の日本理学療法士協会は登録件数での 3 位，2 位の医学書院は登録件数での 1 位であることから，JaLC DOI の登録件数が多い Prefix に対するアクセスが多く行われている傾向が見られる。ただし，DOI 登録件数が 2 位の Registrant である国立国会図書館 (Prefix:10.11501) のコンテンツへのアクセス数は 4,670 件 (65 位) であり，必ずしも DOI 登録件数が多い Registrant のコンテンツへのアクセスが多いとは限らない。

表 2: アクセス数の多いコンテンツ: Prefix 単位 (上位 5 件，n=1,387,321)

順位	Prefix	Registrant	件数
1	10.14900	日本理学療法士協会	143,164
2	10.11477	医学書院	93,261
3	10.11280	日本消化器内視鏡学会	29,664
4	10.11405	日本消化器病学会	26,671
5	10.11236	日本公衆衛生学会	26,595

DOI 単位での上位 15 件を表 3 に示す。これらの項目のうち，11 位は脳科学辞典，14 位はライフサイエンス 領域融合レビューのコンテンツであり，残る 13 項目は J-STAGE 上のコンテンツである。アクセス数の多い項目は医学分野のコンテンツであり，診療ガイドライン (1，2，5，8 位) や実践ガイド (10 位)，症例報告研究 (4，9，12，15 位) が該当する。これらのうち，「日本消化器内視鏡学会 (Prefix:10.11280)」のコンテンツ (1，2，5，8，10 位) はすべて表 4 の 4 番目の URL 上で参照されている。それ以外は日本語

表 3: アクセスの多いコンテンツ: DOI 単位 (上位 15 件，n=1,387,321)

順位	DOI	タイトル	件数
1	10.11280/gee.54.2075	抗血栓薬服用者に対する消化器内視鏡診療ガイドライン	4,392
2	10.11280/gee.55.3822	内視鏡診療における鎮静に関するガイドライン	3,559
3	10.11309/jssst.31.4_9	猫にはわかる量子プログラミング	2,862
4	10.11340/skinresearch1959.13.228	尋常性乾癬を多発した 1 家系	2,256
5	10.11280/gee.56.1598	大腸 ESD/EMR ガイドライン	2,174
6	10.11236/jph.61.3_130	某ファミリーレストラングループにおける客席禁煙化前後の営業収入の相対変化　未改装店，分煙店の相対変化との比較	1,760
7	10.11467/isss2003.7.1_11	大学における Web メールとターミナルサービスの研究	1,585
8	10.11280/gee.56.310	胃癌に対する ESD/EMR ガイドライン	1,583
9	10.11256/jjdi.14.134	健康食品・サプリメントによる健康被害の現状と患者背景の特徴	1,415
10	10.11280/gee.56.89	消化器内視鏡の感染制御に関するマルチソサエティ実践ガイド	1,357
11	10.14931/bsd.1408	ストレス	1,220
12	10.11405/nisshoshi1964.97.575	プロトンポンプ阻害剤により視力障害をきたした 2 症例	1,155
13	10.11353/sesj1988.13.61	ミドリムシに対する強磁場の影響	1,067
14	10.7875/leading.author.2.e008	植物における免疫誘導と病原微生物の感染戦略	943
15	10.11213/tonyobyo.52.255	みかん缶詰・アイスクリームの大量摂取を契機に清涼飲料水ケトーシスと同様の病態を来たした 1 例	911

表 4: 参照元 URL (上位 15 件，n=1,387,321)

順位	参照元	件数	概要
1	(リファラなし)	380,838	—
2	http://search.jamas.or.jp/index.php	117,863	検索
3	https://www.google.co.jp/	40,764	検索
4	http://www.jges.net/index.php/member_submenu/archives/122	10,507	学協会
5	http://www.google.co.jp/	6,973	検索
6	http://dx.doi.org/	3,101	IDF
7	https://www.google.co.jp	2,175	検索
8	http://demo.jamas.or.jp/index.php	2,026	検索
9	http://personalsearch.jamas.or.jp/index.php	1,892	検索
10	http://jipsti.jst.go.jp/johokanri/	1,428	学協会
11	https://www.google.com/	1,297	検索
12	http://ja.wikipedia.org/wiki/乾癬	1,279	Wikipedia
13	http://t.co/0U615rEDzE	1,068	Twitter
14	http://ja.wikipedia.org/wiki/ペットボトル症候群	949	Wikipedia
15	http://www.ls-japan.org/modules/documents/index.php?content_id=39	890	学協会

版 Wikipedia が参照元であり，4 位は「乾癬」(表 4 の 12 番目)，15 位は「ペットボトル症候群」(表 4 の 14 番目)，9 位は「サプリメント」,「クロレラ」,「ウコン」，12 位は「プロトンポンプ阻害薬」，13 位は「ミドリムシ」の項目から参照されている。その他，3 位の参照元は表 4 の 13 番目であり，Twitter でのツイートをクリックしてアクセスが行われている。

参照元の完全修飾ドメイン名単位での上位 15 件を表 5 に示す。アクセス数が多い参照元として，CiNii(1 位)，医中誌 Web(3 位)，NCBI(7 位)，国立国会図書館サーチ (12 位) のような分野特化型の検索サービスやデータベース，Google(4 位) や Yahoo! JAPAN(5 位) のような検索エンジン，日本語版 Wikipedia(6, 10 位)，大学ウェブサイト (11 位)，researchmap(14 位)，Twitter(15 位) があることが分かる。この結果は完全修飾ドメイン名での集計であるため，同一サービスへのアクセスが複数に分かれている場合がある。たとえば，日本語版 Wikipedia はデスクトップ版とモバイル版で件数が分かれている。なお，Wikipedia からのアクセス数は全体で 29,860 件であり，言語版ごとの内訳は，日本語版が 29,795 件，英語版が 59 件，フランス語版が 3 件，韓国語版が 2 件，ポーランド語版が 1 件である。

表 5: 参照元の完全修飾ドメイン名 (上位 15 件, n=1,387,321)

順位	参照元	件数	概要
1	ci.nii.ac.jp	529,381	検索
2	(リファラなし)	380,838	—
3	search.jamas.or.jp	170,711	検索
4	www.google.co.jp	82,209	検索
5	search.yahoo.co.jp	74,860	検索
6	ja.wikipedia.org	20,971	Wikipedia
7	www.ncbi.nlm.nih.gov	11,597	検索
8	www.jges.net	10,974	学協会
9	dx.doi.org	10,192	IDF
10	ja.m.wikipedia.org	8,823	Wikipedia
11	ir.lib.shizuoka.ac.jp	7,988	大学
12	iss.ndl.go.jp	5,737	検索
13	jlc.jst.go.jp	5,022	学協会
14	researchmap.jp	4,366	researchmap
15	t.co	3,444	Twitter

4 おわりに

本研究では，2014 年 4 月から 2015 年 9 月までの JaLC DOI リンクのアクセスログ分析を行った。

分析の結果から，(1) どのようなコンテンツが参照されているかについては，J-STAGE 上のコンテンツ，特に医学分野コンテンツのアクセスが多いことが明らかになった。(2) どのような場所から参照されているかについては，URL 単位では，学協会ウェブサイト，日本語版 Wikipedia の項目，Twitter でのツイートなどであり，完全修飾ドメイン名単位では，CiNii や医中誌 Web，NCBI，国立国会図書館サーチのような分野特化型の検索サービスやデータベース，Google や Yahoo! JAPAN のような検索エンジンに加え，日本語版 Wikipedia，大学ウェブサイト，researchmap，Twitter などから参照されていることが明らかになった。

今後の課題として，時期ごとのアクセス数の多いコンテンツ，利用者属性とアクセス先コンテンツの関係，リファラに含まれている検索クエリなど，JaLC DOI リンクの参照状況について詳細な分析を行なう。

参考文献

[1] The International DOI Foundation. “Digital Object Identifier System”. Digital Object Identifier System. http://www.doi.org/,(参照 2015-10-23).

[2] CrossRef. “crossref.org”. crossref.org. http://www.crossref.org/, (参照 2015-10-23).

[3] Japan Link Center. “ジャパンリンクセンター (JaLC)”. ジャパンリンクセンター (JaLC). http://japanlinkcenter.org/, (参照 2015-10-23).

[4] Bilder, Geoffrey. “Many Metrics. Such Data. Wow.”. CrossTech. 2014-02-24. http://crosstech.crossref.org/2014/02/many-metrics-such-data-wow.html,(参照 2015-10-23).

[5] 吉川次郎, 高久雅生, 逸村裕. “日本語版 Wikipedia における DOI リンクの予備的分析”. 第 23 回 (2015 年度) 情報知識学会年次大会. 東京, 2015-05-23/24. 情報知識学会誌. 2015, Vol.25, No.2. p.160-165. doi:10.2964/jsik_2015_011, (参照 2015-07-13).

[6] 吉川次郎, 佐藤翔, 高久雅生, 逸村裕. “日本語版および英語版 Wikipedia における DOI リンクの重複分析”. 第 14 回情報メディア学会年次大会. 京都, 2015-06-27. 第 14 回情報メディア学会研究大会発表資料. 2015, p.27-30. http://hdl.handle.net/2241/00125076, (参照 2015-07-15).