

ВИКОРИСТАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ СИСТЕМИ АНАЛІЗУ ФАЙЛІВ НА ДИСКУ

Вінницький національний технічний університет

Анотація. У статті розглядається проблема неструктурованості, дублювання та надмірного накопичення файлів на цифрових носіях, що ускладнює ефективне управління інформаційними ресурсами. Проаналізовано принципи роботи програмного забезпечення для аналізу файлових систем, включаючи індексацію, хешування, класифікацію та візуалізацію результатів. Описано технологічні засади таких систем, зокрема використання багатопотокових алгоритмів, криптографічних хеш-функцій, елементів машинного навчання та інтеграцію з базами даних. Розглянуто практичні галузі застосування: корпоративний сектор, освіта, охорона здоров'я, державне управління та побут. Автори підкреслюють, що сучасні аналітичні інструменти для обробки файлів є ключовими для підвищення продуктивності, інформаційної безпеки та оптимізації ІТ-інфраструктури.

Ключові слова: аналіз файлів, файлові системи, хешування, дублікати, візуалізація даних, оптимізація дискового простору, інформаційна безпека.

Abstract. This article addresses the issue of file duplication, unstructured storage, and excessive data accumulation on digital media, which hinder efficient information management. It examines the operating principles of file analysis software, including indexing, hashing, classification, and data visualization. The technological foundation of such systems is discussed, covering multithreaded algorithms, cryptographic hash functions, machine learning elements, and database integration. The article explores practical application areas such as corporate environments, education, healthcare, public administration, and home use. The authors emphasize that modern analytical tools for file processing are essential for increasing productivity, ensuring data security, and optimizing IT infrastructure.

Keywords: file analysis, file systems, hashing, duplicates, data visualization, disk space optimization, information security.

Вступ

Проблема накопичення, дублювання та неструктурованості файлів на цифрових носіях є серйозною перешкодою для ефективного використання інформаційних ресурсів. За даними аналітичних досліджень, понад 30% простору на корпоративних та особистих жорстких дисках зайнято файлами, які не використовуються або дублюються [1]. Це не лише уповільнює роботу систем, а й ускладнює пошук необхідної інформації, підвищує витрати на зберігання та ускладнює забезпечення кібербезпеки. Традиційно для оптимізації файлової структури застосовувалися базові вбудовані засоби операційних систем, однак їх можливостей недостатньо для глибокого аналізу великих обсягів даних [2].

Сучасний підхід орієнтований на використання спеціалізованого програмного забезпечення, здатного автоматично сканувати файлові системи, класифікувати дані, виявляти дублікати та створювати звіти щодо стану дискового простору. Швидкий розвиток технологій аналізу даних і візуалізації сприяє впровадженню таких систем у бізнес-середовище, ІТ-інфраструктури та навчальні заклади. Особливої актуальності ці інструменти набувають в умовах віддаленої роботи, коли централізований контроль за даними є критично важливим.

Отже, аналіз файлових структур за допомогою інтелектуальних програмних засобів перетворюється на необхідний компонент сучасного управління цифровими ресурсами.

Принципи роботи аналізу файлів на диску

Програмне забезпечення для аналізу файлів на диску ґрунтується на комплексному підході до вивчення структури файлової системи та логіки використання дискового простору. Основним етапом є глибоке сканування жорсткого диска або інших носіїв даних, під час якого утворюється індекс файлів – спеціальна база даних, що містить інформацію про кожен об'єкт файлової системи: його назву, тип, розмір, дату створення та останнього доступу, місце розташування, а також права доступу. Така деталізація дозволяє програмі формувати повне уявлення про структуру збережених даних [3].

Після цього система переходить до аналітичної обробки зібраної інформації, застосовуючи алгоритми хешування та порівняння вмісту файлів. Завдяки цьому вона здатна виявляти дублікати,

великі або застарілі файли, які часто займають значний обсяг пам'яті без практичної користі. Сучасні рішення можуть не лише виявляти ідентичні об'єкти за контрольними сумами, а й знаходити схожі файли за вмістом або структурою. Це особливо корисно в умовах великих даних, де ручне виявлення надлишкової інформації є неможливим.

Після аналітичного етапу користувач отримує результати у зручному для сприйняття форматі. Більшість програм оснащені візуальними засобами – від кругових діаграм до мап використання простору, які дозволяють наочно побачити, які каталоги чи типи файлів займають найбільше місця [4]. Інтерфейс дозволяє фільтрувати та сортувати дані за різними критеріями, що значно пришвидшує прийняття рішень щодо подальших дій.

Окремі програмні рішення також інтегруються з корпоративними політиками управління файлами, підтримують хмарне збереження, автоматичне архівування або видалення застарілих документів. Таким чином, програмне забезпечення перетворюється з пасивного засобу аналізу на активний інструмент оптимізації цифрового середовища.

Усі ці механізми працюють у комплексі, забезпечуючи користувача ефективними засобами контролю над файловою системою, мінімізацією надлишкових даних і підвищенням загальної продуктивності роботи інформаційної інфраструктури.

Технологічне забезпечення програмного аналізу файлів

Функціонування сучасного програмного забезпечення для аналізу файлів на диску базується на поєднанні кількох важливих технологій, що забезпечують глибоку, точну та швидку обробку великих обсягів даних. Однією з ключових є система індексації та сканування, що реалізується за допомогою багатопотокових алгоритмів, здатних обробляти десятки тисяч файлів одночасно, мінімізуючи навантаження на систему користувача. Такі алгоритми використовують як рекурсивне сканування директорій, так і побудову графів файлових зв'язків, що дозволяє аналізувати структуру збереження у просторі [5].

Для ідентифікації дублікатів і класифікації даних активно застосовуються криптографічні хеш-функції (наприклад, MD5, SHA-1, SHA-256), що забезпечують унікальний цифровий відбиток кожного файлу. У поєднанні з алгоритмами машинного навчання це дає змогу не лише виявляти ідентичні файли, але й класифікувати їх за типом вмісту, навіть якщо вони мають різні назви або формати [6]. Застосування кластерного аналізу дозволяє виявляти групи схожих даних та оптимізувати простір, наприклад, об'єднуючи однотипні архіви або великі медіафайли.

Візуалізація результатів аналізу, як правило, реалізується на основі бібліотек для побудови графіки в реальному часі, таких як D3.js або Chart.js, які інтегруються з графічним інтерфейсом користувача [7]. Це дає змогу динамічно оновлювати аналітичні дані та забезпечити інтерактивну взаємодію з файлами. Користувач може не лише переглядати статистику, а й миттєво виконувати дії — переміщення, видалення або архівування — безпосередньо з панелі керування.

У більш складних рішеннях використовуються бази даних типу SQLite або PostgreSQL для збереження результатів попередніх аналізів, що дозволяє порівнювати зміни в структурі диска з часом. Водночас підтримка командного рядка та скриптів (наприклад, PowerShell або Bash) забезпечує можливості для автоматизації процесів у великих IT-середовищах.

Розвиток технологій штучного інтелекту також відкриває нові горизонти для аналізу файлів: наприклад, розпізнавання зображень, вмісту PDF або обробка метаданих документів у фоновому режимі. У результаті програмні рішення стають не лише інструментами аналізу, але й інтелектуальними асистентами з управління цифровими ресурсами.

Галузі застосування програмного аналізу файлів

Програмне забезпечення для аналізу файлів на диску знаходить широке застосування в різноманітних сферах, де зберігання та ефективне управління великими обсягами цифрових даних є критично важливими. Насамперед такі рішення активно використовуються в корпоративному секторі, де компанії оперують терабайтами інформації, розміщеної як на локальних серверах, так і в хмарних середовищах [8]. У цьому контексті системи аналізу допомагають виявляти дублікати, оптимізувати структуру каталогів, контролювати доступ до конфіденційних даних і скорочувати витрати на зберігання.

Важливе місце займає застосування таких технологій в освітніх установах і наукових центрах. Тут вони використовуються для впорядкування архівів студентських робіт, методичних матеріалів,

результатів досліджень, а також резервного копіювання важливої інформації. Аналіз файлової системи дозволяє швидко визначати застарілі або непотрібні файли, що значно полегшує адміністрування навчальних серверів.

У сфері охорони здоров'я системи аналізу файлів забезпечують оптимізацію медичних архівів, зокрема цифрових зображень (рентген, МРТ), історій хвороб та звітів лабораторних досліджень. Це не лише сприяє збереженню чутливої інформації, але й підвищує швидкість доступу до неї, що критично в умовах оперативної медичної допомоги.

У державному секторі такі програми застосовуються для забезпечення інформаційної безпеки, контролю за дотриманням політик зберігання документів та виявлення несанкціонованих файлів. Крім того, системи аналізу використовуються в юридичній практиці для аудиту цифрових носіїв та збору електронних доказів.

У побутовому використанні подібне програмне забезпечення дозволяє звичайним користувачам керувати простором на особистих комп'ютерах, виявляти зайві файли, очищати кеш, проводити дефрагментацію та організовувати резервні копії.

Отже, гнучкість і функціональність сучасних систем аналізу файлів забезпечує їх актуальність у різних галузях – від освіти і медицини до бізнесу та державного управління.

Висновки

У сучасних умовах цифрового перевантаження ефективно управління файловими системами є важливим чинником підвищення продуктивності, інформаційної безпеки та раціонального використання ресурсів. Програмне забезпечення для аналізу файлів на диску виконує не лише функцію очищення або сортування даних – воно стає стратегічним інструментом у процесах оптимізації IT-інфраструктури, захисту інформації, підтримки впорядкованості архівів та дотримання політик зберігання.

Застосування таких систем демонструє високу ефективність як у корпоративному, так і в навчальному та побутовому середовищі. Завдяки інтеграції сучасних технологій – алгоритмів хешування, візуалізації, автоматизації та навіть штучного інтелекту – користувачі отримують зручні, гнучкі та точні інструменти для контролю над власними цифровими ресурсами.

Таким чином, розвиток і впровадження програмних систем аналізу файлів є не лише актуальним, а й необхідним етапом на шляху до безпечного, впорядкованого та ефективного інформаційного середовища.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Statista Research Department. "Share of unused and duplicate files on corporate storage." Statista, 2023.
2. Microsoft Docs. "Manage duplicate files and optimize storage using Windows tools." docs.microsoft.com, 2022.
3. TreeSize by JAM Software. "How TreeSize analyzes disk space." jam-software.com, 2023.
4. WinDirStat Project. "Visualization and statistics of disk usage." windirstat.net, 2021.
5. White, T. "Hadoop: The Definitive Guide." O'Reilly Media, 2021.
6. IBM Research. "File system optimization using AI and hashing algorithms." IBM White Papers, 2022.
7. D3.js Official Documentation. "Interactive Data-Driven Documents." d3js.org, 2023.
8. Gartner. "Use Cases for File Analysis Tools in Different Sectors." Gartner Insights, 2023.

Назарчук Богдан Юрійович – студент групи 4ПІ-216, ФІТКІ, Вінницький національний технічний університет, Вінниця, e-mail: bogdannazarchuk7@gmail.com

Черноволик Галина Олександрівна – к.т.н., доцент, завідувач кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця, e-mail: lina2433@gmail.com @gmail.com.

Nazarchuk Bohdan Yuriiovych - student of group 4PI-21b, FITCI, Vinnytsia National Technical University, Vinnytsia, e-mail: bogdannazarchuk7@gmail.com

Chernovolyk Halyna Alexandrovna - PhD in Engineering, Associate Professor, Head of the Department of Software, Vinnytsia National Technical University, Vinnytsia, e-mail: lina2433@gmail.com @gmail.com.