УДК 004.63:004.056

РОЗРОБКА WINDOWS-АНАЛОГІЧНОГО ІНСТРУМЕНТУ ДЛЯ ВИЗНАЧЕННЯ ТИПУ ФАЙЛУ НА ОСНОВІ МАГІЧНИХ БАЙТІВ

Яковенко Д.О., Нестеренко Є.В
email: dmytro.iakovenko@nure.ua yevhen.nesterenko@nure.ua
email: arkadii.snihurov@nure.ua
Харківський національний університет радіоелектроніки
каф ІКІ ім. В.В. Поповського
м. Харків, Україна.

File type identification is a crucial task in cybersecurity and digital forensics. In Linux, the file command determines file types based on magic bytes, but Windows lacks a native equivalent. This research focuses on developing a Windows-based tool for identifying file types by analyzing magic bytes. The proposed solution extracts the first bytes of a file, compares them with a predefined database of known signatures, and determines the file type. This tool can assist in malware analysis, file integrity verification, and digital forensics. The implementation includes a command-line interface and potential integration into Windows Explorer.

Визначення типу файлу є критично важливим для багатьох завдань, зокрема для цифрової криміналістики, де точне розпізнавання типів файлів може допомогти виявити приховані або підроблені дані. Для цього використовуються Magic bytes. Це унікальна послідовність байтів на початку файлів, яка визначає його формат незалежно від розширення.

Відомо, що в операційній системі Linux для цієї мети використовується команда File, яка аналізує файли на основі їх магічних байтів [1].

Однією з основних проблем використання команди file є її відсутність у стандартному наборі утиліт Windows, що змушує дослідників та аналітиків перемикатися на Linux-середовище або використовувати емуляцію. Це створює додаткові складнощі для автоматизації процесів аналізу файлів у Windows-системах. Крім того, file у Linux має обмежену ефективність при визначенні бінарних файлів та файлів із нестандартною структурою, оскільки її метод аналізу базується переважно на магічних байтах [2]. Це може призводити до помилкової ідентифікації або повної відсутності результату для нових або модифікованих форматів. У сучасних умовах, коли постійно з’являються нові типи файлів і методи їхнього шифрування, така обмеженість стає критичною, особливо в сферах кібербезпеки, цифрової криміналістики та автоматизованої обробки даних.

Однак для користувачів операційної системи Windows аналогічних інструментів, які б проводили таку ідентифікацію, не так багато.

Метою вирішення даної проблеми було запропоновано розробити власну команду File під операційну систему Windows. Що забезпечує точніше визначення типів файлів без потреби у Linux-середовищі [3]. Це критично важливо для цифрової криміналістики та кібербезпеки.

Розроблений інструмент використовує заздалегідь підготовлену базу даних Magic bytes для підтримки різних форматів файлів. Порівняння магічних байтів з таблицею сигнатур реалізовано на базі мови програмування Python [4].

Порядок розробки інструменту для визначення типу файлу на основі магічних байтів у Windows:

- зчитує початкові байти файлу;
- порівнює ці байти з відомими сигнатурами файлів, збережених у таблиці сигнатур, або вставлені у сам код програми;
- виводить результат, визначаючи тип файлу, md5 hash.

Результат роботи продемонстровано на рисунку 1.

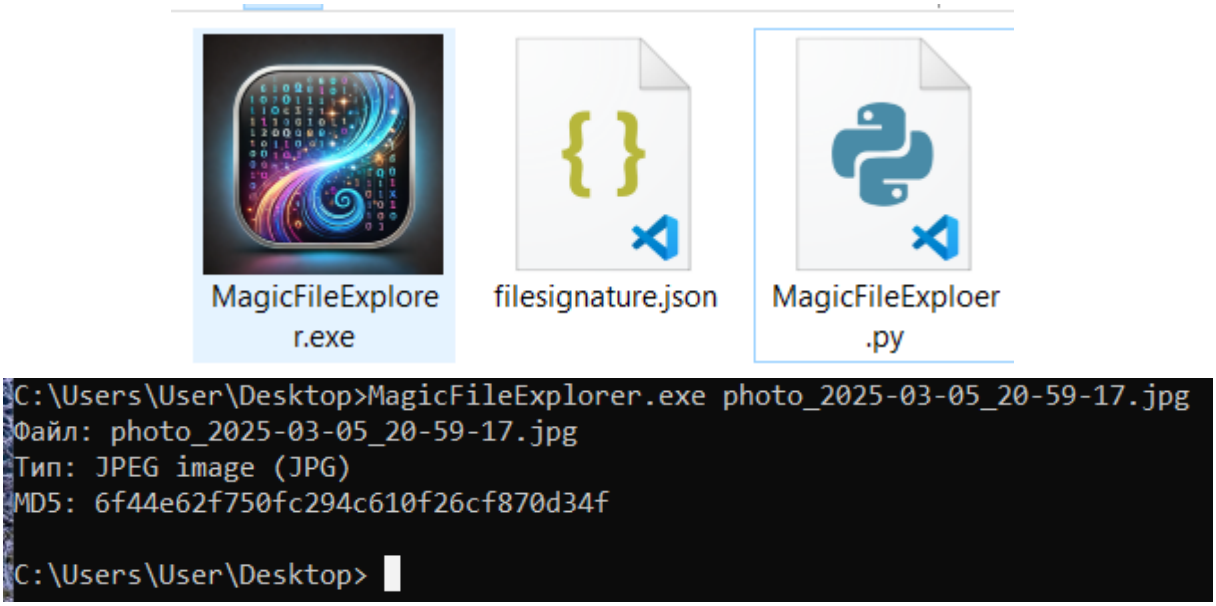

Рисунок 1 – Результат роботи програми

Команда file, що використовується у Linux, є стандартним інструментом для визначення типів файлів на основі їхнього вмісту. Вона аналізує магічні байти та мета-дані файлу, проте має певні обмеження. По-перше, її точність варіюється в межах 85-90%, оскільки методи аналізу базуються лише на початкових байтах файлу, що може призводити до хибних визначень, особливо у випадках архівів та складних форматів даних. По-друге, швидкість її роботи є відносно невисокою, що може бути критичним при масовому аналізі файлів

На відміну від file, запропонована утиліта для Windows забезпечує точніше та швидше визначення типів файлів завдяки використанню не лише магічних байтів, але й додаткового аналізу, наприклад, хешування MD5 та зіставлення з розширеною базою даних форматів. Це дозволяє підвищити точність до понад 95% та забезпечити підтримку більшої кількості бінарних і нестандартних форматів.

Окрім цього, нова утиліта підтримує динамічне оновлення бази сигнатур, що дозволяє легко додавати нові формати без необхідності внесення змін у вихідний код.

Список використаних джерел:

1. IBM Documentation. File command [Електронний ресурс]. URL: https://www.ibm.com/docs/fr/aix/7.1?topic=f-file-command (дата звернення 01.03.2025).

2. Wikipedia. List of file signatures [Електронний ресурс]. URL: https://en.wikipedia.org/wiki/List_of_file_signatures (дата звернення 29.02.2025).

3. Гонтарь, І. А., and А. В. Снігуров. "Методика проведення системного аудиту на Linux серверах." (2023).

4. Python Software Foundation. Python Downloads [Електронний ресурс]. URL: https://www.python.org/downloads/ (дата звернення 27.02.2025).