

日本語版および英語版WikipediaにおけるDOIリンクの重複分析

著者	吉川 次郎, 佐藤 翔, 高久 雅生, 逸村 裕
著者別名	KIKKAWA Jiro, TAKAKU Masao, ITSUMURA Hiroshi
内容記述	第14回情報メディア学会研究大会 2015年6月27日 同志社大学 今出川キャンパス 良心館
雑誌名	第14回情報メディア学会研究大会発表資料
ページ	27-30
発行年	2015
URL	http://hdl.handle.net/2241/00125076

日本語版および英語版WikipediaにおけるDOIリンクの重複分析

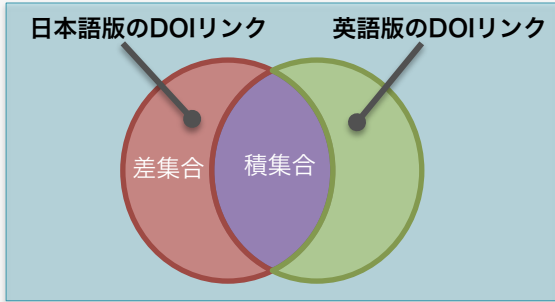
吉川 次郎 (筑波大学大学院) jiro@slis.tsukuba.ac.jp 佐藤 翔 (同志社大学) 高久 雅生 (筑波大学) 逸村 裕 (筑波大学)

背景 Webにおける学術情報流通の多様化とデジタル識別子の役割

目的 日本語版WikipediaにおけるDOIリンクが記述された経緯について明らかにする
(日本語版に独自に記述されたものであるか、英語版の翻訳を通じて記述されたものであるかを明らかにする)

- 考察
- 日本語版におけるDOIリンクのほとんどが英語版項目の翻訳を通じて記述されたものである
 - Wikipedia上での学術情報の参照を分析するには各言語版由来の学術情報の参照と他言語版の翻訳を通じた参照を区別する必要がある / Altmetricsスコア算出に用いると多重集計に繋がる可能性がある

分析1 DOIリンク単位での日本語版と英語版の重複



■ 日本語版と英語版の「項目数」と「DOIリンク数」

項目数	日本語版 2015年3月13日時点	英語版 2015年3月4日時点
のべDOIリンク数*	27,201	1,473,728
異なり項目数	9,135	166,368

* 標準名前空間、dx.doi.org または doi.orgを含むリンク

■ 異なりDOIリンクで見た日本語版と英語版の重複

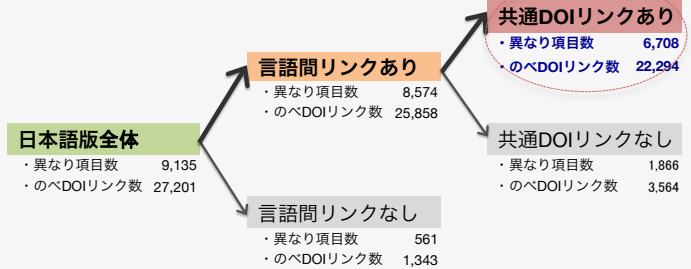
条件	日本語版	%	英語版	%
差集合	4,900	20.4	500,209	96.3
積集合	19,171	79.6	19,171	3.7
合計	24,071	100	519,380	100

分析1の結果

日本語版に記述されているDOIリンクのうち、**79.6%** 英語版と共通のDOIリンクが占める割合 (24,071件のうち19,171件)

96.3% 英語版に記述されているDOIリンクのうち、日本語版から参照されていないものの割合 (519,380件のうち500,209件)

分析2 英語版の翻訳を通じて記述されたDOIリンクの特定



「言語間リンクあり」かつ「共通DOIリンクあり」の条件で、6,708項目から300項目を無作為抽出、DOIリンク874件を分析

■ 分析2の手順: 「毒の歴史」を例に

毒の歴史

<http://dx.doi.org/10.1163/156851602760586671>

<http://dx.doi.org/10.1353/jmh.2003.0006>

(最新 | 前) 2010年11月2日 (火) 07:27 4K (会話 | 投稿記録) m.. (37,145バイト) (+37,145) .. (en:History of poison (21:35, 14 October 2010 UTC) から最後の節以外を翻訳している) (感謝)

(cur | prev) 13:16, 26 October 2010 Rambam rashi (talk | contribs) .. (38,529 bytes) (+1) .. (undo | thank)

History of Poison

<http://dx.doi.org/10.1163/156851602760586671>

<http://dx.doi.org/10.1353/jmh.2003.0006>

日本語版での初出時点 + 編集コメント

該当時点での英語版編集履歴

- ✓ 日本語版初出時点以前の英語版における当該DOIリンクの有無
- ✓ 日本語版初出時点の編集コメントに翻訳である旨の記述の有無

分析2の結果

97.9% 日本語版での初出時点以前に、英語版で当該DOIリンクが記述されている。日本語版項目執筆時に参照された可能性がある (874件のうち856件)

88.1% 日本語版の初出時点での変種コメントに翻訳である旨の記述がある。英語版項目の翻訳を通じて記述されたDOIリンクである (874件のうち770件)