

ТЕРНОПІЛЬСЬКИЙ ДЕРЖАВНИЙ ТЕХНІЧНИЙ  
УНІВЕРСИТЕТ ім. І. ПУЛЮЯ

# АРХІВАЦІЯ

The background of the slide is a grayscale image of a landscape. In the foreground, there are several utility poles with power lines stretching across the scene. The sky is a uniform, light gray, and the ground is dark and hazy. The overall mood is industrial and somewhat somber.

Павло Древницький

# Що таке архівація даних ?



Під архівацією розуміють стиснення даних. Стиснення даних — це процедура перекодування даних, яка проводиться з метою зменшення їх об'єму, розміру, обсягу. Стиснення буває без втрат (коли можливе відновлення вихідних даних без спотворень) або з втратами (відновлення можливе з незначними спотвореннями).

Архів — файл, що містить у собі один або декілька файлів та метадані. Файли можуть бути як стиснені (без втрат), так і мати початковий розмір та структуру, але першочергове завдання архіву тримати у собі саме стиснуті файли. Метедані можуть містити інформацію про початковий розмір файлів, інформацію про формат файлів, структуру директорій, коментарі до файлів, інформацію для відновлення архіву і т. д. Архіви файлів створюються за допомогою спеціалізованих програм — архіваторів, які можуть бути як окремими програмами, так і частиною інших програм.



# Алгоритми стиснення даних

## □ Стиснення інформації способом кодування серій (RLE)

Найбільш відомий, простий підхід і алгоритм стиснення інформації оборотним шляхом - це кодування серій послідовностей (Run Length Encoding - RLE). Суть методів даного підходу полягає в заміні ланцюжків або серій байтів що повторюються або їх послідовностей на один лічильник, що кодує байт і числа їх повторень.

## □ Арифметичне кодування

Арифметичне кодування є методом, що дозволяє стиснути символи вхідного алфавіту без втрат за умови, що відомий розподіл частот цих символів і є найбільш оптимальним, тому що досягається теоретична межа ступені стиснення. Передбачувана необхідна послідовність символів, при стисненні методом арифметичного кодування розглядається як деякий двійковий дріб з інтервалу  $[0, 1)$ . Результат стиснення записується як послідовність двійкових цифр із запису цього дробу.

## □ Алгоритм Лемпеля-Зіва-Велча (Lempel-Ziv-Welch - LZW)

Записується в перші 256 слотів рядки, що складаються з одного символу, номер якого дорівнює номеру слоту. Алгоритм переглядає вхідний потік, розбиваючи його на підстроки і додаючи нові слоти у кінець словника. Прочитаємо кілька символів у рядок  $s$  і знайдемо в словнику рядок  $t$  - самий довгий префікс  $s$ . Нехай він знайдений у слоті з номером  $n$ . Виведемо число  $n$  у вихідний потік, перемістимо покажчик вхідного потоку на  $\text{length}(t)$  символів уперед і додамо в словник новий слот, що містить рядок  $t+c$ , де  $c$  - черговий символ на вході (відразу після  $t$ ). Алгоритм перетворить потік символів на вході в потік індексів слотів словника на виході.



# Алгоритми стиснення даних

## □ **Двоступінчасте кодування. Алгоритм Лемпеля-Зіва (LZ-compression)**

Пакувальник постійно зберігає деяку кількість останніх оброблених символів у буфері. Під час обробки вхідного потоку символи, що знову поступили у потік, попадають в кінець буфера, зсуваючи попередні символи і витісняючи самі старі. Розміри цього буфера (sliding dictionary) варіюються в різних реалізаціях кодуючих систем. Потім, після побудови хеш таблиць, алгоритм виділяє (шляхом пошуку у словнику) саму довгу початкову підстроку вхідного потоку, що збігається з однією з підстрок у словнику, і видає на вихід пари (length, distance), де length - довжина знайденої в словнику підстроки, а distance - відстань від неї до вхідної підстроки (тобто фактично індекс підстроки у буфері, відмінусований від його розміру). У випадку, якщо така підстрока не знайдена, у вихідний потік просто копіюється черговий символ вхідного потоку.

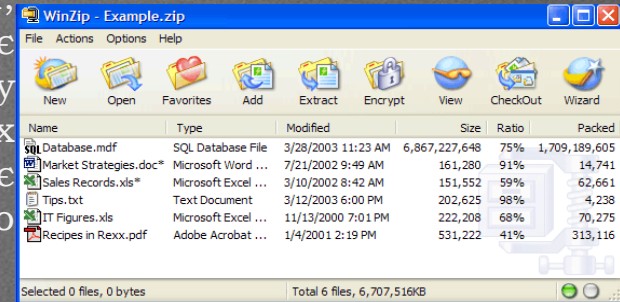
## □ **Стиснення даних кодом Хафмана**

Стискаючи файл по алгоритму Хафмана перше що треба зробити - це прочитати файл повністю і підрахувати скільки раз зустрічається кожний символ з розширеного набору ASCII. Якщо будуть враховуватися всі 256 символів, то не буде різниці в стисканні текстового чи EXE файлу. Після підрахунку частоти входження кожного символу, необхідно переглянути таблицю кодів ASCII і сформувати бінарне дерево.

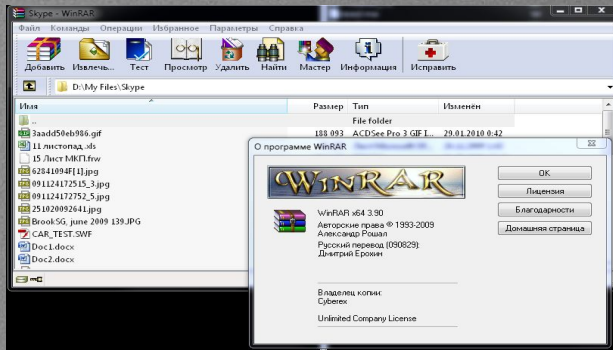


# Огляд архіваторів

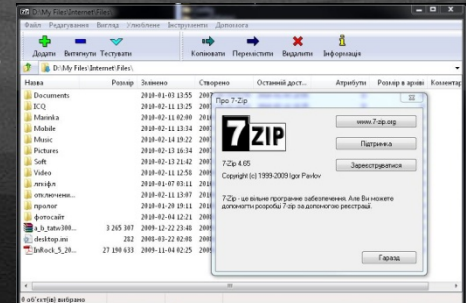
Формат ZIP був спочатку створений Філом Кацем, засновником компанії PKWARE. На даний момент існує безліч алгоритмів компресії, які виграють у ZIP і у швидкості, і у компресії, і у кількості додаткових можливостей. Незважаючи на це, ZIP надалі є популярним методом стиснення даних, резервного копіювання й обміну даними.



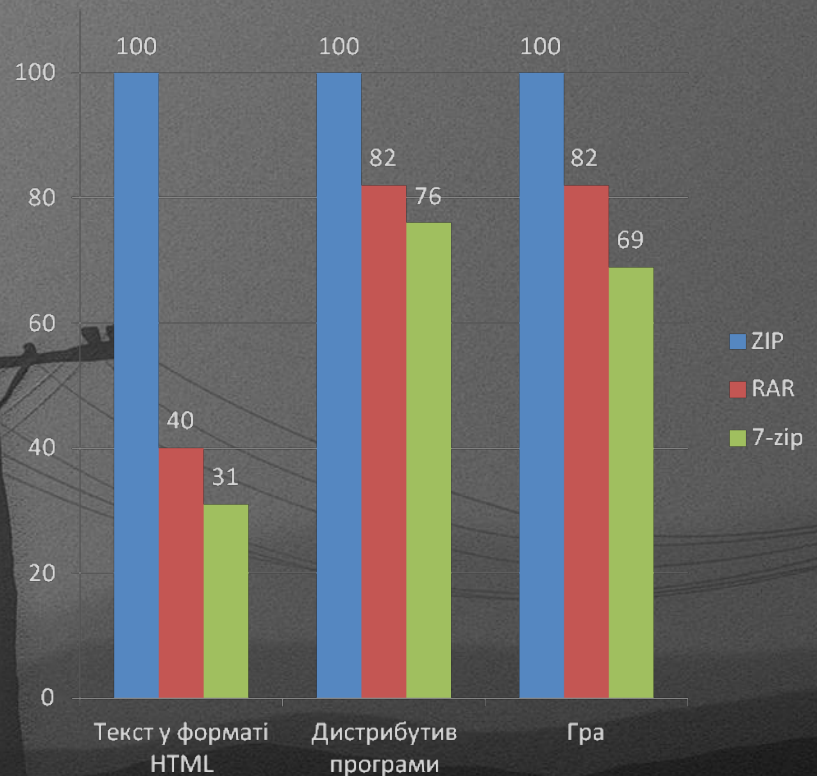
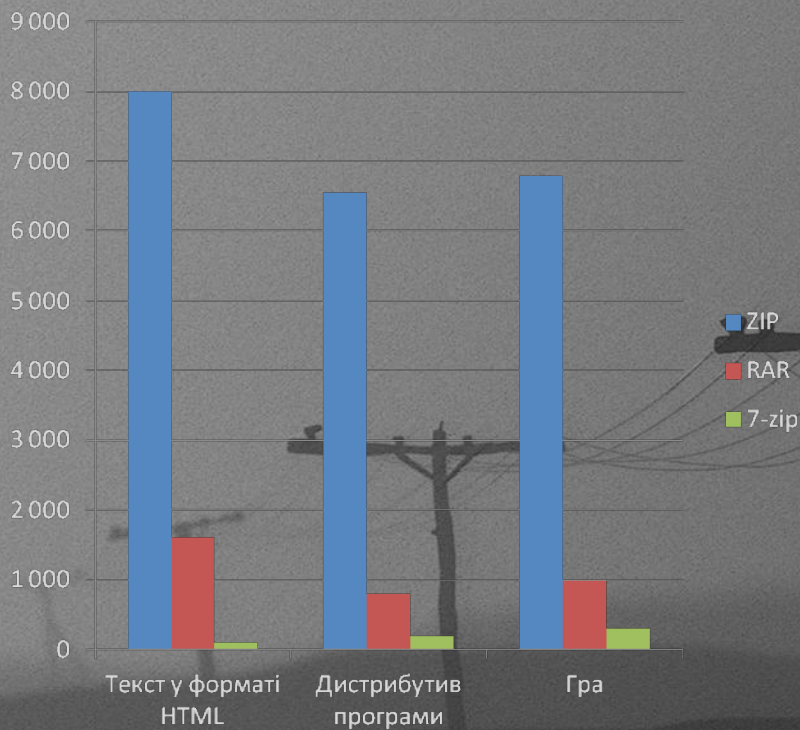
Формат RAR розроблений російським програмістом Євгенієм Рошалом (звідси і назва RAR: Roshal Archiver). Операції компресії із форматом RAR повільніші, ніж за використання ранніх алгоритмів стиснення, таких як ZIP і gzip, але зазвичай забезпечують вищу ступінь компресії.



7-zip — Розроблений Ігорем Павловичем файловий архіватор з високим ступенем стиснення. У більшості випадків ступінь стиснення вище, чим у RAR, за винятком деяких мультимедіа-даних. Швидкість стиснення при цьому нижча чим у конкурентів (як правило, не більше ніж на 30%).



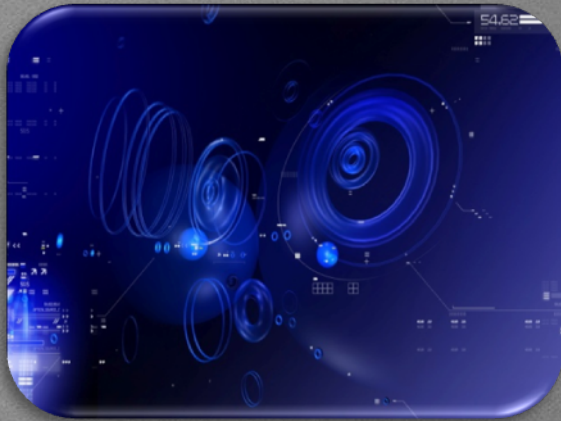
# Порівняння архіваторів



Середня швидкість стиснення (КБ/с)      Ступінь стиснення  
(процент по відношенню до ZIP)



# Стиснення аудіо-файлів без втрати інформації



Стиснення звукових даних — тип стиснення даних, кодування, що застосовується для зменшення обсягу аудіофайлів або заради можливості зменшення смуги пропускання для потокового аудіо. Алгоритми стиснення звукових файлів реалізуються у комп'ютерних програмах, що називаються аудіокодеками. Розробка спеціальних алгоритмів стиснення звукових даних мотивована тим, що загальні алгоритми стиснення неефективні для роботи зі звуком і унеможливають роботу в реальному часі.

Складність стиснення звуку без втрат полягає в тому, що записи звуку є надзвичайно складними у своїй структурі. Одним з стандартних методів стиснення - пошук взірців і їх повторень - не ефективний для більш хаотичних даних, якими є оцифрований звук.

FLAC — популярний свобідний кодек для стиснення аудіо-даних. На відміну від кодеків які працюють з втратами, наприклад MP3, FLAC не видаляє ніякої інформації з аудіопотоку і підходить як для прослуховування музики на високоякісній звуковідтворюючій апаратурі, так і для архівування аудіоколекції. FLAC ділить вхідний потік на блоки і кодує їх незалежно один від одного. Блок пакується у фрейм і додається до потоку. Базовий кодер



# Стиснення аудіо-файлів з втратою інформації



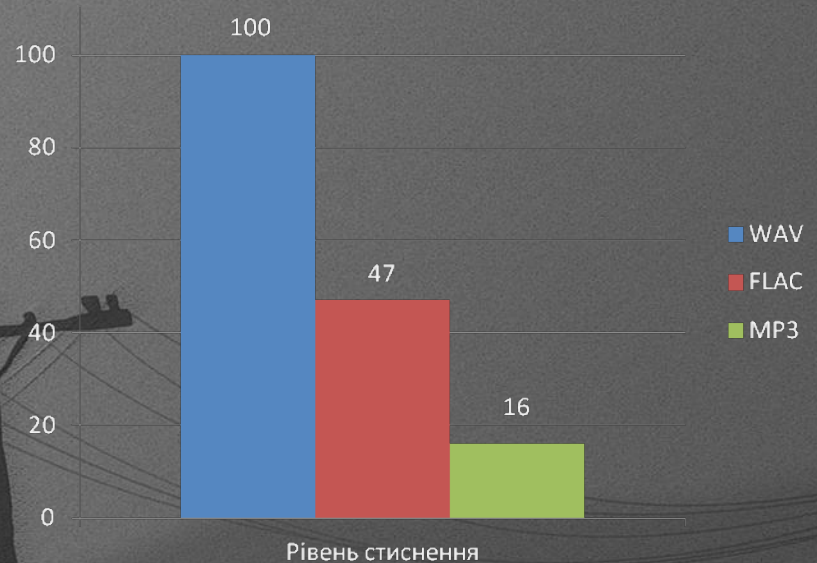
Стиснення з втратами має надзвичайно широке застосування. Окрім комп'ютерних програм, стиснення з втратами використовується в потоковому аудіо, в DVD, цифровому телебаченні і радіо та потоковому медіа в інтернеті. Новацією цього методу стиснення було використання психоакустики для виявлення компонентів звучання, що не сприймаються слухом людини. Прикладом можуть слугувати або високі частоти, які сприймаються лише при достатній їх потужності, або тихі звуки, що виникають одночасно або одразу після голосніших звуків і тому маскуються ними — такі компоненти звучання можуть бути передані менш точно, або і взагалі не передані. На даний момент MP3 є найвідомішим і популярним з розповсюджених форматів цифрового кодування аудіоданих. Він широко використовується у мережах для передачі музичних фрагментів. Формат підтримується практично всіма портативними аудіо-плеєрами, домашніми кінотеатрами, музичними центрами і різноманітним комп'ютерним програмним забезпеченням. Після застосування вищезгаданих методів стиснення аудіо з втратами (розподіл смуги звукових частот на підполоси, використання психоакустичної моделі та сполученого стерео)



# Порівняння аудіо-форматів



Швидкість стиснення (в секундах)



Рівень стиснення аудіо-треку (процент по відношенню до WAV)



# Список використаних джерел

1. [http://uk.wikipedia.org/wiki/Стиснення\\_даних](http://uk.wikipedia.org/wiki/Стиснення_даних)
2. [http://uk.wikipedia.org/wiki/%D0%90%D1%80%D1%85%D1%96%D0%B2\\_%28%D1%96%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29](http://uk.wikipedia.org/wiki/%D0%90%D1%80%D1%85%D1%96%D0%B2_%28%D1%96%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%29)
3. <http://structur.h1.ru/arch.htm>
4. <http://ru.wikipedia.org/wiki/ZIP>
5. <http://uk.wikipedia.org/wiki/Rar>
6. <http://ru.wikipedia.org/wiki/7-Zip>
7. <http://ru.wikipedia.org/wiki/FLAC>
8. [http://uk.wikipedia.org/wiki/%D0%A1%D1%82%D0%B8%D1%81%D0%BD%D0%B5%D0%BD%D0%BD%D1%8F\\_%D0%B0%D1%83%D0%B4%D1%96%D0%BE](http://uk.wikipedia.org/wiki/%D0%A1%D1%82%D0%B8%D1%81%D0%BD%D0%B5%D0%BD%D0%BD%D1%8F_%D0%B0%D1%83%D0%B4%D1%96%D0%BE)
9. <http://www.n-audio.com/articles/format-2.htm>



Дякую за увагу

Дякую за увагу

