

氏名(本籍)	吉川 次郎				
学位の種類	博士(図書館情報学)				
学位記番号	博甲第 9984 号				
学位授与年月日	令和 3 年 3 月 25 日				
学位授与の要件	学位規則第4条第1項該当				
審査研究科	図書館情報メディア研究科				
学位論文題目	Wikipediaにおける学術文献の参照記述に関する研究				
主査	筑波大学	教授	博士(教育学)	芳鐘	冬樹
副査	筑波大学	教授	文学修士	逸村	裕
副査	筑波大学	准教授	博士(情報学)	高久	雅生
副査	筑波大学	准教授	博士(教育学)	辻	慶太
副査	国立情報学研究所	教授	工学博士	武田	英明

論文の要旨 (2,000字程度)

電子ジャーナルを始めとしてウェブ上での学術情報流通は増加を続けており、伝統的な研究者以外の多様なコミュニティによる学術文献の利用が見られるようになってきている。本論文では市民参加型コミュニティの一つであるオンライン百科事典 Wikipedia を取り上げ、Wikipedia 上で学術文献への参照記述を追加する編集および編集者の分析を通じて、Wikipedia における参照記述追加の持続可能性を明らかにすることを目的としている。

本論文は6章から構成される。

第1章では研究背景として、学術ビッグデータとも言われる学術情報の増加と、研究者や専門家以外による学術文献の利活用として Wikipedia を取り上げ、その編集方針である検証可能性との関係から、学術文献への参照記述が多数みられ、重要なものとなっていることを指摘している。研究課題として、1) DOI (Digital Object Identifier) リンクを追加する編集および編集者の分析、2) 学術文献の参照記述を追加する編集の特定手法の構築、3) 参照記述の追加に関する時系列分析の3点を設定し、研究課題にこたえることを通じて、参照記述追加の持続可能性を考察するとしている。

第2章では文献レビューとして、Wikipedia 上の学術文献の参照記述の分析、参照記述の同定に関する方法論、Wikipedia の持続可能性、専門家や高等教育機関による Wikipedia の編集に関する関連研究について述べている。

第3章は研究課題1に対応し、DOIに着目して、DOIリンクを追加する編集および編集者の分析を行っている。対象として2017年3月時点の英語版 Wikipedia 記事における DOI リンクを抽出し、編集履歴から各ページでの初出時点を特定するとともに編集者の種別を判定している。対象とする学術文献は ISSN を用いて ESI (InCites Essential Science Indicators) の22研究分野と照合可能なものに限定している。

結果、ESI 研究分野と照合可能な約93万件の参照記述を取得でき、編集者の種別 User (編集アカウント保持者) と Bot (機械的な編集を行う自動化プログラム用アカウント) による編集全体のジ

二係数は 0.93 であり、全 22 分野に対しても、個別の分野の多くでも偏りが大きいことを示している。この要因として、既存の参照記述に DOI リンクを大規模かつ機械的に追加する者の存在が挙げられ、追加が少数特定の者に依存していることが明らかになったとしている。一方で、参照記述を新規追加する際に DOI リンクが記述される割合の推計値は高い分野でも約 4 割から 7 割弱に留まり、参照記述の初出時点を特定する方法論としては限界があることを示している。

第 4 章は研究課題 2 に対応し、参照記述の初出時点を高い精度で特定可能な方法論を提案し、その結果を評価している。具体的には、本文データ群への 3 つの手法の適用により初出時点候補を取得したうえで、編集日時が最も古いものを初出時点とすることにより、参照記述の初出時点データセットを構築している。3 つの手法は、識別子による照合（手法 A）、文献タイトルの全体または先頭 W 語による照合（手法 B）、学術文献データベースおよび Wikipedia 上の参照記述における文献タイトル間の類似度を用いた照合（手法 C）である。

評価実験では、研究課題 1 でのデータセットをもとに参照記述の初出時点データセットを構築し、精度を調査した結果、手法単体では手法 B (84.5%)、手法 C (70.7%)、手法 A (62.9%) の順に精度が高いこと、そして、複数の手法を組み合わせた場合では、提案手法 3 つの手法すべてを組み合わせたときの精度が 93.3% で最も高いことを示している。また、提案手法での研究分野ごとの精度は 20 分野で 90% 以上、残る 2 分野でも 80% 以上であり、分野を問わず高い精度で初出時点を特定できている。

第 5 章は研究課題 3 に対応し、参照記述の追加における持続可能性を明らかにするために、参照記述の初出時点を対象とした時系列分析を行っている。対象データとして、第 4 章で構築した参照記述の初出時点データセットを用いたうえで、大学での授業を通じた編集者を分析するために Wiki Education 参加者の特定を併せて行っている。

結果、参照記述の追加における時系列推移に関しては、編集者数、追加件数ともに伸び悩みは見られず、近年においても一定の規模を維持していることを示している。また、2011 年以降は毎年 4 月と 11 月に編集者が多くなる点で周期的な特徴が見られ、その要因として Wiki Education 参加者の影響が確認されたとしている。また、参照記述が追加されるまでのタイムラグに関しては、いずれの時差も後年になるにつれて短くなる傾向が見られたとしている。これらの結果から、参照記述の追加では新規参加者がコンスタントに生じ続けており、編集者数や追加件数には伸び悩みが生じていないと解釈でき、Wiki Education の活動が今後も継続するとすれば、この傾向は長期的に持続する可能性が高いとしている。

第 6 章では結論として、3 章から 5 章までの結果と考察をもとに、新規参加者の状況や編集者数と追加件数の規模、時系列推移の点から、参照記述の追加は、それ以外の編集も含めた全体に比べて持続可能性が高いと結論づけている。本研究の意義として、初出時点の特定手法が多様な研究の基盤として活用可能な点に加えて、Wikipedia コンテンツの質の間接的な向上につながる出典や情報源となる参照記述が継続的に追加されているという新たな知見を得た点を指摘している。

審 査 の 要 旨 (2,000 字以上)

【批評】

本論文は、オンライン百科事典 Wikipedia 上の記事において出典や参考文献として記載される学

術文献への参照記述がどのように追加編集され、どのような編集者がかかわってきたかを多角的に分析し、追加編集および編集者の規模、時系列推移の観点から Wikipedia 全体における参照文献の記載が持続可能かを明らかにすることを研究目的としている。

Wikipedia は多くの編集者が参加協働して執筆されるオンライン百科事典であり、ウェブ上でも調査学習の出発点として多くの利用者とアクセスを集める知識資源となっている。Wikipedia 上の参照記述は、検証可能性を満たすことにより、百科事典の記事内容の質を間接的に保証する存在となっているだけでなく、学術情報流通の中でも近年の *Altmetrics* のような代替的な研究評価指標でも取り上げられる対象となって久しい。このような役割をもつ参照記述がどのような過程を経て追加編集されているかを明らかにする研究はほとんど見当たらず、本研究の新奇性は高く、社会的な意義も認められ、時機に適ったものと考えられる。

2 章では関連する研究領域を、1) Wikipedia 上の学術文献の参照記述の分析、2) 参照記述の同定に関する方法論、3) Wikipedia の持続可能性、4) 専門家や高等教育機関による Wikipedia の編集に関する研究の 4 つにわけ、それぞれの既往研究事例とその成果を論じたうえで、本研究の位置づけと差異を説明している。学術情報流通の研究領域全体における位置づけとしてはやや物足りなくも感じられるものの、本論文における研究課題との関係の中で先行する研究成果がよく整理されており、先行する研究を把握して理解する力は十分と認められる。また、各章における先行研究も適切に参照され示されている。

3 章では、Wikipedia 上において DOI リンクを追加した編集とその編集者を特定して、編集者の偏りや占有度等の指標をもとに、編集追加の大部分を占有するような大量編集を行う一部の編集者の存在を明らかにしている。追加件数が最多の編集者は全分野からなる文献参照全体の約 5 分の 1 に及ぶような編集に関与しており、他にも機械的に DOI リンクを付加する Bot と呼ばれる少数の編集アカウントが大量の参照記述追加を行う様子を明らかにした。このように DOI リンクとそのメタデータを活用した編集が行われている様子を示したことは、ウェブ上で大量に流通する学術情報の新しいエコシステムとしての DOI システムの役割を示唆したとも言えるが、この点についての考察は十分でなく、より詳細な考察が望まれる。

4 章では、Wikipedia 上の参照記述を特定する方法論を提案している。参照記述の特定手法として複数の参照記述候補を抽出する手法が説明されており、筆者自身が人手で作成したデータセットを通じて評価し、参照記述の特定手法として 9 割超の高い精度が実現されており、汎用的な提案手法としての性能の高さが示されている。同時に、候補抽出手法におけるパラメータ選択や手法それぞれの得失についても丁寧に考察しており、方法論選択の妥当性も示されており、提案手法には学術的な価値が認められる。さらに、このようにして初出時点が特定されたデータセットは 5 章以降で分析されており、データセットそのものの有用性も示されている。

5 章では、前章で特定した参照記述の初出時点データセットを対象として、研究課題 1 では明らかにならなかった参照記述の初出時点までさかのぼって参照記述を新規追加した編集および編集者の実態を明らかにするための時系列分析を行っている。時系列推移を示す分析手法により、4 月と 11 月に編集追加が増加する特徴的な時系列トレンドを示している。また、参照記述が追加されるまでのタイムラグ分析を通じて、2007 年以降に参照記述が追加されるまでの時間差が減少していることも示している。これらの現象は従来明らかにされていなかったものであり、参照記述の追加が百科事典記事の質を間接的に担保する役割を考えると、そのコンテンツの質をとらえる新しい手法を示した意義があると言える。一方で、参照記述が追加される百科事典記事の主題や新鮮度との関係など、さらに考察を深める事柄は今後の課題として残されている。

6 章では結論として、参照記述の追加およびそれに関わる編集者の新規参加の度合い、規模、時系列推移の観点から、Wikipedia における参照記述の持続可能性は高いとしている。この結論を導くための持続可能性に関わる要因は、Wikipedia 上の編集の蓄積を分析したものであり、持続可能性に関わる要因の網羅性にはやや疑問も残るものの、得られた分析結果から結論に至る論理は一貫している。体系的で緻密な分析から妥当な結論が導かれており、今後の研究に対する意義も大きい。また、論文全体の章構成は一貫しており、文章表現にも大きな不備は見当たらない。

以上を総合的に判断すると、本論文は図書館情報学の学位論文として十分な内容を有すると認められる。

【最終試験結果】

2020 年 12 月 23 日、図書館情報メディア研究科学学位論文審査委員会において、審査委員全員出席のもと、本論文について著者に説明を求めた後、関連事項について質疑応答を行った。引き続き、「図書館情報メディア研究科博士後期課程（課程博士）の学位論文審査に関する内規」第 23 項第 3 号に基づく最終試験を行い、審議の結果、審査委員全員一致で合格と判定された。

【結論】

よって、本学位論文の著者は博士（図書館情報学）の学位を受けるに十分な資格を有するものと認める。