Automating citation formatting in scientific publications using ChatGPT

Aldeniz Rashidov
Department of Automation, information and control systems
Technical University of Gabrovo
Gabrovo, Bulgaria
https://orcid.org/0000-0003-4946-890X

Abstract - In academic writing, the accuracy of citation formatting in scientific publications is essential for maintaining the integrity and consistency of scientific communication. However, manually formatting citations according to different styles, such as IEEE, APA, or MLA, can be time-consuming and error-prone. This paper presents an innovative approach to automate citation formatting in scientific publications using ChatGPT. It proposes an algorithm that incorporates a sequence of instructions and guidance, combined with the capabilities of ChatGPT, and greatly simplifies the process of formatting citations according to different styles. The proposed approach involves training ChatGPT with a dataset containing citation guides and examples from different formatting styles to improve its ability to generate correctly formatted citations. This work presents a comparative characterization of existing automated citation formatting systems and the proposed algorithm with ChatGPT. Their functionalities are analyzed and their advantages and disadvantages are highlighted. In addition, a SWOT analysis of the systems is performed, which examines their strengths, weaknesses, opportunities and threats. The analysis highlights the effectiveness and advantages of the proposed solution with ChatGPT. The results show that automation using ChatGPT not only facilitates accurate citation formatting, but also offers a practical tool for improving the quality and relevance of scientific publications. ChatGPT can significantly reduce formatting errors and improve the efficiency of academic writing, offering a scalable solution for researchers and institutions.

Keywords - Citation, Formatting, ChatGPT, Automation, Publications, References, Algorithms.

I. INTRODUCTION

Accuracy and consistency in the formatting of citations in scientific publications are essential to ensure the reliability and reproducibility of research [1,2]. Proper citation not only acknowledges the sources of information, but also helps avoid plagiarism and makes it easier for readers to trace the sources used [3]. Modern citation automation systems, such as EndNote, Zotero, and Mendeley, offer numerous features that greatly facilitate the citation process [4]. However, they still have drawbacks, such as the need for manual intervention and limited support for certain styles [4-6].

Current trends in citation automation include integration with cloud services, real-time collaboration, and opportunities to use artificial intelligence (AI) [7]. Despite these innovations, existing systems are often limited by interface complexity, the need for training, and financial costs. Furthermore, the lack of full automation and the need for regular updates of stylistic requirements often lead to errors and inconsistencies [8].

Creating an algorithm that includes a sequence of instructions and guidance, combined with ChatGPT's capabilities, can greatly facilitate the generation of properly formatted references and assist authors in their process. This algorithm can be made available to authors via a link to a GPT assistant, which will allow them to easily generate correctly formatted references according to different citation styles without the need for mandatory integration with existing reference management systems.

The goal of this work is to present an algorithm to automate the formatting of citations using ChatGPT. This approach aims to assist researchers in generating accurate and compliant references, minimizing errors and improving the efficiency of scientific writing. In addition, the work provides a comparative characterization of existing systems and the proposed algorithm, as well as a SWOT analysis of their strengths, weaknesses, opportunities and threats. These analyses are intended to assist in understanding the advantages and disadvantages of different approaches, while helping to minimize errors and improve the effectiveness of scientific writing.

II. CITATION FORMATTING ALGORITHM WITH THE ASSISTANCE OF CHATGPT

The algorithm for automating citation formatting using ChatGPT was developed to facilitate the process of creating correctly formatted references following specific style requirements such as IEEE, APA, and others [9-15]. Through clearly defined steps, the algorithm allows authors to set formatting rules, input data, and use ChatGPT to finalize formatting and check conformance to the chosen style. The individual steps in the algorithm are discussed in detail, including how to set rules, process the input data, and generate the final citations. Each stage is presented with detailed instructions and examples that demonstrate the practical application of the algorithm. The block diagram of the algorithm is presented in Fig. 1.

A. Basic steps of the citation formatting algorithm with the assistance of ChatGPT

1) Setting formatting rules:

- Attach a file with examples and detailed information about formatting rules specific to the selected style (e.g., IEEE, MLA, etc.).
- An alternative option to provide formatting rules by:

  - *Integrated* citation style selection *interface*. Users can select their preferred citation style from a list of predefined options;

  - *An integrated wizard* that can consult a database of rules. This helper provides additional guidelines and examples, making it easier for the user to choose the correct style;

- *Manual entry of* specific rules into the system. For users with more specific requirements, the system allows manual entry and configuration of formatting rules.

This step is only necessary if ChatGPT does not have pre-loaded style data. When attaching or setting the rules, the author must declare that they are providing information about the rules and that they will ask a specific question later. If the Algorithm is integrated as a ChatGPT assistant and configured with this knowledge, then this step is omitted. Providing clear and detailed citation rules is essential for adapting the algorithm to the specific requirements of each style

2) Entering the citations and ask a question to ChatGPT:

- Entering the original citations into the system, including the necessary information for each citation (authors, title, source, etc.).
- Asking a specific question to ChatGPT to format the entered citation according to a specific style.
- The question may include clarification of the type of source (scientific article, book, website, etc.) as well as additional instructions, if available.

Entering original citations into the system involves providing the necessary information for each citation, such as authors, title, source, and other relevant data. This can be done by manual input or by importing data from files or other sources. This is an important step as the accuracy of the data here will determine the quality of the references generated. Often, data from various bibliographic databases or personal libraries may be included in this phase. The accuracy of the data entered is key to the subsequent correct formatting. In the absence of accuracy, errors may occur that will affect the overall scientific work. The author should declare that ChatGPT should only respond after asking a specific question.

3) Adapting the rules from ChatGPT:

- ChatGPT analyzes the question and the formatting rules provided and adapts the instructions according to the specific citation style.
- The algorithm automatically recognizes the specific requirements for information arrangement, use of italics, spacing and punctuation.

ChatGPT's ability to adapt and apply a variety of citation rules is a key advantage, allowing the system to serve a variety of styles and requirements. This reduces the likelihood of errors resulting from misinterpretation of style requirements.

4) Citation formatting:

- ChatGPT generates the formatted citations automatically using the provided data and the specified rules.

Automating formatting saves time and reduces manual work, which is especially useful for large bibliographies. However, it is important to perform a follow-up review for possible inconsistencies.

5) Review and correction:

- Review formatted citations for conformance to set rules and standards.

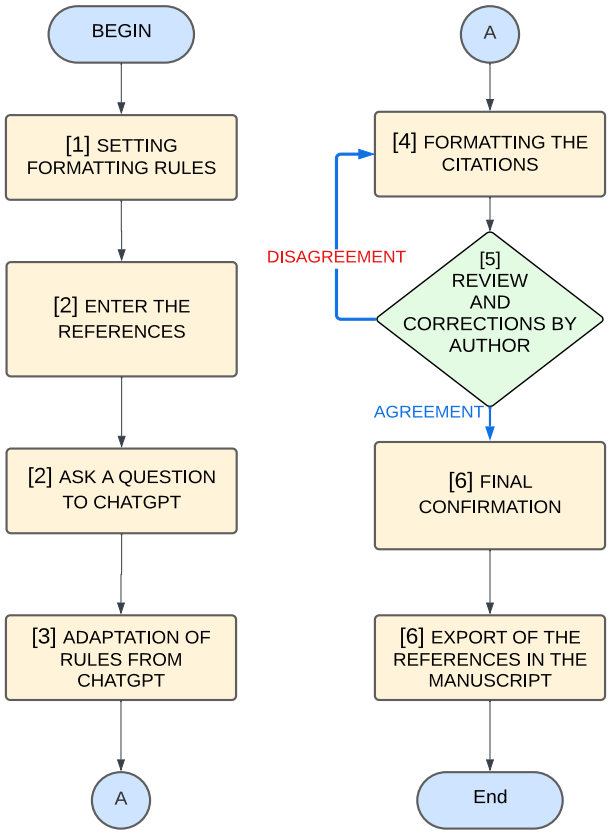

Fig.1. Block diagram of the citation formatting algorithm with the assistance of ChatGPT

- In case there is disagreement or errors:

  - *Reporting* identified issues and discrepancies to ChatGPT;

  - Problem *analysis*.

  - *Re-generating* the formatted citations from step 4 with corrections and changes reflected.

This step is critical to ensure the quality of the final product. Automation does not eliminate the possibility of errors, so manual inspection remains an important part of the process.

6) Final confirmation:

- Confirmation that final formatted citations are correct and conform to the specified style.
- After final review and confirmation that all citations are formatted correctly, the author exports the completed reference list to a text file or directly to his or her manuscript.

Final confirmation and export of citations ensures the process is complete and ready for inclusion in the scientific work.

Different academic and professional fields use different citation styles that have specific requirements for the arrangement of bibliographic data. Each citation style has its own specific rules for arranging information, punctuation, italics, and other elements. This creates challenges in automating the process, as the algorithm must precisely follow these rules for each style. The developed algorithm has been

integrated as ChatGPT assistant (under "My GPTs" in GPT-4 Omni) and trained to automatically recognize and apply the rules for styles such as APA, MLA, Chicago, IEEE, Harvard style, and many more styles [9-15]. Updates to ChatGPT play a key role in improving the accuracy and efficiency of the algorithm. With each new version of ChatGPT, the model gains access to a greater amount of data and advanced algorithms, which will allow better understanding and application of different citation styles. This not only reduces the likelihood of errors, but also increases compliance with current citation standards. However, ChatGPT updates can present challenges, such as the need to retrain the model and adapt to new functionalities.

B. *Example application of the algorithm for automating citation formatting using ChatGPT*

This section presents a practical example of the application of the algorithm to automate citation formatting using ChatGPT. Through a demonstration of the actual process, including setting formatting rules, entering citation data, and interacting with ChatGPT, the basic actions and communication with the system are illustrated step by step. This example is intended to show how authors can effectively use the algorithm to generate correctly formatted references, as well as to highlight the conveniences and advantages of automation in academic writing. The example follows the basic steps of the algorithm, presenting them in the context of a specific application.

1) Setting formatting rules:

- The author attaches a file with examples and detailed information about formatting rules specific to the IEEE style. He explains, "Please, I'm just attaching the rules here; I'll ask you a specific question later. Wait for me to type the citations!".

2) Entering the citations and asking a question to ChatGPT:

- The author enters the details of their sources, including authors, title of the article, journal name, year of publication, etc. This is done by importing data from a text file. The author explains, "Please, I am only attaching the citations here; I will ask you a specific question afterwards. Wait for me to enter all the data".
- After data entry, the author asks a specific question to ChatGPT: "Please format the following citations for a scientific publication in a conference proceeding, according to IEEE style. It contains mainly scientific papers and books, but there are a few websites as well!".

3) Adapting the rules from ChatGPT:

- ChatGPT analyzes the provided rules and the question asked, adapting the instructions according to IEEE style.
- The algorithm automatically recognizes specific requirements, such as information ordering, use of italics, spacing and punctuation, etc.

4) Citation formatting:

- ChatGPT generates the formatted citations automatically using the entered data and the rules. For example, book citations are formatted with italics for titles, and for scientific articles the strict order of authors, year, and title is followed.

5) Review and correction:

- The author reviews the formatted citations and notices that a title is missing from one of them. He reports to ChatGPT: "I noticed a title missing in one of the citations. Can you add it if you have the information?".
- ChatGPT checks and updates the missing data, then generates an updated version of the citations.

6) Final confirmation and export:

- After final review and confirmation that all citations are formatted correctly, the author exports the finished reference list directly into his or her manuscript.

III. SWOT ANALYSIS OF POPULAR SYSTEMS AND THE ALGORITHM WITH CHATGPT FOR AUTOMATING CITATION FORMATTING

The purpose of this section is to provide an in-depth analysis of existing systems for automating citation formatting, as well as the proposed algorithm with ChatGPT, using SWOT analysis. For the purpose of this SWOT analysis, several leading citation formatting automation systems are selected: EndNote, Zotero and Mendeley. These systems were chosen due to their significant popularity and prevalence in academia, the wide range of functionality offered, support for multiple citation styles, and the ability to be customizable and scalable. By examining the Strengths, Weaknesses, Opportunities and Threats of each system, this analysis aims to highlight the key advantages and disadvantages of different solutions. It also identifies potential areas for improvement and opportunities for future development, as well as challenges that the systems may encounter. In doing so, it provides a clear assessment of the effectiveness and applicability of the algorithm with ChatGPT compared to other well-established solutions, offering useful information for selecting the most appropriate solution for academic and professional needs. Tables 1÷4 present the SWOT analysis.

TABLE I. STRENGTHS OF THE SYSTEMS

System	Strengths
EndNote	- Integration with Microsoft Word. - Extensive database of styles.
Zotero	- Free and open source. - Intuitive interface.
Mendeley	- Cloud storage and synchronization. - Large number of active users.
Algorithm with ChatGPT	- High flexibility and adaptability. - Full automation through AI. - Easy integration. - Continuous improvement through learning. - Ease of use and convenience for users.

TABLE II. WEAKNESSES OF THE SYSTEMS

System	Weaknesses
EndNote	- Paid system. - Requires extensive training time.
Zotero	- Limited maintenance. - Lack of new features.
Mendeley	- Acquired by Elsevier, concerns about system independence. - Limited functionality for some styles.
Algorithm with ChatGPT	- Dependence on training data quality. - Need for periodic training and updating.

TABLE III. OPPORTUNITIES OF THE SYSTEMS

System	Opportunities
EndNote	- Supporting of cloud functions. - Integration with AI technologies.
Zotero	- Extensibility via plugins. - Opportunities for collaboration and user cooperation.
Mendeley	- Cloud services extension. - Integration with AI.
Algorithm with ChatGPT	- Expandability to new styles and features. - Integration with various academic platforms. - Reference recommendation.

TABLE IV. THREATS OF THE SYSTEMS

System	Threats
EndNote	- Competition from free solutions. - Technological changes leading to functional obsolescence.
Zotero	- Funding and development through donations. - Growing competition.
Mendeley	- Elsevier dependency. - Data security and privacy concerns.
Algorithm with ChatGPT	- Competition from new AI technologies. - Concerns about the security and ethical aspects of AI.

Several important conclusions can be drawn from the SWOT analysis. Systems for automating citation formatting offer different advantages and challenges, each with its own unique strengths and weaknesses.

The proposed algorithm with ChatGPT stands out for its high flexibility and full automation capability, making it a competitive solution against established systems such as EndNote, Zotero and Mendeley. However, it is also subject to some limitations, such as dependency on the quality of the training data and the need for regular updates.

The analysis highlights the potential opportunities for development and integration of the algorithm with other technology solutions, which could extend its functionalities and improve its performance. Threats related to the competitive market and technological change need to be carefully monitored to ensure that the algorithm remains relevant and competitive.

It can be concluded that the algorithm with ChatGPT offers significant potential to improve the process of automating citation formatting, while requiring careful management and continuous development to overcome existing challenges and maximize opportunities.

IV. COMPARATIVE CHARACTERIZATION OF POPULAR SYSTEMS AND THE ALGORITHM WITH CHATGPT FOR AUTOMATING CITATION FORMATTING

Table 5 presents a comparative analysis of major systems for automating citation formatting: EndNote, Zotero, Mendeley, and the proposed algorithm with ChatGPT. These systems have been chosen to provide an objective and comprehensive view of the most used tools on the market that play a significant role in the scientific community. This comparative analysis provides an objective assessment of each system, focusing on key aspects such as cost, integration, cloud storage, ease of use, supported citation styles, and other relevant criteria. By highlighting the main advantages and disadvantages of these systems, as well as identifying their unique features, the analysis aims to help users select the most appropriate system for their needs in different academic and professional contexts. Furthermore, it provides a clear picture of how the algorithm with ChatGPT is positioned against established solutions in the market, facilitating informed user choice.

Based on the comparative analysis of systems for automating citation formatting, several key conclusions can be drawn. Systems such as EndNote, Zotero, and Mendeley offer a variety of features and capabilities that meet the needs of different user groups.

TABLE V. COMPARATIVE SYSTEMS ANALYSIS

Criteria	EndNote	Zotero	Mendeley	Algorithm with ChatGPT
Price	Paid	Free	Free (with paid features)	Ability to integrate with free and paid tools
Type of software	Closed Code	Open Code	Closed Code	Flexible cross-platform integration
Integration	MS Word, Google Docs	MS Word, Google Docs, OpenOffice	MS Word, LaTeX, BibTeX	Easy integration with major word processing programs
Cloud storage	Yes	Yes	Yes	Potential for integration with cloud services
Ease of use	Requires training time	Intuitive interface	Easy to use	High user suitability
Supported styles	Many, including IEEE, APA, MLA	Many, including IEEE, APA, MLA	Many, including IEEE, APA, MLA	All basic styles, option for custom styles
Support	Paid maintenance	Limited official support	Good maintenance	Maintenance and additional resources
Additional features	Advanced bibliography features	Multiple plugins and extensions	Tools for collaboration and resource sharing	Automatic reference recommendation, citation analysis
Data security	High	High	Ownership concerns from Elsevier	High, when integrated with secure systems
Adaptability and flexibility	Limited to the capabilities of the software	High	High	Very high, AI and machine learning based
Accuracy	High, requires manual verification	Good, but possible mistakes in specific styles	Satisfactory, but sometimes bugs	Very high, full automation
Automation	Partial, requires manual update	Partial but with many automated functions	Partial but well integrated	Full automation through AI and machine learning

EndNote is notable for its powerful integration features with word processing programs and its extensive stylesheet database, but it requires payment and extensive training time. Zotero and Mendeley, as free options, offer ease of use and cloud storage, but have their limitations in terms of support and functionality.

The algorithm with ChatGPT stands out for its unique ability for full automation and flexibility in adapting to different citation styles. It offers a high degree of adaptability and can be easily integrated with a variety of academic and professional tools. The main advantages of this algorithm include its ability to automatically generate references, its ability to be improved through training, and its user-friendliness.

Disadvantages of the algorithm with ChatGPT include the dependence on the quality of the training data and the need for updating to keep style standards up to date. Despite these challenges, the algorithm demonstrates significant potential to improve the citation formatting process while providing a scalable and state-of-the-art solution.

The findings of this analysis highlight that the choice of the most appropriate system depends on the specific needs and preferences of the user. The algorithm with ChatGPT offers innovative capabilities and is particularly suitable for users seeking an automated and easy-to-use solution that evolves with technological advances and the needs of the academic community.

CONCLUSION

This work presents an innovative algorithm for automating citation formatting in scientific publications using ChatGPT. The main advantages of this algorithm include high flexibility and adaptability, which allows complete automation of the process and facilitates authors in creating accurate and tailored references. This approach significantly reduces the time and effort required for manual formatting while minimizing the risk of errors. The algorithm with ChatGPT offers a modern and scalable solution for automating citation formatting that can be of great benefit to researchers and institutions seeking to improve the quality and relevance of their scientific publications.

Along with the advantages, the algorithm has some limitations, such as the dependence on the quality of the training data and the need for regular updates to keep the style standards up-to-date. However, ChatGPT's potential for self-improvement through training with new data opens up opportunities for continued improvement in its accuracy and efficiency.

The next steps in algorithm refinement include expanding the training data base and improving accuracy and consistency with different citation styles. Also, ways to better understand and handle context in citation will be explored, especially in cases with more specific or unusual requirements. Additional future opportunities for integration with other AI technologies and how this may enrich the functionality of the algorithm may be considered. These steps will help overcome existing limitations and expand the possibilities for applying the algorithm in different academic and professional environments.

REFERENCES

[1] T. M. Annesley, "Giving Credit: Citations and References", Clinical Chemistry, vol. 57, no. 1, pp. 14–17, Jan. 2011. [Online]. Available: https://doi.org/10.1373/clinchem.2010.158048.

[2] A. Santini, "The Importance of Referencing", J. Crit. Care Med. (Targu Mures), vol. 4, no. 1, pp. 3-4, Feb. 2018. https://doi.org/10.1515/jccm-2018-0002.

[3] R. Ahmad, M. Afzal, and M. A. Qadir, "Pattern Analysis of Citation-Anchors in Citing Documents for Accurate Identification of In-Text Citations", IEEE Access, vol. 5, pp. 5819-5828, 2017. https://doi.org/10.1109/ACCESS.2017.2689925.

[4] C. Ivey and J. Crum, "Choosing the Right Citation Management Tool: Endnote, Mendeley, Refworks, or Zotero", J. Med. Libr. Assoc., vol. 106, no. 3, pp. 399–403, Jul. 2018. https://doi.org/10.5195/jmla.2018.468. Epub 2018 Jul 1.

[5] S. Kelly, S. Skripsky, and R. Winslow, "Not just for citations: Assessing Zotero while reassessing research", 2016. [Online]. Available: https://doi.org/10.37514/PER-B.2016.0834.2.14.

[6] D. Kalita, "Resource Description Capacity of Reference Management Software: A Case Study", Creativity, Innovation and Transformations in Libraries (SALIS 2016), pp. 232-237, 2016.

[7] "10 Best AI Tools for Academic Research in 2024", Academia Insider. [Online]. Available: https://academiainsider.com/10-best-ai-tools-for-academic-research-in-2024/. [Accessed: July 30, 2024].

[8] J. Tu, H. Hadan, D. Wang, S. Sgandurra, R. H. Mogavi, and L. Nacke, "Augmenting the Author: Exploring the Potential of AI Collaboration in Academic Writing", presented at GenAICHI: CHI 2024 Workshop on Generative AI and HCI, 2024. [Online]. Available: https://arxiv.org/abs/2404.16071. [Accessed: July 30, 2024]. https://doi.org/10.17605/OSF.IO/FVTMQ.

[9] "Reference Examples", APA Style, American Psychological Association. [Online]. Available: https://apastyle.apa.org/style-grammar-guidelines/references/examples#textual-works. [Accessed: July 30, 2024].

[10] "MLA Style", Modern Language Association. [Online]. Available: https://www.mla.org/MLA-Style. [Accessed: July 30, 2024].

[11] "The Chicago Manual of Style", The Chicago Manual of Style Online. [Online]. Available: https://www.chicagomanualofstyle.org/home.html. [Accessed: July 30, 2024].

[12] "IEEE Xplore Digital Library", IEEE. [Online]. Available: https://ieeexplore.ieee.org/Xplore/home.jsp. [Accessed: July 30, 2024].

[13] "Harvard referencing", The University of Sheffield Library. [Online]. Available: https://www.librarydevelopment.group.shef.ac.uk/referencing/harvard.html. [Accessed: July 30, 2024].

[14] "Citation Generator", Cite This For Me. [Online]. Available: https://www.citethisforme.com/us/citation-generator. [Accessed: July 30, 2024].

[15] "Citation Style Manuals", Virginia Tech University Libraries. [Online]. Available: https://guides.lib.vt.edu/find/citation-style-manuals/. [Accessed: July 30, 2024].