

WikipediaにおけるDOIリンクの経年変化の予備的分析

著者	吉川 次郎, 高久 雅生
著者別名	KIKKAWA Jiro, TAKAKU Masao
内容記述	第22回情報知識学フォーラム「オープンサイエンスの障壁への挑戦」 日時：2017年12月2日（土）13:00-18:00 会場：国立情報学研究所 12階会議室
雑誌名	第22回情報知識学フォーラム予稿
ページ	329-336
発行年	2017
URL	http://hdl.handle.net/2241/00149285

Wikipedia における DOI リンクの経年変化の予備的分析 A Time Series Analysis of DOI Links on Wikipedias

吉川 次郎^{1*} 高久 雅生²
Jiro KIKKAWA^{1*}, Masao TAKAKU²

¹ 筑波大学大学院 図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of Tsukuba

〒305-8550 茨城県つくば市春日 1-2

E-mail: jiro@slis.tsukuba.ac.jp

² 筑波大学 図書館情報メディア系

Faculty of Library, Information and Media Science, University of Tsukuba

〒305-8550 茨城県つくば市春日 1-2

E-mail: masao@slis.tsukuba.ac.jp

*連絡先著者 Corresponding Author

本研究では、2015 年、2016 年、2017 年 3 月時点の日本語版、英語版、中国語版 Wikipedia における DOI リンクを対象に、経年変化に着目した分析を行なった。結果、(1) 各言語版における DOI リンクの件数は経年的に増加していること、(2) 各言語版において独自に参照されている DOI リンクは経年的に増加していること、(3) 日本の学術情報の参照記述は、他の言語版に比べて、日本語版において多いこと、(4) 日本の学術情報の参照記述は、各言語版において独自に参照されており、経年的に増加していることが明らかになった。

The authors extracted Digital Object Identifier (DOI) links on Japanese, English, and Chinese Wikipedias as of March 2015, 2016, and 2017 for a time series analysis. As a result, it is revealed that (1) the number of DOI links on each language Wikipedia is increasing year by year, (2) the number of DOI links uniquely referenced on each language Wikipedia is increasing year by year, (3) scholarly information of Japan tends to be referenced more on Japanese Wikipedia than other languages, and (4) there are unique references to scholarly information of Japan on each language Wikipedia and it is increasing year by year.

キーワード: Wikipedia, Digital Object Identifier (DOI), 学術情報流通

Keyword: Wikipedia, Digital Object Identifier (DOI), Scholarly Communication

1 はじめに

学術情報流通の急速かつ大規模な電子化により、今日、ウェブを通じた学術情報の利用が盛んに行われている。加えて、近年、「オープンサイエンス」として、最終的な成果物である学術論文だけでなく、研究データの利活用を志向する機運が高まっている。

ウェブを通じた学術情報基盤が整備され、普及することで、研究者や専門家に限らず、多様な人々やコミュニティによる学術情報の利活用が生じることが考えられる。実際に、世界最大規模の DOI (Digital

Object Identifier, デジタルオブジェクト識別子) の登録機関である Crossref の報告 [1] によると、2015 年時点でのアクセスログ分析の結果から、Crossref DOI の参照元について、Web of Science 等の学術文献データベースに次いで、誰でも編集できるオンライン百科事典である「Wikipedia」が 5 番目にアクセスの多い参照元である。

筆者らは、日本語版 Wikipedia, 英語版 Wikipedia, 中国語版 Wikipedia (以下、それぞれ、日本語版、英語版、中国語版とする) における DOI リンクの分析を行なってきた [2]。分析の結果、2015 年 3 月時点

の日本語版において、約 29,000 件の DOI リンクの参照記述が存在し、その大部分が Crossref DOI であることが明らかになった。さらに、二言語版間での重複状況の分析から、英語版ページの翻訳を通じて日本語版や中国語版にコピーされた DOI リンクの参照記述が多いことが明らかになった。以上から、学術情報の参照が行われているウェブ上のコミュニティのひとつに Wikipedia があること、日本語版や中国語版における DOI リンクの参照記述は、英語版との重複率が高いことが分かる。

しかし、DOI リンクをはじめとした、Wikipedia における学術情報の参照記述に関して、経年変化、すなわち、月日の経過につれてどのように件数が変化しているのかに着目した分析事例は少ない。

日本国内では、2012 年に発足した Japan Link Center (ジャパンリンクセンター、以下、JaLC とする) を通じて、2017 年 9 月現在、約 470 万件の日本の学術情報に DOI が登録されている [3]。このため、DOI リンクを用いて参照可能な日本の学術情報は年々増加しており、日本語版をはじめとする各言語版において、それらの参照記述が増加している可能性があるものの、分析事例は見当たらない。

もし、日本語版において、日本の学術情報に対する参照記述の大幅な増加や、参照コンテンツの多様化が見られれば、Wikipedia において日本の学術情報を参照する動きが活発化していると言える。また、他の言語版において同様の傾向が見られれば、DOI リンクを通じて、日本語版以外の言語版において日本の学術情報が参照されていることを示す結果であるといえる。

以上から、本研究では、2015 年 3 月、2016 年 3 月、2017 年 3 月の 3 つの時点における日本語版、英語版、中国語版 Wikipedia 上で参照されている DOI リンクを対象に、経年変化に着目した分析を通じて、以下のリサーチクエスチョンに答えることを目指す：

- RQ1 日本語版における DOI リンクの参照記述は、経年的に増加している
- RQ2 各言語版において独自に参照されている DOI リンクは、経年的に増加している
- RQ3 日本語版は、他の言語版に比べて、日本の学術情報の参照記述が多い
- RQ4 日本語版以外の言語版において、日本の学術情報の参照記述は経年的に増加している

2 関連研究

Wikipedia 上の学術情報を対象とした関連研究について述べる。

Nielsen [4] は、2007 年 4 月時点の英語版を対象に、参照記述が多い学術雑誌と、当該雑誌の Journal Citation Reports における Impact Factor の値との関係を分析した。その結果、Nature, Science などの学術雑誌、天文学分野の学術雑誌の参照記述が多いこと、必ずしも Impact Factor の値が高い学術雑誌の参照記述が多いわけではないことを指摘している。

Lin ら [5] は、2014 年 3 月時点においてページ数の多い 25 の言語版を対象に、DOI 名を用いて、PLOS のコンテンツの参照状況の分析を行なった。結果、調査時点での PLOS の全コンテンツのうち、4% が Wikipedia 上で参照されていること、さらに、それらのうち 47% は英語版以外の言語版で参照されていることを明らかにしている。

Halfaker [6] は、英語版およびオランダ語版における DOI, PubMed, ISBN, arXiv の参照記述を分析し、両言語版ともに、参照記述が最も多いのは ISBN で、2 番目が DOI であることを明らかにしている。また、1ヶ月ごとの参照記述の追加件数のデータを公開している。

その他、前述のとおり、筆者らによる、2015 年 3 月時点の日本語版、英語版、中国語版における DOI リンクを対象とした分析事例がある [2]。

3 DOI とは

DOI とは、コンテンツの電子データに付与される国際的な識別子であり、解決可能、持続的、相互運用可能なリンクを提供するための仕組みである。以下、識別子としての DOI を「DOI 名」と呼ぶ。

DOI 名は、「10.2964/jsik.21.06」のような文字列であり、「10.」で始まる Prefix、「/」(スラッシュ)、Suffix、から構成される。DOI 名を「https://doi.org/」(または「https://dx.doi.org/」)の末尾に加えると URI として機能し、当該コンテンツの URI にリダイレクトされる。本研究では、この URI を通じたハイパーリンクを「DOI リンク」と呼ぶ。

DOI 名の登録は、DOI 登録機関 (Registration Agency, 以下、RA とする) によって行なわれる。以下、各 RA によって登録された DOI 名を、「Crossref

DOI」,「JaLC DOI」のように「RA名 + DOI」の形式で呼ぶ。2017年時点で、RAは10機関 [7] あり、このうち、本研究に特に関係するものは、Crossref、DataCite、ISTIC、JaLCの4機関である。

4 対象・方法

4.1 分析対象

本研究では、Wikipediaのダンプデータを使用する。分析対象は、2015年3月13日時点の日本語版、同年3月4日時点の英語版、同年3月1日時点の中国語版、2016年3月5日時点の日本語版、英語版、中国語版、2017年3月1日時点の日本語版、英語版、中国語版において、百科事典のページを意味する、名前空間の値が「0」（以下、標準名前空間とする）のページ上で参照されているDOIリンクと、当該の参照記述を含むページである。

英語版を分析対象とする理由は、最大規模の言語版であることに加え、筆者らの過去の分析 [2] から、日本語版には、英語版のページの翻訳を通じたDOIリンクの参照記述が多数存在することが明らかになったことから、英語版の翻訳を通じた参照記述の件数に関する経年変化を分析するためである。中国語版を対象とする理由は、日本語版と同じくアジア圏の言語版であり、ページ数が日本語版と同等程度であることに加えて、中国語版も日本語版と同様に、英語版の翻訳を通じた参照記述が多く存在するため、中国語版との比較を通じて、日本語版の特徴を明らかにするためである。

分析対象に関する基礎データを表1、表2に示す。それぞれの表における「増加数」は、前年との差を指す。「増加率」は、増加数が前年の件数に占める割合（「増加数」/「前年の件数」*100）である。

4.2 分析方法

各時点、各言語版におけるDOIリンクと、DOIリンクを含むページを対象に、(1) RAごとの集計、(2) 二言語版間での重複状況の分析、(3) 各言語版で独自に参照されているDOIリンクの分析を行なった。

(1) RAごとの集計では、国際DOI財団が提供するAPIであるdoiRA [8]を用いた。doiRAは、任意のDOI名を指定すると、RAの情報を返すAPIである。同APIにおいて、DOI名が無効な場

合は「Invalid DOI」、DOI名が存在しない場合は「DOI does not exist」などの値が返されることがあるが、集計結果では、すべて「Error」とした。

JaLCは、自らがRAとしてDOI名を登録する、すなわち、JaLC DOIの登録以外に、CrossrefやDataCiteのメンバーとして、JaLCを経由してCrossref DOIやDataCite DOIを登録する場合がある。前者を「JaLC経由のCrossref DOI」、後者を「JaLC経由のDataCite DOI」と呼ぶこととする。このため、日本の学術情報の一部は、JaLC経由のCrossref DOIまたはJaLC経由のDataCite DOIであるが、RAごとの集計では、これらの日本の学術情報を特定することができない。そこで、平成29年8月時点のJaLCメタデータ [9]を用いて、JaLC経由のCrossref DOIおよびJaLC経由のDataCite DOIを特定し、集計を行なった。

JaLCメタデータから、JaLCが登録に関与した全てのDOI名を抽出し、RAの情報を取得した結果、JaLC DOIは2,376,724件、JaLC経由のCrossref DOIは2,275,557件、JaLC経由のDataCite DOIは32件、Errorは10,633件であった。Errorは無効なDOI名であると考えられるため、除外した。

(2) 二言語版間での重複状況の分析では、各時点、各言語版のDOIリンクの差集合と積集合を取得し、分析を行なった。図1に示すように、差集合は一方の言語版のみに記述されているDOIリンクであり、積集合は両言語版に共通のDOIリンクである。

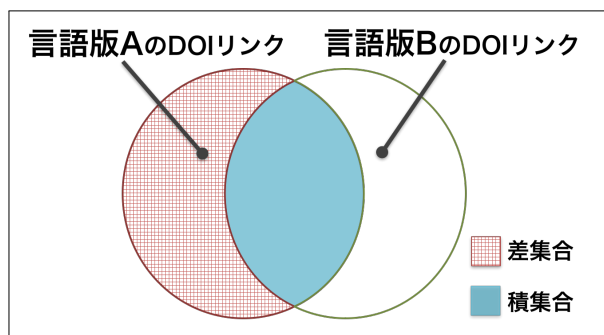


図1: 二言語版間における重複状況の分析の概要

(3) 各言語版において独自に参照されているDOIリンクの分析では、(2)で取得した差集合、すなわち、他言語版の翻訳を通じてコピーされたものではないDOIリンクの参照記述を各言語版で独自に参照されているDOIリンクとみなし、RAごとの集計と、日本の学術情報の分析を行なった。

表 1: 各言語版, 各時点における DOI リンクの件数 (標準名前空間のみ)

項目 言語版/時点	延べ DOI リンク数					異なり DOI リンク数				
	2015 年	2016 年	増加数 増加率	2017 年	増加数 増加率	2015 年	2016 年	増加数 増加率	2017 年	増加数 増加率
日本語版	28,787	35,267	6,480 22.5%	42,657	7,390 21.0%	25,438	30,902	5,464 21.5%	37,147	6,245 20.2%
英語版	1,474,251	1,737,956	263,705 17.9%	1,997,100	259,144 14.9%	518,162	610,905	92,743 17.9%	700,020	89,115 14.6%
中国語版	36,675	79,704	43,029 117.3%	96,438	16,734 21.0%	28,177	33,380	5,203 18.5%	39,789	6,409 19.2%

表 2: 各言語版, 各時点における全ページ数, DOI リンクの参照記述を含むページ数 (標準名前空間のみ)

項目 言語版/時点	全ページ数					DOI リンクの参照記述を含むページ数				
	2015 年	2016 年	増加数 増加率	2017 年	増加数 増加率	2015 年	2016 年	増加数 増加率	2017 年	増加数 増加率
日本語版	1,531,187	1,614,690	83,503 5.5%	1,690,706	76,016 4.7%	9,751	11,542	1,791 18.4%	13,286	1,744 15.1%
英語版	11,570,729	12,305,050	734,321 6.3%	13,056,443	751,393 6.1%	166,495	192,245	25,750 15.5%	218,893	26,648 13.9%
中国語版	1,446,256	1,555,463	109,207 7.6%	1,656,595	101,132 6.5%	9,677	11,200	1,523 15.7%	13,866	2,666 23.8%

5 分析結果

5.1 各言語版における DOI リンクの参照記述の経年変化

表 1 から, 各言語版における DOI リンクの件数は, 経年的に単調増加していることが分かる。延べ DOI リンク数と異なり DOI リンク数を比較すると, 英語版の件数は 3 分の 1 程度に減少しており, 参照記述の重複が多いことが分かる。また, 中国語版における延べ DOI リンク数についても, 2015 年から 2016 年までの 1 年間における増加率は 117.3% であるが, 異なり DOI リンク数での増加率は 18.5% であるため, 参照記述の重複が多いと考えられる。他方で, 日本語版は延べ DOI リンク数と異なり DOI リンク数における増加率の差が他の言語版に比べて小さいため, 参照記述の重複は少ないと考えられる。また, 異なり DOI リンク数で見ると, 日本語版と中国語版における DOI リンクの参照記述は, 同程度の規模である。

これらの結果を踏まえて, 以降の分析では, 異なり単位での集計のみを行なう。異なり DOI リンク数における増加率の結果として, 日本語版は約 20%,

英語版は約 15~18%, 中国語版は約 19% の割合で経年的に DOI リンクの参照記述が増加している。

表 2 から, 各言語版におけるページの件数は, 年間で約 5~8% ほどの割合で単調増加していることが分かる。DOI リンクの参照記述を含むページの件数も同様に単調増加しているが, 増加率は約 15~24% であるため, いずれの言語版においても, DOI リンクの参照記述を含むページ数の増加率のほうが, 言語版自体のページ数の増加率よりも高い。

5.2 RA ごとの集計結果

各時点での各言語版における DOI リンクを RA ごとに集計した結果を表 3 に示す。表の着色箇所は, 件数が 0 であるものを示す。表における「(JaLC 経由)」は, JaLC 経由の Crossref DOI, JaLC 経由の DataCite DOI を意味する。すべての時点, すべての言語版において, JaLC 経由の DataCite DOI の参照記述は存在しない。

集計結果から, 全言語版において参照されている DOI リンクの 95% 以上は Crossref DOI である。Crossref DOI を除くと, 年間に数百件程度のペース

表 3: 各言語版, 各時点における, 異なり DOI 名単位での RA ごとの集計結果

言語版 RA/時点	日本語版			英語版			中国語版		
	2015 年	2016 年	2017 年	2015 年	2016 年	2017 年	2015 年	2016 年	2017 年
Airiti	0	0	0	1	4	11	0	1	3
CNKI	0	0	0	0	4	8	0	4	9
Crossref	24,672	29,726	35,607	514,313	605,819	693,913	27,824	32,942	39,273
(JaLC 経由)	1,607	2,421	3,103	2,200	2,541	2,885	73	88	122
DataCite	10	12	21	230	516	957	5	13	26
(JaLC 経由)	0	0	0	0	0	0	0	0	0
ISTIC	0	0	2	38	73	118	42	60	87
JaLC	475	823	1,116	5	20	33	0	1	4
KISTI	0	0	0	1	3	5	0	0	0
mEDRA	5	9	16	274	389	548	9	17	26
OP	2	2	4	67	76	118	3	4	5
Public	3	3	4	84	87	87	7	7	7
Unknown	0	0	0	0	0	1	0	0	0
Error	271	327	377	3,149	3,914	4,221	287	331	349
合計	25,438	30,902	37,147	518,162	610,905	700,020	28,177	33,380	39,789

で増加しているものとして, 英語版における DataCite DOI, mEDRA DOI, 日本語版における JaLC DOI が挙げられる。中国語版では, 中国の RA である, ISTIC DOI の参照記述が比較的多い。

各言語版における日本の学術情報の参照記述については, まず, JaLC DOI の参照記述が特に日本語版に多いことから, 日本語版において日本の学術情報の参照が多いと言える。他方で, 英語版および中国語版は, JaLC DOI の参照記述は数十件程度と僅かであるが, 2017 年時点における JaLC 経由の Crossref DOI の参照記述について, 英語版が 3,103 件, 中国語版が 122 件であることから, これらの言語版においても日本の学術情報の参照記述が存在し, 経年的に件数が増加していることが分かる。

5.3 二言語版間での重複状況

2015 年時点, 2016 年時点, 2017 年時点での各言語版における異なり DOI リンク単位での重複状況を表 4 に示す。表における対象は, 差集合および積集合の取得対象となる 2 つの言語版の名称を示す。たとえば, 「日本語版-英語版」は「日本語版と英語版」を意味する。このとき, 差集合は, 日本語版で参照されており, 尚且つ, 英語版で参照されていない DOI リンクである。積集合は, 日本語版と英語版

に共通して参照されている DOI リンクである。割合は, それらが日本語版全体に占める割合を示す。

分析結果から, 2015 年時点で, 日本語版における DOI リンクの約 80% が英語版と重複しており, 中国語版における DOI リンクの約 93% が英語版と重複している。この時点において, 日本語版および中国語版は, 英語版との重複率が高いことが分かる。一方で, 2016 年, 2017 年時点と比較すると, 日本語版における, 英語版との重複率は, 徐々に低下していることが分かる。中国語版における, 英語版との重複率は, 経年的に僅かに低下しているものの, 概ね横ばいである。

日本語版と中国語版は, DOI リンクの参照記述の件数が同等程度であるが, 「日本語版-英語版」の差集合は, 2015 年から 2017 年までの間に 4,234 件の増加が見られる。一方で, 「中国語版-英語版」の差集合は 985 件の増加であるため, 特に日本語版において, 英語版の翻訳を通じたものではない, 独自の DOI リンクの参照記述が増えていると考えられる。

5.4 各言語版において独自に参照されている DOI リンクの分析結果

各言語版において独自に参照されている DOI リンクの分析として, 二言語版間での重複状況の分

表 4: 各時点, 各言語版における異なり DOI リンク単位での重複状況

2015 年

条件/対象	日本語版-英語版	英語版-日本語版	中国語版-英語版	英語版-中国語版	日本語版-中国語版	中国語版-日本語版
差集合	5,218	497,942	2,013	491,998	20,763	23,502
割合 (%)	20.5	96.1	7.1	95.0	81.6	83.4
積集合	20,220	20,220	26,164	26,164	4,675	4,675
割合 (%)	79.5	3.9	92.9	5.0	18.4	16.6
合計	25,438	518,162	28,177	518,162	25,438	28,177
割合 (%)	100.0	100.0	100.0	100.0	100.0	100.0

2016 年

条件/対象	日本語版-英語版	英語版-日本語版	中国語版-英語版	英語版-中国語版	日本語版-中国語版	中国語版-日本語版
差集合	7,327	587,330	2,407	579,932	25,436	27,914
割合 (%)	23.7	96.1	7.2	94.9	82.3	83.6
積集合	23,575	23,575	30,973	30,973	5,466	5,466
割合 (%)	76.3	3.9	92.8	5.1	17.7	16.4
合計	30,902	610,905	33,380	610,905	30,902	33,380
割合 (%)	100.0	100.0	100.0	100.0	100.0	100.0

2017 年

条件/対象	日本語版-英語版	英語版-日本語版	中国語版-英語版	英語版-中国語版	日本語版-中国語版	中国語版-日本語版
差集合	9,452	672,325	2,998	663,229	30,668	33,310
割合 (%)	25.4	96.0	7.5	94.7	82.6	83.7
積集合	27,695	27,695	36,791	36,791	6,479	6,479
割合 (%)	74.6	4.0	92.5	5.3	17.4	16.3
合計	37,147	700,020	39,789	700,020	37,147	39,789
割合 (%)	100.0	100.0	100.0	100.0	100.0	100.0

析で取得した差集合を用いて、RA ごとの集計を再度行なった結果を表 5 に示す。集計結果において、Crossref, JaLC, ISTIC 以外の RA は「その他」として合算した。

表 5 の結果から、各言語版において独自に参照されている DOI リンクの大部分は Crossref DOI であることが分かる。また、日本語版で独自に参照されている DOI リンクには JaLC DOI が多いことから、日本語版において、日本の学術情報に対する独自の参照記述が多いことが分かる。

JaLC DOI および ISTIC DOI に着目すると、表 3, 表 5 の結果から、各時点の各言語版における JaLC DOI および ISTIC DOI は、他の言語版とほとんど重複しておらず、各言語版で独自に参照されていることが分かる。たとえば、2017 年時点の日本語版における JaLC DOI は 1,116 件であり、英語版との重

複は 3 件、中国語版との重複は 1 件である。同時点の英語版における JaLC DOI は 33 件で、日本語版との重複は 3 件、中国語版との重複は 1 件である。同時点の中国語版における JaLC DOI は 4 件で、日本語版との重複は 1 件、英語版との重複は 1 件である。同様に、各時点、各言語版における ISTIC DOI は他言語版とほとんど重複しておらず、各言語版で独自に参照されていることが分かる。

各言語版において独自に参照されている DOI リンクについて、日本の学術情報を特定したうえで集計を行なった結果を表 6 に示す。

表 6 の結果から、日本語版以外の言語版においても、JaLC 経由の Crossref DOI が独自に参照されており、それらの件数が経年的に増加していることが明らかになった。なお、特に件数が多いのは日本語版と英語版で、中国語版は件数が少ない。このこ

表 5: 各時点, 各言語版での差集合における RA ごとの集計結果

2015 年

RA/対象	日本語版-英語版	英語版-日本語版	中国語版-英語版	英語版-中国語版	日本語版-中国語版	中国語版-日本語版
Crossref	4,536	494,176	1,766	488,254	20,015	23,167
ISTIC	0	38	41	37	0	42
JaLC	474	4	0	5	475	0
その他	5	642	6	639	16	20
Error	203	3,082	200	3,063	257	273
合計	5,218	497,942	2,013	491,998	20,763	23,502

2016 年

RA/対象	日本語版-英語版	英語版-日本語版	中国語版-英語版	英語版-中国語版	日本語版-中国語版	中国語版-日本語版
Crossref	6,271	582,364	2,121	574,997	24,284	27,501
ISTIC	0	73	56	69	0	60
JaLC	820	17	0	19	822	0
その他	6	1,059	11	1,044	21	41
Error	230	3,817	219	3,803	309	312
合計	7,327	587,330	2,407	579,932	25,436	27,914

2017 年

RA/対象	日本語版-英語版	英語版-日本語版	中国語版-英語版	英語版-中国語版	日本語版-中国語版	中国語版-日本語版
Crossref	8,042	666,348	2,683	657,322	29,155	32,822
ISTIC	1	117	80	111	2	87
JaLC	1,113	30	3	32	1,115	3
その他	16	1,706	21	1,680	38	69
Error	280	4,124	211	4,084	358	329
合計	9,452	672,325	2,998	663,229	30,668	33,310

とから、日本語版以外の言語版においても、DOI リンクを通じて、日本の学術情報の参照が独自に行われていると言える。

RQ3 日本語版は、他の言語版に比べて、日本の学術情報の参照記述が多い

RQ4 日本語版以外の言語版において、日本の学術情報の参照記述は経年的に増加している

6 考察と今後の課題

本研究では、2015 年 3 月、2016 年 3 月、2017 年 3 月の 3 つの時点における日本語版、英語版、中国語版 Wikipedia 上で参照されている DOI リンクを対象に、経年変化に着目した分析を通じて、以下のリサーチクエストに答えるための分析を行った:

RQ1 日本語版における DOI リンクの参照記述は、経年的に増加している

RQ2 各言語版において独自に参照されている DOI リンクは、経年的に増加している

RQ1 は、表 1、表 2 の結果から、日本語版における DOI リンクの参照記述の件数、DOI リンクの参照記述を含むページの件数ともに、経年的に単調増加していることが明らかになった。また、日本語版以外の言語版においても同様の傾向が見られることが明らかになった。具体的には、DOI リンクの参照記述は、1 年あたり、日本語版では約 20%、英語版では約 15~18%、中国語版では約 19%の割合で増加していることが明らかになった。

RQ2 は、表 4 の結果から、2015 年時点での日本語版および中国語版における DOI リンクの参照記述

表 6: 各時点, 各言語版での差集合における日本の
学術情報の集計結果

対象/項目	Crossref	JaLC	合計
	DOI	DOI	
2015 年 日本語版-英語版	1,448	473	1,921
2016 年 日本語版-英語版	2,231	819	3,050
2017 年 日本語版-英語版	2,878	1,111	3,989
2015 年 英語版-日本語版	2,041	4	2,045
2016 年 英語版-日本語版	2,351	17	2,368
2017 年 英語版-日本語版	2,660	30	2,690
2015 年 中国語版-英語版	12	0	12
2016 年 中国語版-英語版	18	0	18
2017 年 中国語版-英語版	33	3	36
2015 年 英語版-中国語版	2,139	5	2,144
2016 年 英語版-中国語版	2,471	19	2,490
2017 年 英語版-中国語版	3,796	32	3,828
2015 年 日本語版-中国語版	1,595	474	2,069
2016 年 日本語版-中国語版	2,403	821	3,224
2017 年 日本語版-中国語版	3,968	1,113	5,081
2015 年 中国語版-日本語版	61	0	61
2016 年 中国語版-日本語版	70	0	70
2017 年 中国語版-日本語版	87	3	90

は、英語版との重複率が高いことが分かった。2016年、2017年時点との比較から、日本語版における英語版との重複率は、経年的に低下していることが分かった。また、差集合の件数から、日本語版および中国語版において、英語版ページの翻訳ではない、独自に参照されている DOI リンクの件数が経年的に増加していることが明らかになった。

RQ3, RQ4 は、表 3, 表 5 から、日本語版の特徴として、他の言語版に比べて、日本の学術情報の参照記述が多いことが明らかになった。他方で、JaLC DOI 以外の日本の学術情報の参照記述として、JaLC 経由の Crossref DOI を特定したうえで集計を行なった表 6 の結果から、それぞれの言語版において、日本の学術情報が独自に参照されていること、そして、それらの件数が経年的に増加していることが明らかになった。

今後の課題として、(1) 各言語版において独自に参照されている日本の学術情報を対象とした詳細な分析を行なうとともに、なぜこのような経年的な増加が生じているのかを明らかにするために、(2) 各言語版において DOI リンクの参照記述を追加した

編集者に着目した詳細な分析を行ないたい。

参考文献

- [1] Bilder, Geoffrey. “Geoffrey Bilder: Strategic Initiatives Update”. SlideShare. 2015-11-23. <http://www.slideshare.net/CrossRef/geoffrey-bilder-crossref15#24>, (参照 2017-10-23).
- [2] Jiro, Kikkawa.; Masao, Takaku.; Fuyuki, Yoshikane. “DOI Links on Wikipedia: Analyses of English, Japanese, and Chinese Wikipedias”. Lecture Notes in Computer Science. 2016, vol.10075, p.369-380. https://doi.org/10.1007/978-3-319-49304-6_40, (参照 2017-10-23).
- [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). <https://japanlinkcenter.org/top/>, (参照 2017-10-23).
- [4] Nielsen, Finn Arup. “Scientific citations in Wikipedia”. First Monday. 2007, vol.12, no.8, p.1-5. <http://doi.org/10.5210/fm.v12i8.1997>, (参照 2017-10-23).
- [5] Lin, Jennifer.; Fenner, Martin. “An analysis of Wikipedia references across PLOS publications”. altmetrics14 workshop at WebSci. <https://doi.org/10.6084/m9.figshare.1048991>,(参照 2017-10-23).
- [6] Halfaker, Aaron. “Research:Scholarly article citations in Wikipedia”. Meta, a Wikimedia project coordination wiki. https://meta.wikimedia.org/wiki/Research:Scholarly_article_citations_in_Wikipedia, (参照 2017-10-23).
- [7] The International DOI Foundation. “DOI Registration Agencies”. Digital Object Identifier System. http://www.doi.org/registration_agencies.html, (参照 2017-10-23).
- [8] The International DOI Foundation. “DOI Handbook – 5.Applications (5.2.1 Basic Functionality)”. Digital Object Identifier System. https://www.doi.org/doi_handbook/5_Applications.html#5.2.1, (参照 2017-10-23).
- [9] Japan Link Center. “JaLC メタデータ (平成 29 年 8 月 25 日時点)”. Japan Link Center. https://japanlinkcenter.org/top/material/material_metadata.html, (参照 2017-10-23).