

ICADL2016 Tsukuba, Japan December 9th, 2016



DOI Links on Wikipedia

Analyses of English, Japanese, and Chinese
Wikipedias

Jiro KIKKAWA[†] **Masao TAKAKU**[‡] **Fuyuki YOSHIKANE**[‡]

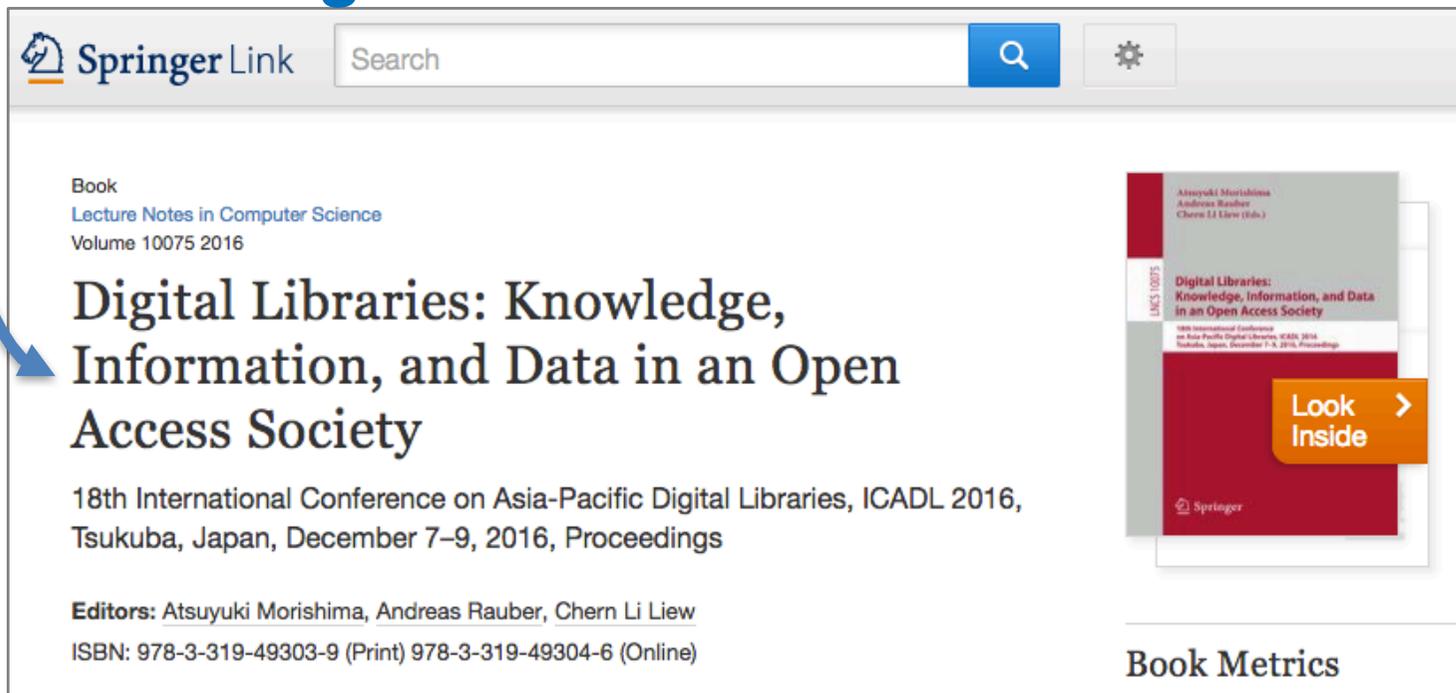
[†] Graduate School of Library, Information and Media Studies, University of Tsukuba

[‡] Faculty of Library, Information and Media Science, University of Tsukuba

{ jiro, masao, fuyuki } @slis.tsukuba.ac.jp

In this research ...

- We analyzed **Digital Object Identifier (DOI) links** among **English**, **Japanese**, and **Chinese Wikipedias** (hereafter, **enwiki**, **jawiki**, and **zhwiki**, respectively).
- An example of DOI link
 - <https://doi.org/10.1007/978-3-319-49304-6>



The screenshot shows the Springer Link interface. At the top, there is a search bar and a settings icon. The main content area displays the book details for "Digital Libraries: Knowledge, Information, and Data in an Open Access Society". The book is part of the "Lecture Notes in Computer Science" series, Volume 10075, published in 2016. The editors are Atsuyuki Morishima, Andreas Rauber, and Chern Li Liew. The ISBN for the print edition is 978-3-319-49303-9 and for the online edition is 978-3-319-49304-6. A "Look Inside" button is visible on the book cover image. The Springer logo is also present at the bottom of the book cover.

Outline

- Background
- Related Work
- About DOI
- Materials and Methods
- Results and Discussion
- Conclusion

Background

- Fast-growing digitization of scholarly communication
 - All people can easily, immediately get scholarly information through the Web
- **DOI** is the de facto standard to identify each electronic document
 - The best-known international standard infrastructure that assigns **persistent and unique identifiers** for any type of objects
 - The total number of DOIs is about **130 million** (as of November 2015)*

* <https://www.doi.org/factsheets/DOIKeyFacts.html>

Why do we analyze DOI links on “Wikipedias” ?

- **The 5th largest referrer of DOI links is Wikipedia**

- CrossRef, the largest DOI Registration Agency, reports that Top 4 referrers of CrossRef DOIs are academic literature databases, and the 5th referrer is Wikipedia (as of 2015) *.

- **Wikipedia seems to ...**

- **Build and enhance a bridge between Web users and scholarly information through DOI links**
- Redound to make the best use of scholarly information — not only by researchers or specialists, but also by more various people such as students and general public
- **But few studies have attempted to analyze scholarly information referenced on Wikipedia**

* <http://www.slideshare.net/CrossRef/geoffrey-bilder-crossref15>

The reasons why this study sets targets on **enwiki**, **jawiki**, and **zhwiki**

- **Enwiki** is the largest language version of Wikipedia, so it is meaningful to identify **its influence on jawiki**
- If some similarities or common points are observed between jawiki and enwiki, **we should check whether it is also seen on other language Wikipedias or not**
- **Jawiki** and **zhwiki** have some similarities in that both are **Asian languages**, and they are **equal in quantity of articles**

Research Questions

RQ1. Which publishers or academic societies have content that is highly referenced on Wikipedia?

RQ2. Does the highly referenced content vary among Wikipedia languages, or is it very similar to other languages?

Related Work

- Analyses of academic/scientific citations on Wikipedia
 - (Nielsen, 2008) analyzed referenced journals in enwiki
 - (Lin & Fenner, 2014) analyzed referenced contents published by PLOS on Wikipedia
- DOI usage analyses by CrossRef
 - by using DOI access log
 - 5th largest referrer of DOI links is Wikipedia (as of 2015)
- Analyses of Wikipedia external links
 - investigate characteristics of external links and dead links

About DOI

- Each DOI consists of a **prefix**, a **slash** (/), and a **suffix**.
 - ex) **10.1002/asi.23209**
- DOI also provides hyperlinks (DOI links) by adding DOI after “**http://doi.org/**” or “**http://dx.doi.org/**.”
DOI links redirect to each original content’s URI.
 - ex) **http://doi.org/10.1002/asi.23209**
→ <http://onlinelibrary.wiley.com/doi/10.1002/asi.23209/abstract>
- A prefix is assigned to a particular DOI registrant, such as publishing companies or academic societies.
 - ex) **10.1002** is **Wiley-Blackwell’s** prefix

About DOI

- DOIs are registered through **DOI Registration Agencies (RAs)**
- Some RAs that handle scholarly resources are **CrossRef**, **JaLC**, and **ISTIC**
 - **CrossRef** is the largest RA
 - **JaLC (Japan Link Center)** is the only RA in Japan
 - **ISTIC** is a RA in China

DOI links on Wikipedia

Lion (enwiki) https://en.wikipedia.org/wiki/Lion#Hunting_and_diet

Because lionesses hunt in open spaces where they are easily seen by their prey, cooperative hunting increases the likelihood of a successful hunt; this is especially true with larger species. Teamwork also enables them to defend their kills more easily against other large predators such as hyenas, which may be attracted by [vultures](#) from kilometres away in open savannas. Lionesses do most of the hunting; males attached to prides do not usually participate in hunting, except in the case of larger quarry such as giraffe and buffalo. In typical hunts, each lioness has a favoured position in the group, either stalking prey on the "wing" then attacking, or moving a smaller distance in the centre of the group and capturing prey in flight from other lionesses.^[99]



Four lionesses catch a cape buffalo. 

99. ^ Stander, PE (1992). "Cooperative hunting in lions: the role of the individual"  (PDF). *Behavioral Ecology and Sociobiology*. **29** (6): 445–54. doi:[10.1007/BF00170175](https://doi.org/10.1007/BF00170175).

<https://dx.doi.org/10.1007/BF00170175>



Materials and Methods

Datasets

- We used following Wikipedia Data dumps
 - the English dump file on **March 4, 2015**
 - the Japanese on **March 13, 2015**
 - the Chinese on **March 4, 2015**
- The extraction conditions
 - only in main namespace pages (namespace = “0”)
 - URIs of external links contained “doi.org” in the el_to column of externallinks.sql
 - the prefix of interwiki links equaled to “doi” in the iwl_prefix column of iwlinks.sql
 - removed non-DOI links

Datasets Overview

Language	No. of total DOI links	No. of unique pages	No. of unique DOI links
enwiki	1,474,230	166,490	519,736
jawiki	28,799	9,750	25,444
zhwiki	36,669	9,676	28,177

Methods

- We performed a detailed analysis of DOI links on each language Wikipedia through the following three analyses:

- 1. Prefix-level analysis**

- 2. Overlap analysis of unique DOI links between two language Wikipedias**

- 3. Comparison of DOI links through interlanguage links and page-revision histories**

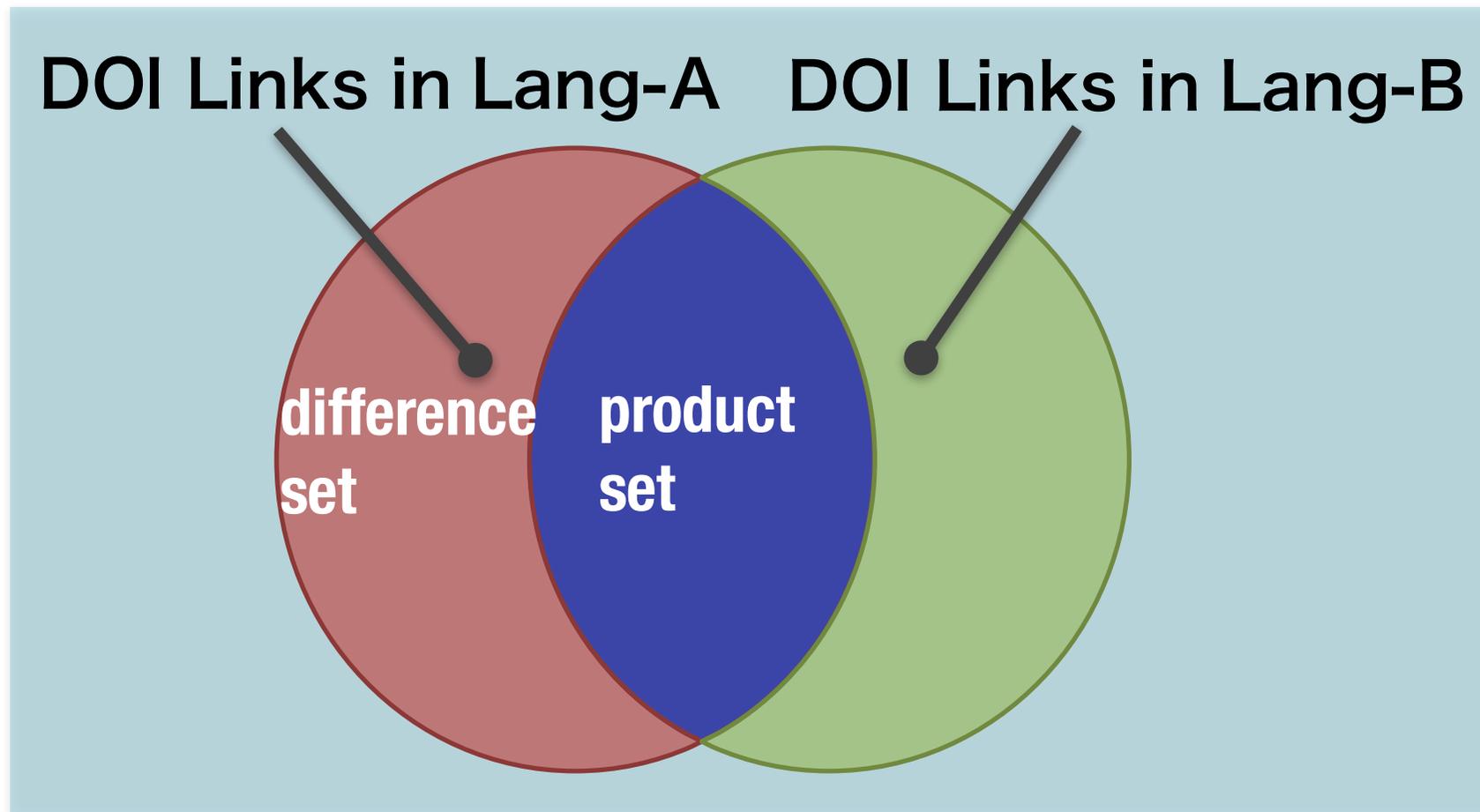
1. Prefix-level analysis

We counted each prefix to clarify which registrant's content is most commonly referenced.

We used CrossRef REST API* to identify registrants from prefixes

* <http://api.crossref.org/>

2. Overlap analysis of unique DOI links between two language Wikipedias



3. Comparison of DOI links through interlanguage links and page-revision histories

Some DOI links seemed to be added to enwiki, before they were first added to jawiki or zhwiki pages. Thus, we extracted common DOI links through the following four steps:

- **STEP1**: We extracted DOI links, written in main namespace pages on each language Wikipedia.
- **STEP2**: We extracted **the pages that have interlanguage links to enwiki (correspondent pages)** and DOI links written on these pages.

ライオン (jawiki)	↔	Lion (enwiki)
獅 (zhwiki)	↔	Lion (enwiki)

correspondent page

3. Comparison of DOI links through interlanguage links and page-revision histories

- **STEP3**: We extracted the pages that have common DOI links with the correspondent page — and the DOI links written on these pages.
- **STEP4**: We extracted the pages that have **10 or more common DOI links with the correspondent page**. This extraction condition, sharing 10 or more DOI links, was set on the basis of data observation.

<u>ライオン (jawiki)</u>		<u>Lion (enwiki)</u>
10.1007/BF00170175	↔	10.1007/BF00170175
10.1007/s10344-005-0008-0	↔	10.1007/s10344-005-0008-0
10.1007/s10592-005-9062-0	↔	10.1007/s10592-005-9062-0
10.1017/S0952836905007508	↔	10.1017/S0952836905007508
10.1038/436927a	↔	10.1038/436927a
10.1086/284097	↔	10.1086/284097
10.1126/science.1073257	↔	10.1126/science.1073257
10.1126/science.271.5253.1215a	↔	10.1126/science.271.5253.1215a
10.1126/science.7652566	↔	10.1126/science.7652566
10.1126/science.7652573	↔	10.1126/science.7652573
.....	

A workflow of comparison of DOI links between different Wikipedia language

STEP1

ALL pages with any DOI links

No. of total DOI links
No. of unique pages

STEP2

The pages with a langlink to enwiki

No. of total DOI links
No. of unique pages

STEP3

The pages with one or more common DOI links to enwiki

No. of total DOI links
No. of unique pages

STEP4

The pages with common DOI links greater than or equal to 10

No. of total DOI links
No. of unique pages

The pages with no langlinks to enwiki

No. of total DOI links
No. of unique pages

The pages with no common DOI links to enwiki

No. of total DOI links
No. of unique pages

The pages with common DOI links less than 10

No. of total DOI links
No. of unique pages

An example of edit summary that mentions translation from enwiki

- (cur | prev) ○ 18:27, 10 November 2010 4 K (talk | contribs) .. (116,036 bytes) **(+95,384)** .. (en:Lion (07:18, 18 October 2010 UTC) から抄訳している) (undo | thank)



Results and Discussion

Result: The number of total DOI links for RAs

RA	enwiki	jawiki	zhwiki
AIRITI	2	0	0
CrossRef	1,463,052	27,900	36,202
DataCite	464	13	6
ISTIC	101	0	44
JaLC	9	549	0
mEDRA	647	5	9
OPOCE	176	2	3
Public	367	6	25
Error	9,412	324	380
Total	1,474,230	28,799	36,669

- Most of DOI links in these Wikipedia are **CrossRef DOIs**
- The second most-referenced DOI links in enwiki are mEDRA DOIs; those in **jawiki are JaLC DOIs**; those in **zhwiki are ISTIC**

Result: Prefix-level analysis

Top-5 Prefixes in enwiki (n=1,474,230)

Rank	Prefix	Registrant	Count	%
1	10.1016	Elsevier BV	245,360	16.6
2	10.1038	Nature Publishing Group	97,943	6.6
3	10.1007	Springer Science+Business Media	87,107	5.9
4	10.1111	Wiley-Blackwell	71,629	4.9
5	10.1093	Oxford University Press	67,657	4.6

Top-5 Prefixes in jawiki (n=28,799)

Rank	Prefix	Registrant	Count	%
1	10.1016	Elsevier BV	4,565	15.9
2	10.1021	American Chemical Society	1,915	6.6
3	10.1007	Springer Science + Business Media	1,796	6.2
4	10.1002	Wiley-Blackwell	1,497	5.2
5	10.1038	Nature Publishing Group	1,497	5.2

Top-5 Prefixes in zhwiki (n=36,669)

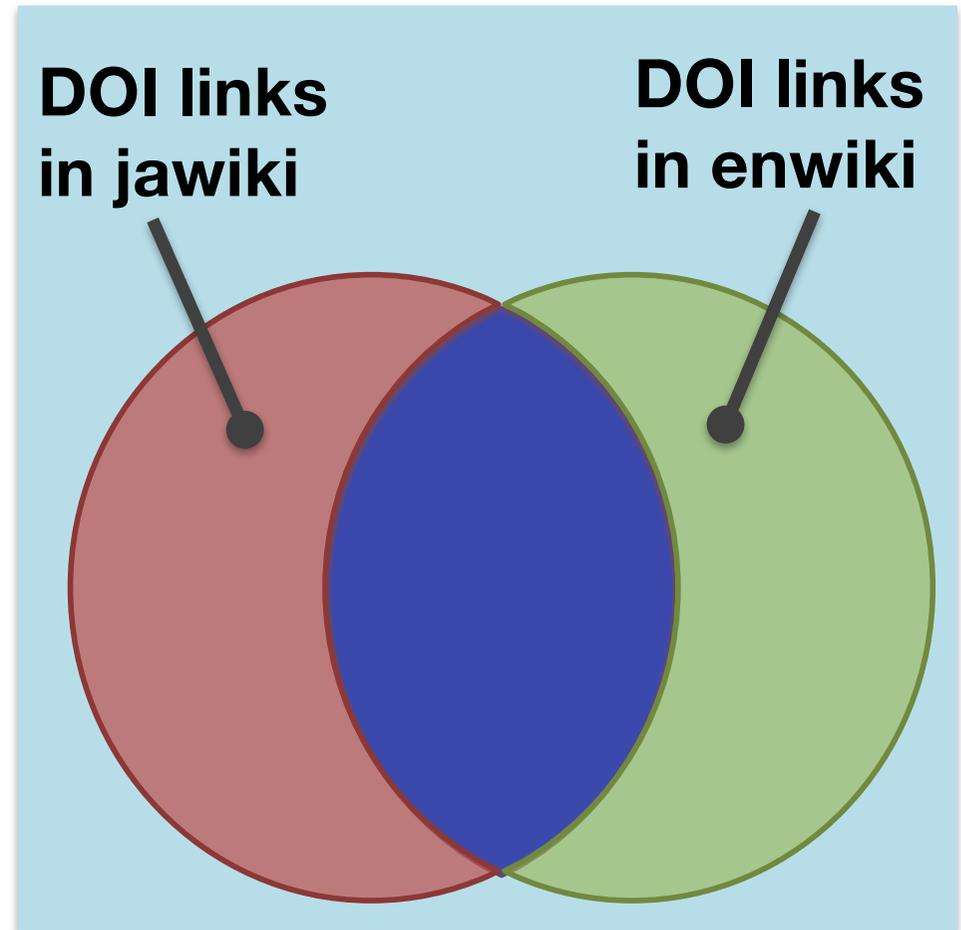
Rank	Prefix	Registrant	Count	%
1	10.1016	Elsevier BV	5,165	14.1
2	10.1021	American Chemical Society	2,588	7.1
3	10.1086	University of Chicago Press	2,530	6.9
4	10.1038	Nature Publishing Group	2,327	6.3
5	10.1002	Wiley-Blackwell	2,180	5.9

Result: Overlap analysis of unique DOI links between two language Wikipedias

Target	jawiki - enwiki	enwiki - jawiki	zhwiki - enwiki	enwiki - zhwiki	zhwiki - jawiki	jawiki - zhwiki
difference set	5,259	499,551	2,022	493,581	20,774	23,507
%	20.7	96.1	7.2	95.0	81.6	83.4
product set	20,185	20,185	26,155	26,155	4,670	4,670
%	79.3	3.9	92.8	5.0	18.4	16.6
total	25,444	519,736	28,177	519,736	25,444	28,177
%	100.0	100.0	100.0	100.0	100.0	100.0

Result: Overlap analysis of unique DOI links between two language Wikipedias

Target	jawiki - enwiki	enwiki - jawiki
difference set	5,259	499,551
%	20.7	96.1
product set	20,185	20,185
%	79.3	3.9
total	25,444	519,736
%	100.0	100.0



Result: Overlap analysis of unique DOI links between two language Wikipedias

Target	jawiki - enwiki	enwiki - jawiki	zhwiki - enwiki	enwiki - zhwiki	zhwiki - jawiki	jawiki - zhwiki
difference set	5,259	499,551	2,022	493,581	20,774	23,507
%	20.7	96.1	7.2	95.0	81.6	83.4
product set	20,185	20,185	26,155	26,155	4,670	4,670
%	79.3	3.9	92.8	5.0	18.4	16.6
total	25,444	519,736	28,177	519,736	25,444	28,177
%	100.0	100.0	100.0	100.0	100.0	100.0

Result: Overlap analysis of unique DOI links between two language Wikipedias

Target	jawiki - enwiki	enwiki - jawiki	zhwiki - enwiki	enwiki - zhwiki	zhwiki - jawiki	jawiki - zhwiki
difference set	5,259	499,551	2,022	493,581	20,774	23,507
%	20.7	96.1	7.2	95.0	81.6	83.4
product set	20,185	20,185	26,155	26,155	4,670	4,670
%	79.3	3.9	92.8	5.0	18.4	16.6
total	25,444	519,736	28,177	519,736	25,444	28,177
%	100.0	100.0	100.0	100.0	100.0	100.0

Result: Comparison of DOI links through interlanguage links and page-revision histories

Language	ALL		The pages with a langlink to enwiki		The pages with one or more common DOI links to enwiki		The pages with common DOI links greater than or equal to 10	
	No. of total DOI links	No. of unique pages	No. of total DOI links	No. of unique pages	No. of total DOI links	No. of unique pages	No. of total DOI links	No. of unique pages
enwiki	1,474,230	166,490	—	—	—	—	—	—
jawiki	28,799	9,570	26,987	9,118	20,599	7,122	6,133	327
zhwiki	36,669	9,676	35,099	9,351	31,161	8,579	12,915	634

Result: The number of DOI links that is identified as translation from enwiki or other language page

- About 88% of the common DOI links in the corresponding pages in jawiki were added by translating from enwiki.
- **A lot of DOI links in jawiki are added by translating from enwiki.**

Language	The pages with common DOI links greater than or equal to 10		The pages translated from enwiki		The pages translated from other language page except English		Unknown	
	No. of total DOI links	%	No. of total DOI links	%	No. of total DOI links	%	No. of total DOI links	%
jawiki	6,133	100.0	5,413	88.3	49	0.8	671	10.9
zhwiki	12,915	100.0	1,479	11.5	408	3.2	11,028	85.4

Result: The number of DOI links that is identified as translation from enwiki or other language page

- **85% DOI links in zhwiki were added with no information about translation in edit summaries.**
- Due to translation guidelines in zhwiki.

Language	The pages with common DOI links greater than or equal to 10		The pages translated from enwiki		The pages translated from other language page except English		Unknown	
	No. of total DOI links	%	No. of total DOI links	%	No. of total DOI links	%	No. of total DOI links	%
jawiki	6,133	100.0	5,413	88.3	49	0.8	671	10.9
zhwiki	12,915	100.0	1,479	11.5	408	3.2	11,028	85.4

Result: The number of DOI links that were added in enwiki before they were first added to the page

- 98% DOI links in jawiki — and about 99% DOI links in zhwiki — that were added to enwiki before they were first added to the page
- **The majority of DOI links in zhwiki are thought to be written through derived enwiki**

Language	The pages with common DOI links greater than or equal to 10		The DOI links were added in enwiki before they were first added to the page		Unknown	
	No. of total DOI links	%	No. of total DOI links	%	No. of total DOI links	%
jawiki	6,133	100.0	6,024	98.2	109	1.8
zhwiki	12,915	100.0	12,808	99.2	107	0.8 ³⁰



Conclusion

Conclusion

RQ1. Which publishers or academic societies have content that is highly referenced on Wikipedia?

- **Elsevier BV** is the largest registrant in all languages. **Nature Publishing Group** and **Wiley-Blackwell** are commonly referenced. The content hosted by these registrants is **shared among the Wikipedia communities**
- Most DOI links in these Wikipedias were **CrossRef DOIs**
- Scholarly contents in Japan tend to be referenced in jawiki, and contents in China tend to be referenced in zhwiki

Conclusion

RQ2. Does the highly referenced content vary among Wikipedia languages, or is it very similar to other languages?

- **Jawiki and zhwiki share the DOI links at a similar high rate with enwiki**
- **The majority of DOI links in jawiki and zhwiki were added by translating from enwiki**
- **These findings allow us to understand how scholarly references are added to Wikipedia and how to count them as altmetrics.**

ICADL2016 Tsukuba, Japan December 9th, 2016



DOI Links on Wikipedia

Analyses of English, Japanese, and Chinese
Wikipedias

Jiro KIKKAWA[†] **Masao TAKAKU**[‡] **Fuyuki YOSHIKANE**[‡]

[†] Graduate School of Library, Information and Media Studies, University of Tsukuba

[‡] Faculty of Library, Information and Media Science, University of Tsukuba

{ **jiro**, masao, fuyuki } @slis.tsukuba.ac.jp