

Вашенко А.В.

студент групи ІКСМ,

Херсонський національний технічний університет, м. Херсон.

Дроздова С. А.

старший викладач кафедри інформаційних технологій,

Херсонський національний технічний університет, м. Херсон.

ВИРІШЕННЯ ПРОБЛЕМИ ЗБЕРІГАННЯ ДАНИХ ЗА ДОПОМОГОЮ ПРОГРАМИ СТИСНЕННЯ ФАЙЛІВ

Багато сфер людської діяльності потребують комп'ютеризації. Вона, у свою чергу, передбачає використання, зберігання та обробку значних обсягів інформації, які можуть займати велику частину простору накопичувачів (жорстких дисків, знімних пристроїв тощо).

Поширення криптовалют, які використовують накопичувачі, призвело до збільшення вартості за одиницю об'єму пам'яті.

На рисунку 1 показаний графік, який відображає динаміку цін на прикладі SSD накопичувача Intel 660p 2 TB (SSDPEKNW020T8X1) [1]

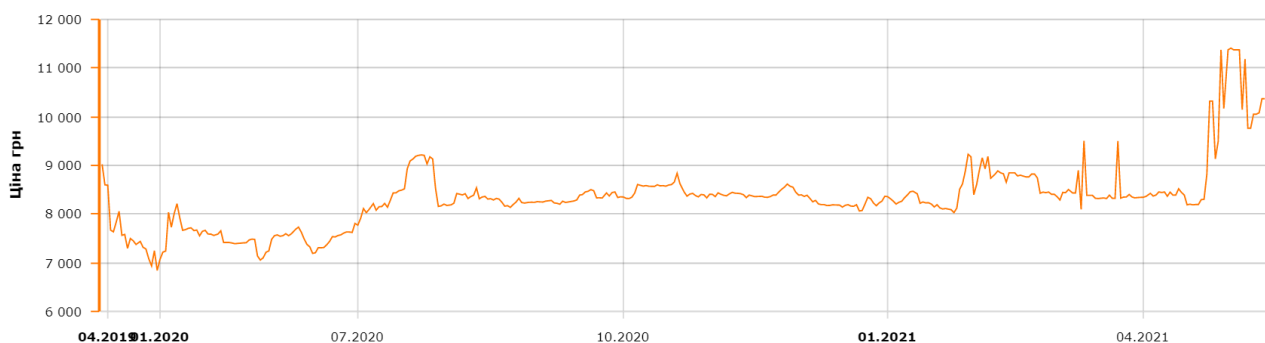


Рис. 1. Динаміка цін накопичувача SSD Intel 660p 2 TB

Вирішенням проблеми зберігання інформація є компресія даних.

Стиснення даних – це процедура перекодування, яка проводиться з метою зменшення займаного ними об'єму. Вона застосовується для більш раціонального використання пристроїв передачі та зберігання даних. Застосування компресії дозволяє раціонально використовувати ресурси пристроїв для зберігання даних.

Стиснення базується на усуненні надлишку інформації, яка міститься у вхідних даних (наприклад, повторення в тексті фрагментів). Подібний надлишок зазвичай усувається заміною повторюваних послідовностей коротшим значенням (кодом). Ще один вид надлишковості пов'язаний з тим, що деякі значення в даних, що стискаються, зустрічаються частіше за інші. В цьому випадку можна замінювати дані, що трапляються часто, коротшими кодами, а ті, що зустрічаються рідко, довгими (ймовірніше стиснення). Стиснення даних, які не мають властивості надлишку, неможливе. З цієї причини стиснення зашифрованих або закодованих даних може бути неефективним або неможливим.

Стиснення може бути виконане з втратами або без втрат. Використання методів стиснення без втрат надає можливість відновити дані при декомпресії з точністю до біта. Стиснення без втрат можливе лише тому, що більшість невідповідних даних мають властивість надлишку. Стиснення без втрат використовується у випадках, коли важливо, щоб відновлені дані були ідентичні оригіналу. Стиснення з втратами – метод стиснення даних, при якому розпакований

файл відрізняється від оригіналу, проте може бути використаний. Часто методи стиснення з втратами включають у себе методи стиснення без втрат. Методи стиснення з втратами найчастіше використовується для стискання мультимедіа-даних.

Для вирішення поставленої проблеми було створене програмне забезпечення «Huffman Data Compression». Для стиснення даних програма використовує алгоритм Хаффмана - алгоритм оптимального префіксного кодування алфавіту з мінімальною надлишковістю [2].

Ідея алгоритму полягає у тому, що знаючи ймовірності появи символів у повідомленні, можна описати процедуру побудови кодів змінної довжини, що складаються з цілої кількості бітів. Символам з більшою ймовірністю ставляться у відповідність коротші коди. Коди Хаффмана володіють властивістю префіксності (тобто жодне кодове слово не є префіксом іншого), що дозволяє однозначно їх декодувати [2].

Даний продукт завдяки простоті використання не накладає обмежень на його поширення серед фахівців напрямів, не пов'язаних з вивченням комп'ютерної техніки. Програма розрахована на одного користувача, не передбачає одночасну роботу багатьох користувачів. Стандартний користувацький інтерфейс, а також відносна простота полегшують її освоєння.

Головне вікно програми відображено на рисунку 2.

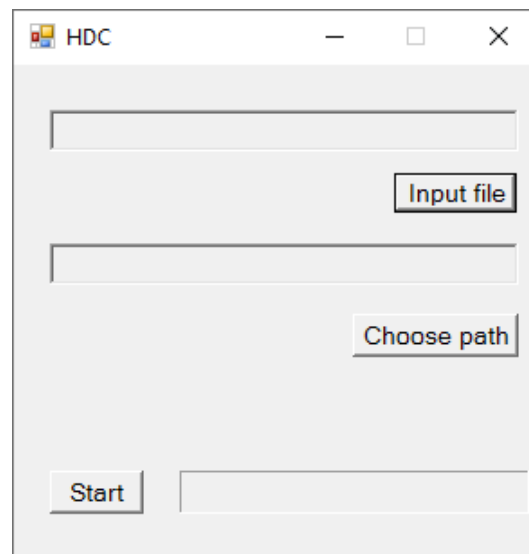


Рис. 2. Головне вікно програми «Huffman Data Compression»

- За допомогою кнопки меню «Input file» обирається файл з файлової системи.
- За допомогою кнопки меню «Choose path» здійснюється необов'язковий вибір папки, в якій буде знаходитися вихідний файл. За замовчуванням буде обрана папка, в якій знаходиться вхідний файл.
- Кнопка «Start» відповідає за обробку файлу. Програма, в залежності від формату файлу, сама обирає як обробляти файл.
- З правого боку від кнопки «Start» знаходиться шкала прогресу.

Для тестування програми було використано файли, що містять класичну літературу, яка розповсюджується безкоштовно, в форматі .txt. За результатами тестувань було виявлено, що середній коефіцієнт стиснення дорівнює 1,938. Програма може стикати дані, які закодовані з використанням більшості розповсюджених текстових кодувань (UTF-8, Unicode, Windows-1251 тощо). Стиснення даних, які є закодованими або зашифрованими, за допомогою програми може виявитись неефективним або неможливим.

Програма написана мовою програмування C++/CLI в інтегрованому середовищі розробки Microsoft Visual Studio Community 2019.

В ході подальший розробок планується додати можливість архівування папок.

Висновки. Використання програм для компресії даних дозволяє раціонально використовувати ресурси накопичувачів. Створене програмне забезпечення допомагає вирішити проблему нестачі дискового простору.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. hotline.ua. *Intel 660p 2 TB (SSDPEKNW020T8X1)*. URL: <https://hotline.ua/computer-diski-ssd/intel-660p-2-tb-ssdpeknw020t8x1/> (date of access: 14.05.2021).
2. Huffman D. A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the IRE*. 1952. Vol. 40, no. 9. P. 1098–1101. URL: <https://doi.org/10.1109/jrproc.1952.273898> (date of access: 14.05.2021).