

# Wikipedia における独自研究テンプレートの利用実態の分析

坂本 寛人<sup>†</sup> 清田 陽司<sup>‡</sup> 山田 剛一<sup>‡</sup> 増田 英孝<sup>‡</sup>

東京電機大学大学院未来科学研究科<sup>†</sup> 東京電機大学未来科学部<sup>‡</sup>

## 1 はじめに

インターネット上の集合知として、オンライン百科事典の Wikipedia が存在する。Wikipedia は 2014 年 1 月現在、日本語版で 890,187 本の記事が存在する。Wikipedia は誰でも記事の閲覧、編集が可能であるが、事実無根の情報を書き込む編集者がいる可能性もある。信頼性を担保する方法のひとつとして、Wikipedia には記事の内容に関する方針が存在し、方針に従って記事を執筆すべきとされている。Wikipedia の方針が守られていない記事には「テンプレート」と呼ばれる定型文を貼ることで問題点があることを示し、修正を促すことが可能となっている。テンプレートは方針を理解している編集者によって貼り付けられ機能しているが、編集者の誤った考えによりテンプレートが貼り付けられる可能性も考えられる。本研究では、どの程度テンプレートが正しく機能しているのかを知るために、テンプレートがどのような状況で付与、削除が行われているかの利用実態の分析を行った。

## 2 Wikipedia の方針とテンプレート

Wikipedia には「テンプレート」が存在する。テンプレートとは、メッセージが書かれた独立したページで、別のページに容易に挿入することができる。テンプレートには複数の種類があるが、その中には「方針から逸脱していることを知らせるテンプレート」が存在する。テンプレートの使用例を図 1 に示す。

### フランチイズ

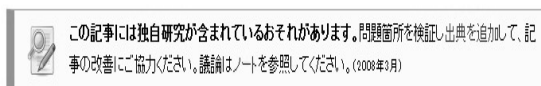


図 1. テンプレートの使用例

Investigation of Actual Conditions of 'Original Research Template' in Wikipedia

<sup>†</sup>Hiroto Sakamoto, Graduate School of Science and Technology for Future Life, Tokyo Denki University

<sup>‡</sup>Yoji Kiyota, Koichi Yamada, Hidetaka Masuda, School of Science and Technology for Future Life, Tokyo Denki University

また、Wikipedia には「三大方針」と呼ばれる記事を執筆する上での方針が存在する。以下が三大方針の内容となっている。

### (1) 中立的な観点

記事に偏りが無く平等な観点から書かれている必要があり、特定の観点を優先または排除して記事を作成してはならない。

### (2) 検証可能性

新聞や論文など、広くから信頼されている情報源の内容のみを記事にする。

### (3) 独自研究は載せない

個人的な知識から新しい用語を定義し、独自の記事を作成してはならない。

上記の 3 つの方針を守って記事を執筆すべきとされている。Anderka ら [1] は、「独自研究は載せない」に違反していることを知らせるテンプレートは、英語版 Wikipedia における「記事を修正すべきことを知らせるテンプレート」の中で 10 番目に多いと述べている。三大方針の中では「検証可能性」についてのテンプレートが最も多く、2 番目が「独自研究は載せない」についてのテンプレートであった。「検証可能性」については出典の有無でテンプレートを貼り付けるので利用は容易だが、独自研究に関しては判断が難しいと我々は考えた。そのため、本研究では「独自研究は載せない」に関するテンプレートが貼られた記事に着目し、記事の履歴からテンプレートが記事にどのように貼られているのかの実態を調査した。

## 3 独自研究テンプレートの実態調査

Wikipedia には全ての記事の履歴のダンプデータが公開されており、誰でも利用することができる。テンプレートの利用実態の分析にはダンプデータを利用した。調査には 2013 年 3 月のダンプデータを利用した。ダンプデータに含まれている情報を表 1 に示す。ダンプデータの情報を元に、独自研究テンプレートが貼られた記事の履歴を調査した。最初に、テンプレートの付与が 1 つの記事に対して何回行われているのか

表1. ダンプデータの情報

	記事総数	履歴総数
全記事	848,726	37,071,400
独自研究テンプレート が含まれる記事	5,793	1,165,193

を調査した。テンプレートの付与回数によって傾向が異なると考え、

- ① テンプレートが一度貼り付けられ、その後削除がされていない記事
- ② テンプレートの付与、削除が繰り返されている記事

の2種類に分類した。さらに、回数ごとにテンプレートの付与、削除に関わったユーザの傾向が異なると考え、「Wikipedia でアカウントを登録しているユーザ（以下登録ユーザ）」と「アカウントを登録していないユーザ（以下非登録ユーザ）」の2種類に分類し、それぞれ調査を行った。次に、テンプレートの除去についても同様に①と②の調査を行った。最後に、テンプレートの付与または削除を行ったユーザが、どの程度記事の編集を行っているかを知るために、ユーザごとの編集回数を調査した。

### 3.1 テンプレート付与時のユーザの情報

テンプレート付与時のユーザごとにおける結果を表2に示す。

表2. テンプレート付与の情報

	①	②	合計
記事数	4,938	855	5,793
追加総数	4,938	2,442	7,380
登録ユーザによる追加	3,526	1,817	5,343
非登録ユーザによる追加	1,412	625	2,037

### 3.2 テンプレート削除時のユーザの情報

テンプレート削除時のユーザごとにおける結果を表3に示す。

表3. テンプレート削除の情報

	数
記事数	855
削除総数	1,592
登録ユーザによる削除	764
非登録ユーザによる削除	828

### 3.3 ユーザごとの編集回数

テンプレートの付与、削除を行ったユーザが、記事に対してどの程度編集を行っているかを計測した。編集の貢献度として、記事に対して1回編集をしているか、2回以上編集を行っているかに分類した。付与に関する結果を表4に、削除に関する結果を表5示す。

表4. 追加したユーザの編集回数

編集回数	登録ユーザ	非登録ユーザ	合計
1回	2,768	1,491	4,259
2回以上	2,571	550	3,121

表5. 削除したユーザの編集回数

編集回数	登録ユーザ	非登録ユーザ	合計
1回	629	192	1,491
2回以上	570	393	963

## 4. 考察

調査の結果、①に分類される記事が5,793件中4,938件と多く、約85%の記事は①であることが分かった。このことから、テンプレートが一度追加された場合は削除されずに残りやすいという傾向が見られた。また、テンプレートを付与する際は非登録ユーザよりも登録ユーザの方が多いことが分かった。テンプレートを付与するユーザは、付与した記事に対しての編集回数が1回のみであることが多く、テンプレートの付与以外の編集は行っていない傾向が見られた。

テンプレートの削除に関しては、登録ユーザよりも非登録ユーザの割合が多いことが分かった。特に非登録ユーザはテンプレートを削除した記事に対しての編集回数が多く見られた。

## 5. おわりに

本研究では、Wikipediaの独自研究テンプレートがどのように付与、削除されているかの利用実態を記事の履歴から分析した。今後は履歴のテキスト情報などを利用し、どのような時にテンプレートの付与や削除が行われるのかを様々な角度から分析を行いたい。

## 参考文献

- [1]Maik Anderka, Benno Stein, and Nedim Lipka: Predicting Quality Flaws in User-generated Content: The Case of Wikipedia, *Proceedings of the 35th international ACM conference on research and development in information retrieval (SIGIR'12)*, 2012.