

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



吉川次郎 (筑波大学大学院)
jiro@slis.tsukuba.ac.jp

高久雅生 (筑波大学)
masao@slis.tsukuba.ac.jp

分析対象

2015年3月, 2016年3月, 2017年3月時点の日本語版, 英語版, 中国語版

Wikipediaにおいて, 標準名前空間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

表1: 各言語版, 各時点におけるDOIリンクの件数

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

項目 言語版/時点	DOIリンクの参照記述を含むページ数					全ページ数				
	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	9,751	11,542	1,791 18.4%	13,286	1,744 15.1%	1,531,187	1,614,690	83,503 5.5%	1,690,706	76,016 4.7%
英語版	166,495	192,245	25,750 15.5%	218,893	26,648 13.9%	11,570,729	12,305,050	734,321 6.4%	13,056,443	751,393 6.1%
中国語版	9,677	11,200	1,523 15.7%	13,866	2,666 23.8%	1,446,256	1,555,463	109,207 7.6%	1,656,595	101,132 6.5%

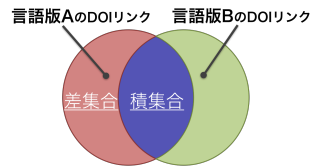
分析方法

1. RA (DOI登録機関) ごとの集計

- RAごとの集計
- DOIリンクを通じて参照されている日本の学術情報の特定・集計:
"JaLC経由の" Crossref DOIまたはDataCite DOI

2. 二言語版間での重複状況

- 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



3. 各言語版で独自に参照されているDOIリンクの分析

- 差集合を「他言語版の翻訳を通じてコピーされたものではない」、すなわち「各言語版で独自に参照されているDOIリンク」とみなしたうえで分析
- RAごとの集計, DOIリンクを通じて参照されている日本の学術情報の分析

分析結果

表3: 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 (JaLC 経由)	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 (JaLC 経由)	10	12	21	230	516	957	5	13	26
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

表6: 差集合における日本の学術情報の集計結果

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

考察

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

- 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ともに, 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

- 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
- 日本語版と中国語版において, 英語版ページの翻訳ではない, 独自参照のDOIリンクが経年的に増加している

RQ3 日本語版は, DOIリンクを通じた日本の学術情報の参照記述が, 他の言語版に比べて多い

- 日本語版の特徴として, DOIリンクを通じた日本の学術情報の参照記述が, 他の言語版に比べて多い

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

- それぞれの言語版において, DOIリンクを通じて日本の学術情報が独自に参照されており, それらの件数は経年的に増加している

表4: 異なり DOI名単位での重複状況

2015年					
条件/対象	日-英	英-日	中-英	英-中	
差集合	5,218	497,942	2,013	491,998	
割合	20.5%	96.1%	7.1%	95.0%	
積集合	20,220	20,220	26,164	26,164	
割合	79.5%	3.9%	92.9%	5.0%	
合計	25,438	518,162	28,177	518,162	
割合	100.0%	100.0%	100.0%	100.0%	

2016年					
条件/対象	日-英	英-日	中-英	英-中	
差集合	7,327	587,330	2,407	579,932	
割合	23.7%	96.1%	7.2%	94.9%	
積集合	23,575	23,575	30,973	30,973	
割合	76.3%	3.9%	92.8%	5.1%	
合計	30,902	610,905	33,380	610,905	
割合	100.0%	100.0%	100.0%	100.0%	

2017年					
条件/対象	日-英	英-日	中-英	英-中	
差集合	9,452	672,325	2,998	663,229	
割合	25.4%	96.0%	7.5%	94.7%	
積集合	27,695	27,695	36,791	36,791	
割合	74.6%	4.0%	92.5%	5.3%	
合計	37,147	700,020	39,789	700,020	
割合	100.0%	100.0%	100.0%	100.0%	

表5: 差集合における RAごとの集計結果

2015年					
条件/対象	日-英	英-日	中-英	英-中	
Crossref	4,538	494,176	1,766	488,254	
ISTIC	0	38	41	37	
JaLC	474	4	0	5	
その他	5	642	6	639	
Error	203	3,082	200	3,063	
合計	5,218	497,942	2,013	491,998	

2016年					
条件/対象	日-英	英-日	中-英	英-中	
Crossref	6,271	582,364	2,121	574,997	
ISTIC	0	73	56	69	
JaLC	820	17	0	19	
その他	0	1,059	11	1,044	
Error	230	3,817	219	3,803	
合計	7,321	587,330	2,407	579,932	

2017年					
条件/対象	日-英	英-日	中-英	英-中	
Crossref	8,042	666,348	2,683	657,322	
ISTIC	1	117	80	111	
JaLC	1,113	30	3	32	
その他	16	1,706	21	1,680	
Error	280	4,124	211	4,084	
合計	9,452	672,325	2,998	663,229	