International Journal of Science and Informatics Technology (IJOSIT)
Vol 1, No 1, April 2026, pp. 30-35
https://www.ijosit.org/index.php/ijosit
ISSN: XXXX-XXXX

Implementation of Format-Based Steganography Method for Secret Message Insertion in Text Media

Angelus F. Luahambowo, Amelia Sanna Maria Hutauruk, Gaby Angel Frasella Sitorus
Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Katolik Santo Thomas Medan Sumatera Utara, Indonesia

Abstract. Information security is a very important aspect in digital-based data exchange, especially to protect confidential messages from unauthorized access. In addition to cryptography techniques, steganography is another approach that aims to hide the existence of the message itself so as not to arouse suspicion. This study aims to implement and analyze format-based steganography methods in text media as a means of inserting confidential messages. The format-based method works by utilizing changes in text format, such as the use of spaces, variations in letter writing, or certain characters, without changing the meaning or readability of the original text. The research stages include the process of embedding confidential messages into the host text and the extraction process to retrieve the hidden messages. The implementation results show that secret messages can be inserted and extracted well without causing noticeable visual changes to the text, making the existence of the message difficult to detect with the naked eye. Based on these results, it can be concluded that format-based steganography in text media is a simple but effective method for maintaining the confidentiality of information in text-based communication, especially in environments where the use of multimedia media such as images or videos is not possible.

Keywords: Rahasia Steganography, Format-Based Steganography, Information Security, Text Media, Secret Message Insertion

This is an open access article under theCC BY-NClicense

Corresponding Author:
Margan Rizkiano Ritonga
Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Katolik Santo Thomas Medan Sumatera Utara, Indonesia

1. Introduction

The rapid development of information and communication technology has changed the way humans interact and exchange information in their daily lives. Digital information has become an integral part of social, academic, business, and government activities. Data exchange via e-mail, digital documents, instant messaging applications, and other online platforms is fast and massive. Behind this convenience, there are various threats to information security, such as wiretapping, data theft, message manipulation, and misuse of information by unauthorized parties. This condition requires a data security mechanism that not only focuses on protecting the content of messages but is also capable of hiding the existence of messages so as not to arouse suspicion. In the context of information security science, this approach is very important because many cyber attacks begin with the detection of secret communications. Therefore, steganography has developed as a relevant information security technique because it is capable of hiding secret messages within certain storage media so that communication can take place covertly and more securely. The urgency of research in the field of steganography has increased along with the high dependence of society on text-based digital communication [1].

Steganography can be applied to various digital media, such as images, audio, video, and text. Among these media, text-based steganography has its own characteristics and challenges [2], [3]. Text media has relatively little data redundancy compared to images or video, so the space for inserting secret messages is very limited. In addition, small changes to the structure or format of text are often easily detected, both by human readers and by automated text processing systems [2], [4]. Various text steganography methods have been developed to overcome these problems, one of which is the format-based steganography method. This method utilizes changes in the format or layout of text, such as the use of spaces, tabs, new lines, capital letter variations, or certain characters, without changing the semantic meaning of the original text. Although format-based methods are considered more subtle and less disruptive to the message content, this approach still has a number of limitations. Some format-based techniques are vulnerable to text normalization processes, such as copying, editing, or document format conversion, which can remove the hidden message. In addition, the message insertion capacity in text media is still relatively small, and the message extraction process does not always produce consistent results. This condition indicates a gap in research on how to implement format-based steganography that is simple, effective, and remains reliable in various conditions of digital text usage [5], [6].

Based on these issues, this study aims to implement a format-based steganography method in text media as an effort to hide secret messages in digital communication. The focus of this study is on designing a systematic process for inserting and extracting secret messages, as well as evaluating the success of the method in maintaining the concealment of messages without changing the meaning of the original text. The main scientific contribution of this research is the

 ISSN: XXXX-XXXX

presentation of a structured and easily replicable implementation of format-based text steganography, which can be used as a reference for further research. In addition, this research provides a practical overview of the use of text steganography as a lightweight, flexible, and suitable alternative for securing information in digital document-based communication. The novelty of this research lies in its conceptual approach to applying format-based steganography in a simple yet functional manner, emphasizing a balance between message concealment, text readability, and ease of implementation. Thus, this research is expected to enrich the study of text-based steganography and make a real contribution to the development of more adaptive information security methods in the digital age.

2. Literature Review and Problem Statement

Format-based steganography is one approach to text steganography that focuses on hiding secret messages by modifying the visual appearance or format of text without changing the content, meaning, or semantic structure of the language. This approach utilizes the characteristics of text layout as a medium for insertion, such as changes in the number of spaces between words, line shifting, word shifting, capitalization variations, and the use of invisible characters (whitespace or invisible characters) [7]. Unlike image or audio steganography, which has high data redundancy, text steganography works on media with limited capacity, thus requiring more subtle and careful insertion techniques [8]. Previous studies have shown that format-based methods are capable of hiding messages because the changes made are difficult for human readers to detect and do not directly affect the readability of the text. In addition, this method can be applied to various types of digital documents, such as simple text files and word processor-based documents, as long as the original format structure is maintained. However, previous research has also revealed a number of limitations of format-based steganography. This method is highly dependent on the consistency of the text format, so that hidden messages have the potential to be lost if the text undergoes reformatting, copying, format conversion, or automatic editing [9]. On the other hand, the message insertion capacity is relatively small because it only utilizes certain format elements, and the message extraction process does not always produce a stable success rate. In addition, most studies still focus on conceptual studies or simple experiments, with limited system implementation and little attention to ease of use [10].

Although format-based steganography has been proven capable of hiding secret messages in text media without changing the meaning of the content, there are still a number of research gaps that need to be addressed. One of the main gaps is the suboptimal use of invisible Unicode characters as a more flexible and difficult-to-detect medium for inserting messages. Many previous approaches still rely on changes in spacing or layout that are vulnerable to text normalization, resulting in low reliability of hidden messages when the text is reprocessed. In addition, the implementation aspect of the system also remains an issue, as most studies have not provided format-based steganography applications that are easy to use, systematic, and replicable by general users. The process of message insertion and extraction is often not designed in a structured manner, resulting in varying and inconsistent levels of message extraction success. Based on these conditions, this research problem is explicitly formulated as the need for a format-based steganography method in text media that can improve the reliability of message insertion and extraction, maintain message concealment, and be easily implemented in application systems. This research aims to fill this gap by implementing format-based steganography using hidden Unicode characters in text media, accompanied by an evaluation of the success of the message insertion and extraction process. The contribution of this research lies in presenting a more practical and systematic implementation approach, which can serve as the basis for developing more reliable and applicable text steganography in the context of digital text communication-based information security.

3. Material and Method

This section describes the materials and methods used in the research to implement a format-based steganography system in text media. The description covers the types of data used, supporting software, and the stages of the procedure for inserting and extracting secret messages. The methods are arranged systematically so that the research process can be replicated and evaluated scientifically.

3.1 Supporting Materials and Devices

The main materials used in this study consist of two types of data, namely container text and secret messages. The container text is free text in [11] .txt format that functions as an insertion medium, while the secret message is text entered

 ISSN: XXXX-XXXX

by the user to be hidden in the container text. Text was chosen as the medium because it is the most common and lightweight form of digital communication used in everyday information exchange.

The software used to develop and run the steganography system is the Python programming language with tkinter library support to build a graphical user interface. The system is developed and run on Windows, Linux, or macOS operating systems without requiring special hardware or additional external libraries, making the system lightweight and easy to replicate.

3.2 Research Method

The method used in this research is format-based steganography by utilizing invisible Unicode characters (zero-width characters) as data bit representations. The basic principle of this method is to insert secret messages into host texts through format changes that do not affect the visual appearance, structure, or semantic meaning of the text. The research process began with the conversion of the secret message into numerical and binary representations. Each character of the secret message was converted to ASCII code, then converted to 8-bit binary form. The entire binary sequence of the message was combined into a single bit sequence. To ensure that the extraction process ran correctly, a 16-bit header was added to serve as a marker for the length of the secret message [12], [13].

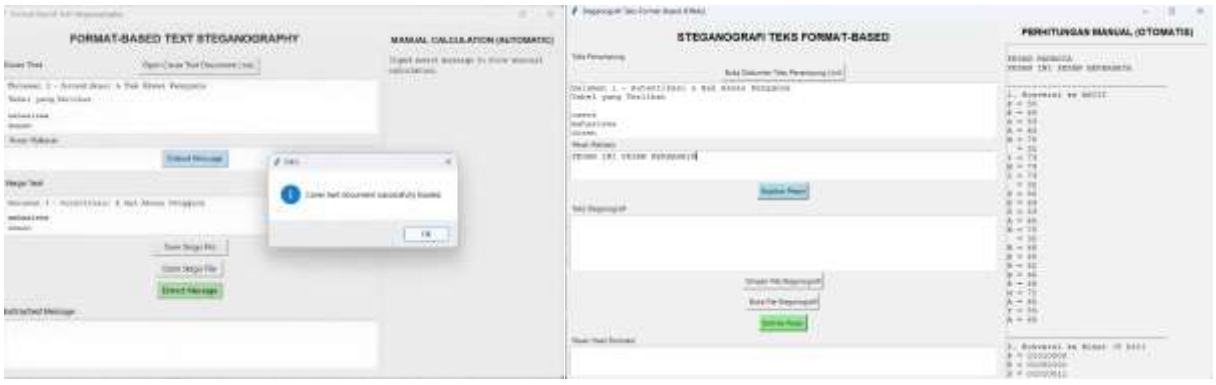
The next step is the message embedding process. The binary sequence consisting of the header and secret message is embedded into the host text using invisible Unicode characters, namely Zero Width Space (\u200b) to represent bit 0 and Zero Width Non-Joiner (\u200c) to represent bit 1. These characters are inserted at the end of the host text so that they do not change the visual appearance of the text. The result of this process is steganographic text that is visually identical to the original text but contains a secret message. The message extraction process is carried out by rereading the steganographic text file and detecting the invisible Unicode characters contained within it. Each Unicode character is converted back to its bit form according to its representation. The first sixteen bits are read as a header to determine the length of the secret message, then the next number of bits are processed as message data. These bits are regrouped into 8-bit units and converted to ASCII characters to obtain the secret message in its original text form. All stages of the insertion and extraction methods are designed in a structured manner so that they can be replicated and evaluated scientifically. This method emphasizes message concealment, text readability, and simplicity of implementation without compromising the reliability of the extraction process.

4 Research And Discussion

This section presents the results of implementing format-based steganography methods in text media and discusses the process of inserting and extracting secret messages. The analysis focuses on how the system works, the concealment of messages, and the system's ability to restore messages without changing the meaning or visual appearance of the host text.

4.1 Implementation of Text Steganography Application

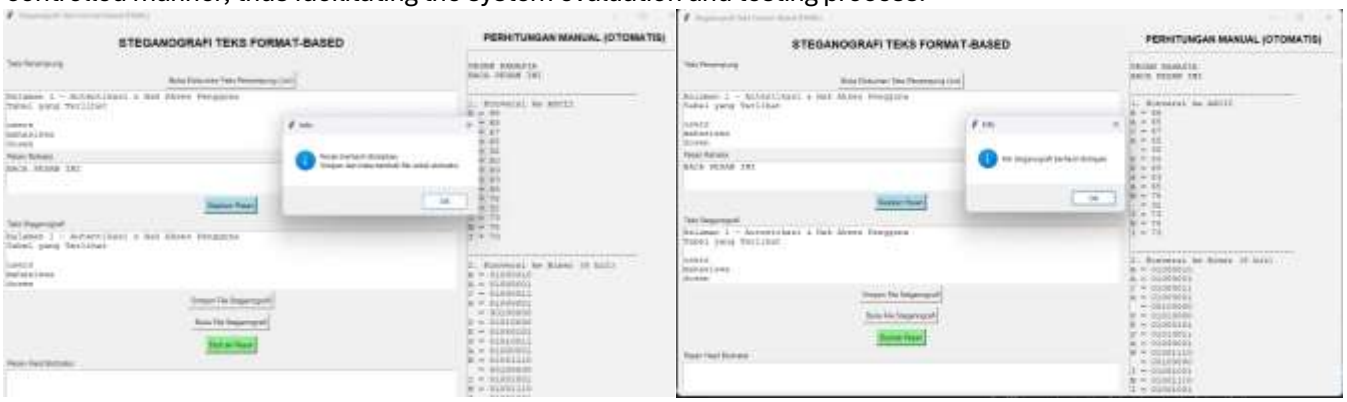
The application developed is a GUI-based desktop application designed to insert secret messages into text documents (.txt) using format-based steganography methods. This method utilizes invisible Unicode characters, namely Zero Width Space (U+200B) and Zero Width Non-Joiner (U+200C), to represent binary bits 0 and 1. The application interface consists of a text container area, secret message input, steganography text area, and message storage and extraction features. Additionally, there is a Manual (Automatic) Calculation panel that transparently displays the technical process of message insertion.



The main interface of a format-based text steganography application that provides features for loading container text, inserting secret messages, storing steganography files, and extracting messages.

4.2 Secret Message Insertion Process

The Manual Calculation Panel (Automatic) displays the technical stages of message insertion in detail, including character conversion to ASCII, ASCII conversion to 8-bit binary, message bit count calculation, header formation, and total bits inserted. This feature is educational and helps users understand the working mechanism of format-based text steganography transparently. The existence of this panel also shows that the insertion process is carried out in a structured and controlled manner, thus facilitating the system evaluation and testing process.

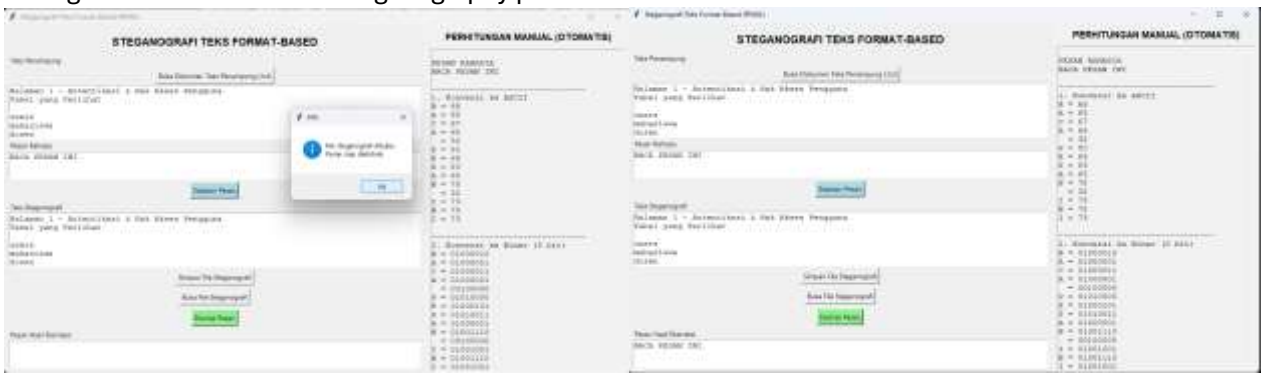


Penampung Figure 2. Process of Inserting Secret Messages into Host Texts

During the evaluation stage of the insertion process, the system shows that the secret message has been successfully embedded into the host text without visibly changing the text content. The process begins with converting the message into ASCII code, followed by transforming it into an 8-bit binary form as a digital data representation. This series of bits is then inserted into the text format structure using a format-based steganography method. The success of the process is indicated by a notification that the message has been successfully inserted and the file needs to be saved for the extraction stage. This shows that the encoding mechanism is running according to procedure, no data loss occurred during conversion, and the system is able to integrate the hidden message into the text medium completely and ready for testing at the extraction stage.

4.3 Steganography Process Evaluation

The steganography process evaluation is conducted to test the success of the system in embedding and extracting secret messages using format-based text steganography methods. This stage aims to ensure that the data conversion process, bit insertion, and message retrieval run according to the system design without causing visual changes to the host text. The evaluation is carried out by testing the embedding and extraction processes sequentially, as well as verifying the conformity of the message before and after the steganography process.



During the insertion stage, the user enters a secret message in the form of the text "READ THIS MESSAGE." The system then converts the message into ASCII code to convert each character into a numerical representation. Next, the ASCII value is converted back into 8-bit binary form to obtain a series of digital bits that represent the entire message. The conversion results show that the total number of message bits generated is 112 bits. The system then adds a 16-bit header that serves as a message length marker during the extraction process. Thus, the total number of bits inserted into the container text is 128 bits, consisting of the header and message content.

The insertion process is carried out using invisible special characters, namely Unicode U+200B (Zero Width Space) to represent bit 0 and U+200C (Zero Width Non-Joiner) to represent bit 1. These characters do not affect the visual appearance of the text, so the steganographic text remains the same as the original text. The bits are inserted at the end of the container text according to the system mechanism. Once the insertion process is complete, the system displays a notification that the steganographic file has been successfully created and is ready to be saved. The next step is the extraction process, where the saved file is reopened to read the hidden characters. The system then identifies the 16-bit header to determine the length of the message, reads the 112-bit message content, converts it from binary to ASCII, and finally returns it to its original text form. The evaluation shows that the extracted message is identical to the original message, so it can be concluded that the steganography process worked well. There was no change in the appearance of the host text, no data loss during conversion, and the system was able to maintain the integrity of the message in its entirety. This proves that the format-based steganography method using zero-width characters is effective for hiding messages in digital text media.

5 Conclusion

Based on the results of implementation and testing, the format-based text steganography method developed is capable of embedding secret messages into text media without changing the visual appearance, structure, or meaning of the host text through message conversion to binary form, addition of message length headers, and utilization of invisible Unicode characters as data bit representations. Test results show that messages can be embedded and extracted accurately according to the original message, making this method effective in maintaining the confidentiality of text-based information. However, format-based methods still have limitations with regard to changes in text format, such as copying, editing, or file conversion, which can remove hidden characters and cause extraction failure. Therefore, further research is recommended to develop techniques that are more resistant to format changes and to improve the capacity and reliability of message insertion.

6. References

[1] T. Ismail, K. Karo, A. Fadhil Herlambang, P. A. Nurfadhilla, and M. Fachrul, "Implementasi Metode LSB (Least Significant Bit) Pada Aplikasi Steganografi Citra Untuk Pengamanan Pesan Teks," 2026.

[2] L. Vinitha Sree, "Steganography Using Dft Based Texture Synthesis," INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, vol. 9, p. 2, 2020, [Online]. Available: www.ijstr.org

[3] D. Sebagai et al., "STEGANOGRAFI GAMBAR MENGGUNAKAN METODE LEAST SIGNIFICANT BIT PADA CITRA DENGAN OPERASI XOR TUGAS AKHIR."

[4] R. Munir, "07-Steganografi."
[5] T. Rabie, "Digital Image Steganography: An FFT Approach," in Communications in Computer and Information Science, 2012, pp. 217–230. doi: 10.1007/978-3-642-30567-2_18.
[6] S. Sundari, M. Vazira, M. Annisa Fitri, and U. Malikussaleh, "Desain dan Implementasi Aplikasi Steganografi Dual Mode Berbasis Java untuk Menyembunyikan Pesan Teks dan Gambar," 2026.
[7] S. Hutahaean, M. Malau, and G. Siboro, "Metode Steganografi EOF untuk Penyisipan Pesan Teks Tersembunyi," Jurnal Quancom, vol. 3, no. 1, pp. 18–24, 2025, doi: 10.62375/jqc.v3i1.434.
[8] A. H. Hasugian, I. Rusydi, and P. Apriani, "TEKNIK STEGANOGRAFI DISCRETE COSINE TRANSFORM DAN ALGORITMA RSA UNTUK MENYISIPKAN PESAN PADA AUDIO," 2024. [Online]. Available: http://ojsamik.amikmitragama.ac.id
[9] S. K. Zhao, S. Kamal, and A. Khalid, "DEEPhide: A Text Steganography Encoder and Decoder Tool using Deep Learning," vol. 6, no. 2, pp. 494–508, 2025, doi: 10.30880/aitcs.2025.06.02.027.
[10] V. Kozachok, A. Kozachok, S. Kopylov, and E. Pavlenko, "Steganographic System Model Based on Text Container," 2021.
[11] M. T. Zahiran Bin Zahari Dr Nazhatul Hafizah Kamarudin, "ZAHIRAN-SEMBUNYI PESANAN TEKS RAHSIA."
[12] J. Petrus, "Implementasi Steganografi pada Citra dengan Metode Bit-Plane Complexity Segmentation Untuk Transformasi Data".
[13] B. Cornelius, "Makalah IF4020 Kriptografi-Teknik Informatika ITB-Semester I Tahun," 2025.