

研究ノート

## BibTeX 書式とコンピュータ言語 Ruby を用いた 無料・発展型の文献管理方法

遠藤大二\*

Development of reference-management system using BibTeX-format.

Daiji ENDOH\*

(Accepted 9 December 2021)

### はじめに

文献作成においては文献管理が重要な役割を果たす。論文においては、論旨の蓋然性を既知の文献で示した上、新規の実験、調査結果および開発などのオリジナル情報を記載する必要がある。適切な参照は、論文に述べた研究が研究者自身の個人的考えではなく、従来の研究に基づいた信頼性の高い新規知見であることを示すのに必要である<sup>1)</sup>。研究論文として参照される文献の調査は従来の研究で満たされていない研究上の課題から研究を立案するため、通常、研究立案段階から始まる。そのため、文献リストは、ある程度長期に保存可能である必要がある<sup>2)</sup>。Mendeley や EndNote などの文献管理ソフトウェア (Reference Management Software; RMS) は、使用者がオリジナルのデータベースを作成して保管することを可能にしている<sup>3,4)</sup>。また、Web 上でそれらのデータベースを複数のユーザーが共有することも可能にしている。

論文で参考文献を参照するためには、RMS を用いて研究者が文献情報をデータベース化する。PubMed および Google Scholar などの文献情報データベースでは、各 RMS にインポート可能な書式がダウンロード可能となっている<sup>2)</sup>。論文作成者は文献情報データベースからデータベースに取り込んだ文献のうち作成中の論文で引用する文献の ID を参照位置に挿入する。論文を仕上げる段階においては、RMS を用いることで、投稿予定の雑誌で指定する書式のリストが作成されるとともに文献を参照している位置に適切な参照記号が挿入される。

RMS ではユーザーごとに作成するデータベース

内で文献データを保管する必要がある、作業場所が異なる場合に備えて Web 上で共有する仕組みが整備されている。複数の研究者が協力して論文を作成する場合には、単一のデータベースを使用者間で共有して管理する必要がある<sup>3)</sup>。そのため、RMS を利用する上では文献データベースの適切な管理と、RMS の操作のある程度の習熟が必要となる。また、RMS を利用する場合には、文献を作成するワードプロセッサは RMS によって指定されるソフトウェアを使用する必要がある<sup>4)</sup>。現在は、Google Docs などを利用する機会があるが、その場合には RMS の利便性が大きく劣ってしまう。

一方、Google Scholar 等においてダウンロード可能な BibTeX 書式は、引用に必要な文献情報を全て含んでいるうえ、書式の 1 行目が ID として利用可能であるため、ダウンロードした文献データに研究者独自の ID を付与する必要が無い。そのため、適切なデータベース化ソフトを用意することにより、文献データベース管理を簡略化することが可能である<sup>5,6)</sup>。本論文では、BibTeX を用いた RMS に代替可能な文献管理プログラムの開発について述べる。

### RMS の処理内容

本研究の動機づけは、習熟を必要としない文献管理方法の必要性だった。共同研究者や指導している学生にも文献を探索して参考文献として利用することを試みていたが、代表的な文献管理ソフトについては、データベースの共有がうまく連動しない、学生が操作方法を習得するのに時間がかかりすぎるなどの問題を抱えていた。確実に進む方法として、参照文献の挿入は、教員が行い、学生はデータベース

\* 酪農学園大学獣医学群獣医学類放射線生物学ユニット

Laboratory of Radiation Biology, Department of Veterinary Medicine, School of Veterinary Medicine, Rakuno Gakuen University

連絡先 (Correspondence) : dendoh@rakuno.ac.jp

獣医学群獣医学類 遠藤大二

を作成を行って、論文完成直前にデータベースをマージするなどの方法を用いるようにした。このような作業手順では、既存の RMS では論文作成時の文献選択の教育が難しく、データベース作成の自由度が高いソフトウェアの必要性を感じていた。著者は研究でプログラム言語を用いていたため、自作での文献管理ソフトウェアの作成に着手した。

自作の RMS の開発にあたっては、まず代表的文献管理ソフトが行っている処理を分析した。処理方法は、EndNote x9 の利用を通じて収集し、Mendeley について確認した<sup>7,8)</sup>。文献管理ソフトの実施している内容としては、大きく二段階に分かれる。第1段階では、文献ライブラリーをユーザーが作成して、そのライブラリー内の文献を参照することを示す参考文献 ID を作成中の文章に挿入する。第2段階は、論文がほぼ完成した段階で行われる。文章中で参照されていた文献から投稿雑誌の規定に基づいて引用文献のリストが作成され、文章中に挿入されていた参考文献 ID は、作成されたリストの番号に置き換えられる。これらの処理は RMS との連携が設定されているワープロソフトで実施されるため、ユーザーには見えづらいが、RMS 指定外の Google Docs での参考文献の処理を行ったことで明らかになった。

RMS 的な文書処理を行うことを想定すると、文章中に挿入する参考文献 ID の形式が重要となる。既存の RMS では、ユーザー個々が作成するライブラリーに基づいて著者-作成年を含む参考文献 ID が生成されて、文章中に挿入されている。この参考文献 ID については、第2段階での文献リストを作成する際に RMS が参照するために、ライブラリー内では唯一性（ユニーク性）を持った ID になっている必要がある。RMS の具体的処理は非公開であるが、参考文献 ID と文献ライブラリーの間を RMS が正しく処理する必要からは参考文献 ID のライブラリー内での唯一性は必須の条件と考えられる。

## RMS 的文献管理ソフトの要件を満たす 文献情報書式

文献管理ソフトが行う処理について整理したところで、新規開発するソフトウェアが利用する文献情報が具備すべき要件を検討した。

まず、文献ライブラリーを作成するための要件としては、ライブラリー内での唯一性を保つことができる ID が必要と考えられた。その ID についてはソフトウェア側で付与する方法と文献検索サイトからダウンロードした時点で文献検索サイト側が提供している ID を利用する方法の2つが考えられた。文献検索サイト側で提供されている唯一性を持つ ID としては、PubMedID (PMID) などがある。ソフトウェア側で文献 ID を付与する場合には、ライブラリーを論文の作成者全員で共有しながら共通の文献 ID を利用するための仕組みも作成する必要がある。著者のソフトウェア開発能力を勘案すると、ダウンロード時点で唯一性のある文献 ID が文献ごとに設定されている形式が理想であることが結論付けられた。

続いて、ダウンロード時点で唯一性のある文献 ID を持つ論文情報形式を検討した。著者は当時最も頻繁に利用していた文献検索サイト Google Scholar において、ダウンロード可能な文献情報書式を検討した。文献を検索した後に表示される個々の文献情報の下には、ダブルクォーテーションのような記号が表示されており、それをクリックすると引用のための書式が表示される（図 1A）。ここで表示された書式のうち、MLA, APA および ISO690 については、雑誌名、著者名などの文献情報が加工されてすでに引用に適した形に変形されており、ライブラリー作成の際には、情報が失われてしまう可能性があるため、新たな文献管理ソフトが用いる文献基本情報の候補としては適さない（図 1B）。一方、BibTeX, EndNote および RefMan 書式は、必要な情報を全て構造化して含んでいるため、文献基本情報の候補としては、適していると考えられる（図 1C-E）。文献管理ソフトを作成する場合にはいずれ

図 1 Google Scholar におけるダウンロード可能な文献書式

### A. 文献表示画面

[HTML] Prediction of PCR amplification from primer and template sequences using recurrent neural network

K Kayama, M Kanno, N Chisaki, M Tanaka, R Yao... - Scientific reports, 2021 - nature.com

We have developed a novel method to predict the success of PCR amplification for a specific primer set and DNA template based on the relationship between the primer sequence and the template. To perform the prediction using a recurrent neural network, the ...

☆ 99 関連記事 全7バージョン BibTeX に取り込む

## B. 「引用」画面に表示される MLA, APA および ISO690 文献形式

×
引用

**MLA** Kayama, Kotetsu, et al. "Prediction of PCR amplification from primer and template sequences using recurrent neural network." *Scientific reports* 11.1 (2021): 1-24.

**APA** Kayama, K., Kanno, M., Chisaki, N., Tanaka, M., Yao, R., Hanazono, K., ... & Endoh, D. (2021). Prediction of PCR amplification from primer and template sequences using recurrent neural network. *Scientific reports*, 11(1), 1-24.

**ISO 690** KAYAMA, Kotetsu, et al. Prediction of PCR amplification from primer and template sequences using recurrent neural network. *Scientific reports*, 2021, 11.1: 1-24.

[BibTeX](#)
[EndNote](#)
[RefMan](#)
[RefWorks](#)

## C. BibTeX 書式

```
@article{kayama2021prediction,
  title={Prediction of PCR amplification from primer and template sequences using recurrent neural network},
  author={Kayama, Kotetsu and Kanno, Miyuki and Chisaki, Naoto and Tanaka, Misaki and Yao, Reika and Hanazono, Kiwamu and Camer, Gerry Amor and Endoh, Daiji},
  journal={Scientific reports},
  volume={11},
  number={1},
  pages={1--24},
  year={2021},
  publisher={Nature Publishing Group}
}
```

## D. EndNote 書式

```
%0 Journal Article
%T Prediction of PCR amplification from primer and template sequences using recurrent neural network
%A Kayama, Kotetsu
%A Kanno, Miyuki
%A Chisaki, Naoto
%A Tanaka, Misaki
%A Yao, Reika
%A Hanazono, Kiwamu
%A Camer, Gerry Amor
%A Endoh, Daiji
%J Scientific reports
%V 11
%N 1
%P 1-24
%@ 2045-2322
%D 2021
%I Nature Publishing Group
```

## E. RefMan 書式

```
TY - JOUR
T1 - Prediction of PCR amplification from primer and template sequences using recurrent neural network
A1 - Kayama, Kotetsu
A1 - Kanno, Miyuki
A1 - Chisaki, Naoto
A1 - Tanaka, Misaki
A1 - Yao, Reika
A1 - Hanazono, Kiwamu
A1 - Camer, Gerry Amor
A1 - Endoh, Daiji
JO - Scientific reports
VL - 11
IS - 1
SP - 1
EP - 24
SN - 2045-2322
Y1 - 2021
PB - Nature Publishing Group
ER -
```

の書式についてもライブラリーを作成することが可能と考えられた。ただし、EndNote および RefMan 書式には、唯一性を保つことができる文献 ID に相当する文字列が無いため、これらを利用する場合には、ライブラリー内で文献 ID を発行する仕組みを構築する必要が生じる。それに対し、BibTeX 書式の場合には、一行目の「@article {kayama2021prediction}」が文献 ID として使用可能であることが考えられた。この形式は、@マークにつづき、文献種別、記号、筆頭著者の姓、出版年および題名の 1 語目で構成されているため、同じ記述が複数存在する可能性は低い。その後、他の文献の BibTeX 形式を参照したが今のところ 1 件も BibTeX の一行目で同じ記述が記録されている文献はなかった。これらのことから、文献情報データベースに Google Scholar を利用し、BibTeX 書式をダウンロードすることにより、唯一性を持つ文献情報を入手可能であることが判明した。

#### BibTeX を用いた RMS 的文献管理ソフトの設計

書式の 1 行目を文献 ID として利用した場合、BibTeX 書式での文献情報は、そのままライブラリーとしての機能を持つことが可能である。BibTeX の一行目に唯一性があるかについては、文献報告は発見されなかったため、証明された事実ではないが、数人が参考文献を作成する上では問題になっていない。事実上唯一性を保った ID の存在により、Google Scholar から BibTeX 書式の文献情報をダウンロードすれば、そのダウンロードした文献情報をテキストとして保存しておくだけでライブラリーとして利用可能となる (図 2)。文献内に文献 ID として BibTeX の一行目を挿入しておくことにより、全く打ち合わせや共通ライブラリーの準備などの必要がなく、利用者間での共通性を保つことができる。文章内に別々の時点で別々の共著者が同じ

文献を参照した場合にも、文献リストを作成する段階でライブラリーを合わせれば、同じ文献としてみなすことが可能になる。この点で、Google Scholar の BibTeX を用いることにより、参照とライブラリーの作成という RMS の第 1 段階の処理は、ソフトウェアを一切使わずに満たすことが可能となる。結果的に RMS の第 1 段階の処理のためのソフトウェアは開発をしなかった。

続いて RMS 処理の第 2 段階の作業に相当するソフトウェアを設計した。第一にライブラリーを読み込み、1 行目の記述をキーとしてデータベースを作成する機能が必要になる。この作業は Ruby 言語の Hash データ形式を用いることにより、実現可能であった。第 2 に論文原稿中に存在する 1 行目の記述を選び出して、文献情報と照合する必要がある。このためには、確実に論文中の参照箇所について、プログラムが間違いなく検出する必要があった。「@article {kayama2021prediction}」のような記述は、論文内の“prediction”と誤解して検出してしまう可能性がある。それを防ぐ方法として、最も使用頻度の低い文字で文献 ID を囲むことにより、誤読を防ぐことを検討した。対象の文字を囲む文字列としては括弧が考えられるが、何種類かの括弧のうち《》の利用率が低いため、文章内の文献の引用は《@article {kayama2021prediction}》のように表現することとした。これにより、文章内の BibTeX の 1 行目は全て検出してライブラリーから選択することが可能となる。ライブラリーから、文章内で引用された文献を選択後は、Ruby によって、文献情報を「著者、題名 (出版年)」などのように整形して表示することにより、雑誌の指定する文献リストが出来上がる。雑誌の指定する著者の ABC 順などについては、Ruby のソート機能を用いて整列させることが容易である。この機能を満たす Ruby ソフトウェアの名称は、BibTeXRubyRMS とする。

図 2 BibTeX 形式でのライブラリー

```
@book{turabian2018manual,
  title={A manual for writers of research papers, theses, and dissertations: Chicago style for students and researchers},
  author={Turabian, Kate L},
  year={2018},
  publisher={University of Chicago Press}
}
@article{borg2000citation,
  title={Citation practices in academic writing},
  author={Borg, Erik},
  journal={Patterns and perspectives: Insights into EAP writing practice},
  pages={26-44},
  year={2000},
  publisher={Centre for Applied Language Studies, University of Reading Reading, UK}
}
```

## BibTeXRubyRMS のコーディング

前述の設計に基づいたプログラムコードは、2015年に原型が作成され、2016年から現在までの酪農学園大学放射線生物学ユニットから作成された卒業論文および投稿論文に利用し、そのたびにマイナーな改善を行っている。基本的な構成としては、一つのフォルダに配置された全てのライブラリーから重複の無い文献情報を Hash として取得するプログラム (construct\_library.rb) と、テキスト形式に変換された論文の文章部分を読み取り、《》内に配置された BibTeX 書式の 1 行目を検出して Hash と照合し、ライブラリー中から引用された文献を選定するプログラム (pick\_refpap.rb) と、引用された文献を雑誌の指定に合わせてリストとし、順番によって決定した番号と《》内に配置された BibTeX 書式の 1 行目を置換するプログラム (list\_replace.rb) で構成される。ソースコードについては、改善が進んでいるため、Github (<https://github.co.jp/>) 上に公開する。開発情報については、著者のホームページで公開する (<https://www.tobunotemaker.com/>)。

### BibTeXRubyRMS の運用と不足している機能

酪農学園大学放射線生物学ユニットでは、BibTeXRubyRMS という名称を意識することなく、所属学生と学外共同研究者が BibTeX 書式で文献情報を保存して論文中に文献 ID を挿入する方法で論文作成を実施している。現在のところ、引用文献リストの作成のためのプログラムは著者が実行している。Ruby プログラムの実行は、Windows および Mac でも可能だが、放射線生物学ユニットでは、実行環境が整っているため、Linux で実施している。共同研究者との連携においても、Google Scholar での検索・文献収集と BibTeX の記録を重複を気にする事なく各自が実施しており、大きな問題は起きていない。現在での最大の問題は Ruby の実行にある。ソースコードは公開するため、一般の研究者にもこの方法を用いてもらいたいが、実行環境が一般的ではないため、Web 上での機能の提供などを実現することが望ましい。

次の問題として、論文原稿を一度テキスト化してから BibTeXRubyRMS での処理をするため、作成中の文献で用いた肩字、添字およびボールド体などの書式を維持することはできない点がある。現在は、リスト作成前の論文ファイルをコピーとして残し、それと照合することにより、手動で問題を回避している。実際には、参考文献と関係の無い誤記が

存在するため、参考文献の引用番号・記号の確認は、文章の確認を兼ねて実施している。結果的に、酪農学園大学放射線生物学ユニットの学生・教員にとっては引用文献の書式の確認は負担感が小さい。

BibTeX 形式は、Google Scholar に掲載されている文献形式であれば、Book, Article, Review などの形式に対応している。しかしながら、ホームページの URL や日本語の論評などについては、書式が設定されていないため、BibTeXRubyRMS が正しく読み取ってリスト化することができない。特にホームページは、ソフトウェアの頒布サイトである場合があるため、参考文献としての掲載が必要になる。そのような場合は、プログラムの中に新たな処理コードを書いて対応するようにした。プログラム作成経験が無い研究者の場合には、クラウドワークス (<https://crowdworks.jp/>) などのサイトで開発者に依頼する事により、適性な価格での修正が可能と考えられる。また、投稿予定の雑誌のリストの書式が最新の BibTeXRubyRMS に対応していない場合にも、ソフト開発者に小さな開発案件として依頼することが可能である。

現状での BibTeXRubyRMS の最大の問題は Ruby プログラムの実行環境が必要という点である。Windows および MacOS の環境でも Ruby プログラムはインストール可能であり (<https://www.ruby-lang.org/ja/documentation/installation/>)、プログラムを実行するための端末プログラムも双方の OS に標準で装備されているため、BibTeXRubyRMS も機能的には問題無く可動するが、プログラムの実行自体に不慣れで不安を持つ研究者は多いことが想定される。今後、著者は論文テキストファイルと文献ライブラリーをアップロードすると BibTeXRubyRMS がリストと文章内の参照記号の生成を実施するサーバーの設置を計画している。進行状況については著者のホームページ (<https://www.tobunotemaker.com/>) で告知をする。

## ま と め

文献作成において重要となる文献管理・論文作成支援ソフトウェアの処理過程を分析したところ、これらのソフトウェアの実施している過程は、①文献のデータベース化、②データベース化された文献のキーとなる文字列の文献への挿入方法の指定、③文献の完成時における文献に挿入されたキーからの挿入された文献の情報のリストアップ、そして、④リストアップされた文献の雑誌規定に基づいたリストの生成、であることが整理された。これらのうち、

文献のデータベース化に必要となるのが①リスト作成に必要な情報を含み定型化された書式の存在と、②取得した情報の全文献にわたる唯一性の確保であることも整理された。これらの整理・分析に基づいて取得形式を検討したところ、Google Scholar データベース上の BibTeX 書式が文献リスト作成に必要なすべての情報を定型化して含んでおり、その一行目が全文献における唯一性を保持していることが判明した。著者は、この分析に基づいて、文献データベースとして BibTeX 書式のデータを記録し、文献内には BibTeX の一行目を参照文献特定符号として挿入する方式と、文献完成時にそれらを参照して、文献リストを作成するとともに、文献内の参照文献特定符号を雑誌で指定している参照記号に変換するソフトを開発し、BibTeXRubyRMS として無償で頒布される。当該ソフトウェアと使用方法は今後の研究論文作成に寄与することが期待される。

#### 参考文献

1. Borg, E: Citation practices in academic writing, Patterns and perspectives: in "Insights into EAP writing practice", 26-44, (2000)
2. Hensley, MK: Citation management software: Features and futures, Reference & User Services Quarterly, 50, 204-208, (2011)
3. Gilmour, R and Cobus-Kuo, L: Reference management software: A comparative analysis of four products, Issues in science and technology librarianship, 66, 63-75, (2011)
4. Basak, SK: Comparison of Researches' Reference Management Software: Refworks, Mendeley, and EndNote, Journal of Economics and Behavioral Studies, 6, 561-568, (2014)
5. Rauch, B, Leidner, J and Colloquium, IGK: Reference Management BibTeX and Beyond, (2007), Citeseer (Pub)
6. Fenner, M, Scheliga, K and Bartling, S: Reference management in "Opening Science", 125-137, (2014), Springer, Cham (Pub)
7. Fitzgibbons, M and Meert, D: Are bibliographic management software search interfaces reliable?: A comparison between search results obtained using database interfaces and the EndNote online search function, The Journal of Academic Librarianship, 36, 144-150, (2010)
8. Reiswig, J: Mendeley, Journal of the Medical Library Association: JMLA, 98, 193, (2010)

#### 利益相反

該当する利益相反は存在しない。