

## Wikipedia上のJ-STAGEコンテンツの分析：研究分野 を中心に

著者	吉川 次郎, 高久 雅生
著者別名	KIKKAWA Jiro, Takaku Masao
内容記述	第15回情報メディア学会研究大会 日時：2016年6月25日（土）9:30-18:30 会場：筑波大学筑波キャンパス春日地区（エリア ）
雑誌名	第15回情報メディア学会研究大会発表資料
ページ	39-42
発行年	2016
URL	<a href="http://hdl.handle.net/2241/00143015">http://hdl.handle.net/2241/00143015</a>

# Wikipedia 上の J-STAGE コンテンツの分析: 研究分野を中心に

Analyses of J-STAGE Articles on Wikipedia: From the View Point of Research Fields

吉川次郎<sup>1</sup>, 高久雅生<sup>2</sup>

Jiro KIKKAWA<sup>1</sup>, Masao TAKAKU<sup>2</sup>

<sup>1</sup> 筑波大学大学院図書館情報メディア研究科,

<sup>2</sup> 筑波大学図書館情報メディア系

<sup>1</sup>Graduate School of Library, Information and Media Studies, University of Tsukuba,

<sup>2</sup>Faculty of Library, Information and Media Science, University of Tsukuba

あらまし: Wikipedia における日本国内の学術情報の参照について、研究分野および言語版ごとの特徴や差異を明らかにするため、日本語版、英語版、中国語版における J-STAGE コンテンツの参照状況の分析を行った。結果、研究分野は生命科学分野の割合が共通して高く、日本語版は理学・工学分野の割合が他言語版に比べて高いこと、日本語版は和文誌や英和混在誌が多いのに対し、英語版および中国語版では英文誌が多いことが分かった。

キーワード: J-STAGE、Wikipedia、学術情報流通、研究分野、学会名鑑

## 1. はじめに

本研究の目的は、Wikipedia 上で参照されている学術情報について、J-STAGE コンテンツを対象に、研究分野および言語版ごとの特徴を明らかにすることである。

学術情報流通の電子化を背景に、今日、ウェブ上で学術情報の参照が行われるようになっている。ハイパーリンクを通じたウェブ上での学術情報の参照を用いて、代替的な評価指標 (Altmetrics) を模索する動きも存在する。世界最大の DOI (Digital Object Identifier、デジタルオブジェクト識別子) 登録機関である Crossref は、Crossref DOI の参照元の分析を通じて、2015 年時点で、Web of Science などの学術文献データベースに次いで、オンライン百科事典の Wikipedia が 5 番目にアクセスの多い参照元であることを報告している [1]。このことから、ウェブ上で学術情報の参照が行われているコミュニティのひとつに Wikipedia がある。

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 [2] は、英語版 Wikipedia (以下、「英語版」) の分析から、Nature、Science などのジャーナルが多いこと、天文学分野のジャーナルが多いこと、Impact Factor の値が高いジャーナルの参照が必ずしも多いわけではないことを指摘している。Lin ら [3] は、英語版における PLOS コンテンツの参照状況の分析を行い、ジャーナル単位では、PLOS Biology、PLOS Medicine、PLOS Genetics が多いことを指摘している。佐藤ら [4] は日本語版 Wikipedia (以下、「日本語版」) を対象に、PubMed、CiNii、機関リポジトリの URI の分析を行った。吉川ら [5] は日本語版における DOI リンクの分析から、日本国外の大手出版社の学術情報が多いこと、自然科学分野が多いことを指摘している。

しかし、Wikipedia において参照されている日本国内の学術情報について、研究分野に着目した分析事例や、Wikipedia の言語版単位での特徴の比較事例は管見の限り見当たらない。これらの分析を行うことで、日本国内の学術情報について、どのようなコンテンツが日本国内/国外のコミュニティで参照されているのかを明らかにするための足がかりとするとともに、日本語版において学術情報の参照が少ない分野の特定および出典記述の充実に向けた応用が可能であると考えられる。そこで本研究では、日本国内最大規模の電子ジャーナルプラットフォームである J-STAGE のコンテンツを対象に、研究分野を中心に、日本語版、英語版、中国語版 Wikipedia の分析を行う。具体的には、(1) 各言語版において、どのような研究分野のコンテンツが参照さ

れているか、(2) 参照コンテンツの言語版ごとの特徴について明らかにする。

## 2. 対象と方法

2016年5月時点での「学会名鑑<sup>1</sup>」および「J-STAGE ジャーナルリスト<sup>2</sup>」を用いて、日本語版、英語版、中国語版 Wikipedia において参照されている J-STAGE コンテンツの分析を行った。

### 2.1 分析対象

2016年3月5日時点の日本語版、英語版、中国語版 Wikipedia のダンプデータを用いて、各言語版において百科事典記事を意味する「標準名前空間」に記述されている外部リンクのうち、(1)「doi.org」を含む項目、(2)「jstage.jst.go.jp」を含む項目(以下、「J-STAGE URI」)を抽出した。さらに、J-STAGE の各資料のトップページなどへの参照を除外し、記事詳細または記事本文の URI(以下、「J-STAGE 記事」)を抽出した。すなわち、(1)は、DOI 登録機関が Crossref または JaLC(Japan Link Center)であり、尚且つ、J-STAGE 記事へのリダイレクトが行われるものを抽出し、(2)は、URIに「article」の文字列を含むものを抽出した。抽出結果を表1に示す。

表 1: 各言語版における J-STAGE 記事の延べ数、資料コードの異なり数

条件/言語版	日本語版	英語版	中国語版
Crossref DOI	33,878	1,723,234	78,692
うち J-STAGE 記事	2,707	4,586	127
JaLC DOI	957	39	0
うち J-STAGE 記事	925	35	0
J-STAGE URI	1,656	617	64
うち J-STAGE 記事	1,579	580	57
<b>J-STAGE 記事の延べ数</b>	<b>5,211</b>	<b>5,201</b>	<b>184</b>
<b>資料コードの異なり数</b>	<b>598</b>	<b>346</b>	<b>70</b>

本研究では学会名鑑を用いて研究分野の特定と分析を行うため、学会名鑑に登録されている項目のみを分析対象とした。学会名鑑には1,934件の学協会、2,616誌のジャーナルが登録されている。うち、767誌に J-STAGE URI が登録されている。資料コード<sup>3</sup>を用いて学会名鑑との照合を行った結果を表2に示す。たとえば、日本語版では5,211件の J-STAGE 記事が参照されているが、学会名鑑を用いて研究分野が特定可能であるものは3,897件(74.8%相当)である。

表 2: 各言語版で参照されているコンテンツのうち、学会名鑑と照合可能であるものの件数

条件/言語版	日本語版	割合 (%)	英語版	割合 (%)	中国語版	割合 (%)
資料コードの異なり数	598	100.0	346	100.0	70	100.0
うち分析対象	<b>551</b>	<b>92.1</b>	<b>202</b>	<b>58.4</b>	<b>45</b>	<b>64.3</b>
J-STAGE 記事の延べ数	5,211	100.0	5,201	100.0	184	100.0
うち分析対象	<b>3,897</b>	<b>74.8</b>	<b>3,569</b>	<b>68.6</b>	<b>99</b>	<b>53.8</b>

### 2.2 分析方法

本研究では、(1) 各言語版において参照の多いジャーナルの分析、(2) 学会名鑑を用いた研究分野ごとの分析、の2つの分析を行った。

<sup>1</sup><https://gakkai.jst.go.jp/gakkai/>

<sup>2</sup>[https://www.jstage.jst.go.jp/pub/html/002.jp\\_menu\\_.html](https://www.jstage.jst.go.jp/pub/html/002.jp_menu_.html) のジャーナル一覧を使用した

<sup>3</sup>たとえば、「情報知識学会誌」の資料コードは「jsik」である。同誌の J-STAGE URI は「<https://www.jstage.jst.go.jp/browse/jsik/>」である。なお、資料コード末尾が「西暦」または「西暦+アルファベット1文字」である場合は、それらの文字列を除去したうえで照合を行った。

(1)は、資料コード単位での集計を行った。また、ジャーナルリストを用いて、和文誌や英和混在誌などの言語種別の特定を行った。(2)は、学会名鑑における各学協会のデータから、J-STAGE URIと研究分野を取得した。学会名鑑には「人文・社会学」、「生命科学」、「理学・工学」の3分野(以下、「大分類」)がある。さらに、「人文・社会学」は10、「生命科学」は9、「理学・工学」は11の下位分類(以下、「小分類」)がある。学会名鑑は学協会単位のデータベースであり、大分類、小分類ともに、学協会自体の研究分野である。本研究では、発行元学協会の研究分野をジャーナルの研究分野とみなした。複数の学協会が発行するジャーナルは、当該学協会の研究分野の和集合を用いた。研究分野を複数もつ場合は、研究分野の数で等分して集計を行った。

### 3. 分析結果

#### 3.1 各言語版において参照の多いジャーナル

各言語版において参照の多いジャーナルを表3、表4、表5に示す。中国語版は分析対象件数が少ないため、上位3件のみを示す。

これらの結果から、言語種別について、日本語版では和雑誌または英和混在誌が上位であるため、日本語で書かれたものが多く参照されていると考えられる一方で、英語版と中国語版は英文誌が多い点に差異が見られる。また、ジャーナル名および学協会名から、日本語版では「地震」や「地学雑誌」、英語版と中国語版は「日本薬学会」のコンテンツが上位に位置している。

表 3: 日本語版において参照の多いジャーナル (上位 5 件、n=3,897)

順位	資料コード	ジャーナル名	学協会名	言語種別	件数
1	zisin1948	地震 第 2 輯	日本地震学会	和文誌	233
2	jgeography1889	地学雑誌	東京地学協会	英和混在誌	151
3	jrsj1983	日本ロボット学会誌	日本ロボット学会	和文誌	124
4	zisin1929	地震 第 1 輯	日本地震学会	和文誌	103
5	suisan	日本水産学会誌	日本水産学会	和文誌	95

表 4: 英語版において参照の多いジャーナル (上位 5 件、n=3,569)

順位	資料コード	ジャーナル名	学協会名	言語種別	件数
1	antibiotics1968	The Journal of Antibiotics	日本感染症医薬品協会	英文誌	406
2	bpb	Biological and Pharmaceutical Bulletin	日本薬学会	英文誌	379
3	bbb1961	Agricultural and Biological Chemistry	日本農芸化学会	英文誌	349
4	cpb1958	Chemical and Pharmaceutical Bulletin	日本薬学会	英文誌	214
5	cpb	Chemical and Pharmaceutical Bulletin	日本薬学会	英文誌	172

表 5: 中国語版において参照の多いジャーナル (上位 3 件、n=99)

順位	資料コード	ジャーナル名	学協会名	言語種別	件数
1	bpb	Biological and Pharmaceutical Bulletin	日本薬学会	英文誌	8
2	jphs	Journal of Pharmacological Sciences	日本薬理学会	英文誌	6
3	bpb1993	Biological and Pharmaceutical Bulletin	日本薬学会	英文誌	5

#### 3.2 各言語版において参照の多い研究分野

各言語版において参照の多い研究分野について、大分類での集計結果を表6、小分類での集計結果を表7に示す。表において、「人文・社会科学」は丸(●)、「生命科学」は星(★)、「理学・工学」は四角(■)で示す。表6から、いずれの言語版においても最も割合の高い項目は「生命科学」であり、最も割合の低い項目は「人文・社会科学」である。日本語版は、他の言語版に比

べて「理学・工学」が占める割合が高く、表7の小分類では「地球惑星科学」の割合が最も高い。この点については、表3から、日本語版では「地震」や「地学雑誌」など「理学・工学」分野かつ「地球惑星科学」分野のジャーナルの参照が多いためと考えられる。英語版は「生命科学」の割合が特に高い。表7から、英語版と中国語版では「薬学」分野の占める割合が高い。

表 6: 大分類での研究分野の集計結果 単位:%

言語版/研究分野	●人文・社会科学	★生命科学	■理学・工学	合計
日本語版	6.4	50.4	43.2	100.0
英語版	2.1	89.4	8.5	100.0
中国語版	7.1	67.5	25.4	100.0

表 7: 小分類での研究分野の集計結果 (上位5分野のみ) 単位:%

言語版 順位/条件	日本語版		英語版		中国語版	
	研究分野	割合 (%)	研究分野	割合 (%)	研究分野	割合 (%)
1	■ 地球惑星科学	20.0	★ 薬学	26.4	★ 薬学	24.2
2	★ 農学	13.8	★ 健康・生活科学	13.1	★ 基礎医学	15.7
3	★ 臨床医学	12.7	★ 臨床医学	12.9	★ 臨床医学	11.6
4	■ 機械工学	6.6	★ 基礎医学	12.0	■ 地球惑星科学	9.1
5	★ 食料科学	5.7	★ 農学	10.3	■ 総合工学	4.5

#### 4. 考察と今後の課題

本研究では、J-STAGEのコンテンツについて、研究分野を中心に、日本語版、英語版、中国語版 Wikipedia での参照状況の分析を行った。結果、(1) 各言語版において、どのような研究分野のコンテンツが参照されているかについては、生命科学分野の占める割合が共通して高いこと、日本語版は理学・工学分野の割合が他の言語版に比べて高いことが明らかになった。(2) 参照コンテンツの言語版ごとの特徴については、言語種別としては、日本語版は和文誌または英和混在誌が多いのに対し、英語版と中国語版では英文誌が多いという差異があること、中国版は J-STAGE コンテンツの参照が少ないことが明らかになった。

今後の課題を述べる。第一に、J-STAGEのコンテンツに対象を限定した場合であっても、学会名鑑に情報のない資料コードの分析は行えておらず、これらを対象とした分析の検討が必要である。第二に、複数の研究分野をもつジャーナルに対し、研究分野の重み付けを行う手法の検討が課題である。第三に、J-STAGE 以外の学術情報に対しても分析を行うことが課題である。

#### 参考文献

- [1] Bilder, Geoffrey. "Geoffrey Bilder: Strategic Initiatives Update". SlideShare. 2015-11-23. <http://www.slideshare.net/CrossRef/geoffrey-bilder-crossref15>, (参照 2016-05-31).
- [2] 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 2016-05-31).
- [3] Lin, Jennifer; Fenner, Martin; "An analysis of Wikipedia references across PLOS publications". altmetrics14 workshop at WebSci. <http://doi.org/10.6084/m9.figshare.1048991>, (参照 2016-05-31).
- [4] 佐藤翔, 吉田光男, 逸村裕. Wikipedia 日本語版からの学術論文の引用状況. 2013 年日本図書館情報学会春季研究集会. 茨城, 2013-05-25. 2013 年日本図書館情報学会春季研究集会発表論文集. 2013, p.27-30.
- [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. <http://doi.org/10.2964/jsik.2015.011>, (参照 2016-05-31).