

## ДОДАТОК А

Перелік джерел посилання за науковими напрямками керівника та науковців  
кафедри програмної інженерії

3. Byzkrovnyi, O., Chupryna, A., Smelyakov, K., Sharonova, N., Repikhov, V. Comparison of Object Detection Algorithms for the Task of Detecting Possible Road Accident. CEUR Workshop Proceedings, 2023, 3387, pp. 13–28

19. Metrics applicable for evaluating software at the design stage / I. Gruzdo та ін. *5th International Conference on Computational Linguistics and Intelligent Systems. Volume I: Main Conference*, м. Lviv, 22–23 квіт. 2021 р. 2021. С. 916–936.

URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0->

[85107237353&partnerID=40&md5=2acca556132c7c02af11bc95e7b5d4ec](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85107237353&partnerID=40&md5=2acca556132c7c02af11bc95e7b5d4ec)

## ДОДАТОК Б

Апробація результатів роботи

### ДОСЛІДЖЕННЯ АЛГОРИТМІВ ОПТИМАЛЬНОГО КОДУВАННЯ І СТИСНЕННЯ ДАНИХ

Мичка С. О.

Науковий керівник - доцент Голян Н. В.

Харківський національний університет радіоелектроніки  
(61166, Харків, просп. Науки, 14, каф. Програмної інженерії,  
тел. (057) 702-14-46)

e-mail: sviatoslav.mychka@nure.ua

File compression algorithms have long been integral to information theory, serving diverse purposes in data storage, transmission, and resource optimization. With the proliferation of the Internet, their significance has only heightened across various domains. This paper describes the research on compression algorithms, which employ increasingly sophisticated techniques to enhance compression efficiency across diverse data types. Specifically, this research aims to analyze and compare the behavior of prominent compression algorithms, including entropy encodings, dictionary methods, and context modelling, with respect to their effectiveness in compressing different data types.

У сучасному світі кількість нових даних і, відповідно, затрати на їх зберігання й транспортування таких об'ємів даних теж зростають. Актуальність розробки та поліпшення алгоритмів стиснення даних полягає, насамперед, у тому, що сильніше стиснення даних із можливістю точно (за використання алгоритмів стиснення без втрат) відновити початкову інформацію дозволяє зменшити витрати на їх зберігання, транспортування й обробку. Алгоритми стиснення без втрат можуть бути ентропійними, словниковими, або поєднувати обидва підходи, а також використовувати інші допоміжні методи, такі як перетворення вихідного повідомлення або моделювання контексту. Тому деякі алгоритми можуть давати більші коефіцієнти стиснення з одними даними, і менші – з іншими, що потенційно може призвести до зайвих витрат ресурсів на зберігання або передавання даних.

Для вирішення цієї проблеми було поставлено задачу дослідження методів оптимального кодування і стиснення даних і порівняння характеристик їхньої роботи, знаходження алгоритмів стиснення, що працюють краще на певних видах даних. Потрібно дослідити алгоритми, їхні поєднання, підходи та методи стиснення без втрат даних різних видів, зокрема текстових, відео, статичних зображень тощо, визначити критерії порівняння тестових даних та алгоритмів, що досліджуються, і використати їх для порівняння ефективності застосування методів стиснення до певних видів файлів. Використовуючи отримані дані, необхідно розробити додаток, що підбиратиме необхідний алгоритм стиснення

для даних, наданих користувачем, а також провести ряд експериментів для оцінки ефективності роботи розробленого додатку.

Було проведено аналіз та обрано для дослідження алгоритми словникового стиснення даних LZ77, LZW і RLE із використанням перетворення Барроуза-Віллера, ентропійного стиснення за допомогою коду Хафмана, поєднання ентропійного та словникового методів у вигляді алгоритму Deflate, а також алгоритми, що застосовують моделювання контексту: PPM і Brotli. Ці алгоритми в наш час отримують розвиток і популярність в першу чергу через необхідність передавати більші об'єми даних мережею Інтернет і зберігати їх у хмарних сховищах.

Було розроблено план проведення експериментів, згідно з яким:

- визначено та поділено на категорії вхідні дані, що використовуватимуться в якості тестових, а саме: за середньою ентропією повідомлення, за наближеністю до тексту натуральною мовою, за повторюваністю символів;
- виконано стиснення тестових даних за допомогою алгоритмів, використовуючи бібліотеки або власні реалізації;
- порівняно результати, використовуючи такі метрики: коефіцієнт стиснення, час кодування, час декодування, кількість використаної оперативної пам'яті;
- збережено отримані дані в вигляді таблиці;
- розроблене програмне забезпечення для надання рекомендацій щодо використання алгоритмів стиснення для різних даних, використовуючи дані, отримані під час стиснення тестових даних;
- порівняно рекомендації щодо тестових даних із фактичними результатами їх стиснення.

Експерименти планується проводити з використанням локального комп'ютера з 4-ядерним процесором Intel Core i7-7700HQ і 16 Гб оперативної пам'яті (DDR4). Очікується, що метрики, отримані під час проведення експериментів дозволять краще зрозуміти принципи роботи алгоритмів із реальними даними. Вони можуть бути корисними для подальших досліджень в області стиснення даних, а також для розробки й покращення практичних рішень, що вже існують. Розроблене програмне забезпечення не лише може бути використане як практична реалізація для демонстрації результатів досліджень, а й потенційно має застосування як API для визначення оптимального алгоритму стиснення даних для їхнього подальшого зберігання або передачі мережею Інтернет.

Список використаних джерел:

1. David J. C. MacKay. (2005). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.
2. Sharonova N., Kyrychenko I., Gruzdo I., Tereshchenko G. (2022). Generalized Semantic Analysis Algorithm of Natural Language Texts for Various Functional Style Types. *CEUR Workshop Proceedings*, 3171, 16–26. <https://ceur-ws.org/Vol-3171/paper4.pdf>

## ДОДАТОК В

### Слайди презентації



# ДОСЛІДЖЕННЯ АЛГОРИТМІВ ОПТИМАЛЬНОГО КОДУВАННЯ І СТИСНЕННЯ ДАНИХ



Мичка С. О., ІПЗм-22-2  
Науковий керівник: к. т. н., доц. каф. ПІ, Голян Н. В.

20 червня 2024

Рисунок В.1 – Титульний слайд

## Дослідження

- У сучасному світі кількість нових даних невідомо зростає. Щороку об'єм даних, що створюються людьми, зростає на 30%.
- Алгоритм стиснення (компресії) даних – це таке кодування даних, яке забезпечує зменшення об'єму, що вони займають
- Основною характеристикою алгоритмів стиснення даних є коефіцієнт стиснення



Рисунок В.2 – Слайд «Дослідження»

## Постановка задачі

- дослідити алгоритми стиснення без втрат даних різних видів;
- визначити критерії порівняння алгоритмів, що досліджуються;
- використовуючи отримані дані, розробити додаток, що підбиратиме необхідний алгоритм стиснення для даних, наданих користувачем;



Рисунок В.3 – Слайд «Постановка задачі»

## Обчислення даних для дослідження

- Для визначення ефективності алгоритмів стиснення даних без втрат визначається коефіцієнт стиснення

$$k = \frac{V_0}{V_{\text{стисн}}}$$

- Вхідні дані необхідно порівнювати за середньою ентропією повідомлення

$$H(X) = -K \sum_{i=1}^n p_i * \log_2 p_i$$



Рисунок В.4 – Слайд «Обчислення даних для дослідження»

## Досліджувані алгоритми

- RLE
- BWT
- MTF
- LZ77
- LZW
- Кодування Хафмана
- Deflate
- BZip2
- Brotli
- PPM



Рисунок В.5 – Слайд «Досліджувані алгоритми»

## Проектування бази даних

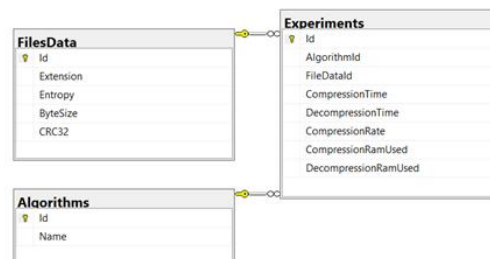


Рисунок В.6 – Слайд «Проектування бази даних»

# Розробка системи

- Було розроблено дві системи, що використовували одну БД
- Для розробки проекту для проведення експериментів використовувався С# і бібліотеки System.IO.Compression, SharpZipLib, SharpCompress
- Демонстраційний проект розроблено за допомогою ASP.NET, Entity Framework, React



Рисунок В.7 – Слайд «Розробка системи»

## Результати експерименту. Ентропія 0

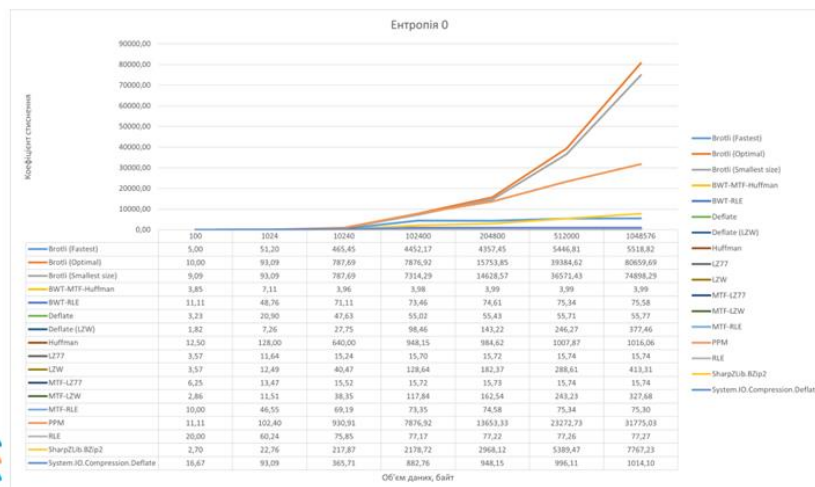


Рисунок В.8 – Слайд «Результати експерименту. Ентропія 0»

## Результати експерименту. Ентропія 1

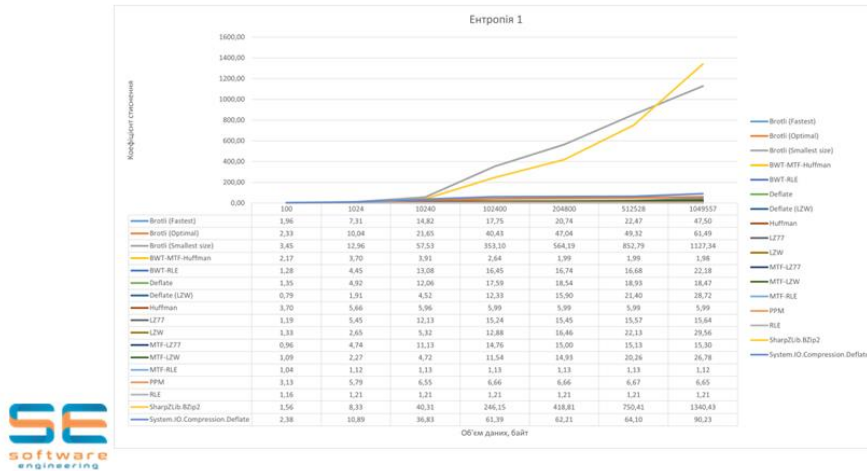


Рисунок В.9 – Слайд «Результати експерименту. Ентропія 1»

## Результати експерименту. Ентропія 4

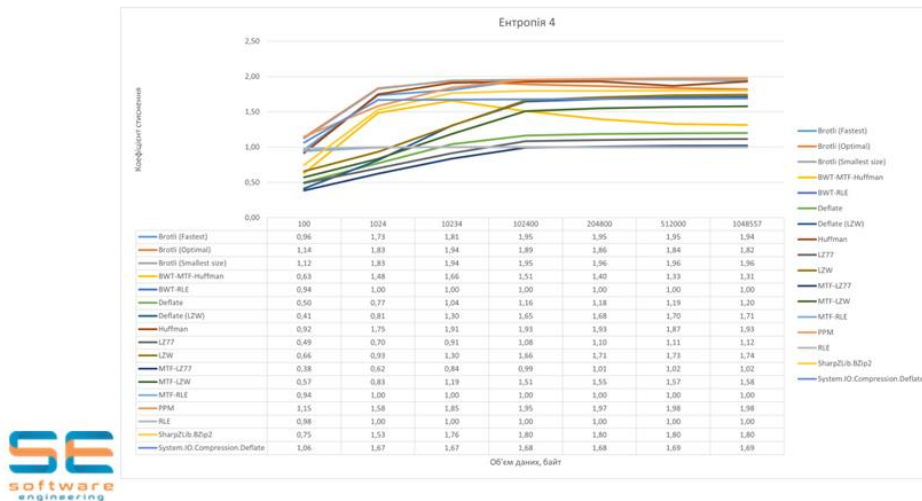


Рисунок В.10 – Слайд «Результати експерименту. Ентропія 4»

## Результати експерименту. Ентропія 7

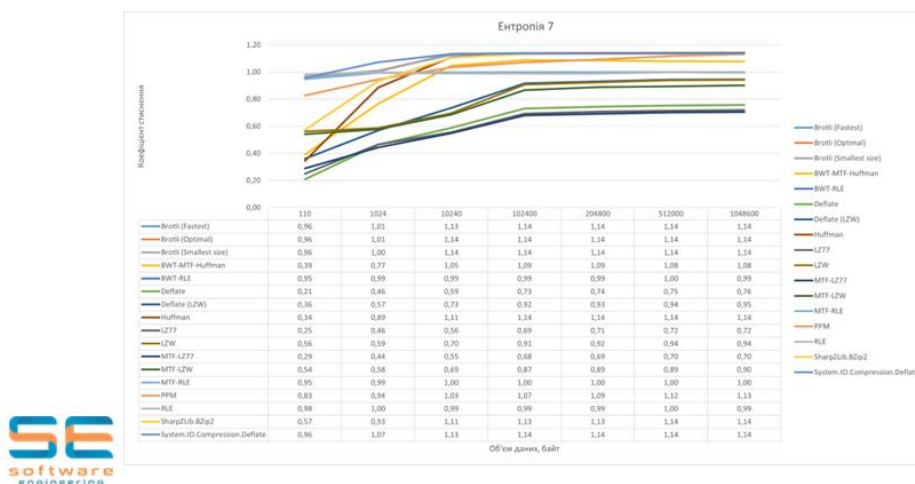


Рисунок В.11 – Слайд «Результати експерименту. Ентропія 7»

## Результати експерименту. Ентропія 8

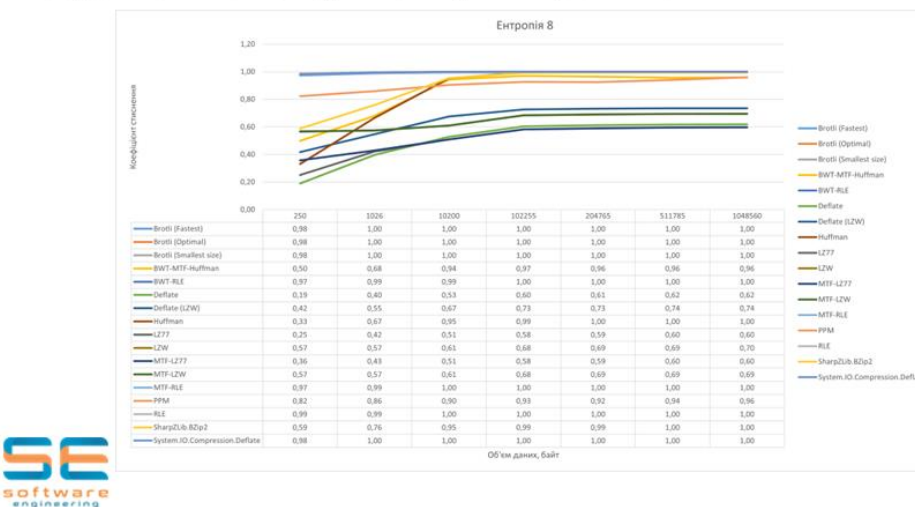


Рисунок В.12 – Слайд «Результати експерименту. Ентропія 8»

## Результати експерименту. Текстовий файл



Рисунок В.13 – Слайд «Результати експерименту. Текстовий файл»

## Результати експерименту. Зображення BMP та SVG

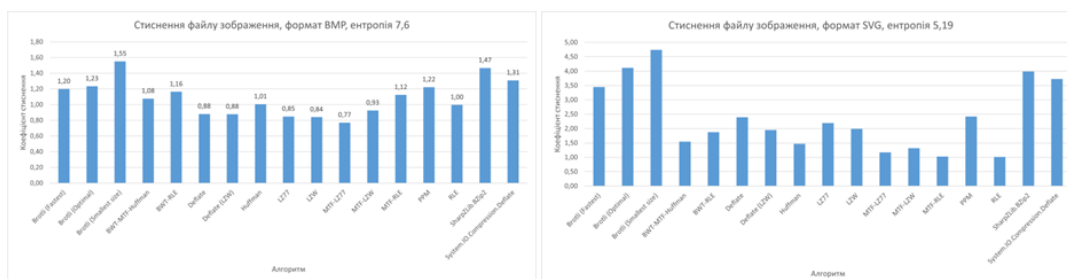


Рисунок В.14 – Слайд «Результати експерименту. Зображення BMP та SVG»

## Результати експерименту. HTML-сторінка



Рисунок В.15 – Слайд «Результати експерименту. HTML-сторінка»

## Аналіз отриманих результатів

- Ентропія, розмір та формат даних мають значення під час стиснення
- Не існує алгоритма, що однаково добре стискає будь-які дані
- На реальних даних BWT та MTF дають кращі результати
- На невеликих об'ємах даних часто переважають ентропійні методи



Рисунок В.16 – Слайд «Аналіз отриманих результатів»

## Висновки

- В ході виконання науково-дослідницької практики було розглянуто алгоритми стиснення без втрат, що найактивніше використовуються в сучасних застосунках для стиснення даних
- Було проведено експерименти та зроблено висновки про особливості роботи алгоритмів
- Було подано тези на 28-й Міжнародний форум «Радіоелектроніка і молодь в XXI ст.»



---

Рисунок В.17 – Слайд «Висновки»

## ДОДАТОК Г

## Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ



Ім'я користувача:  
Кардаш Євген Вікторович каф.ПІ

ID перевірки:  
1016356961

Дата перевірки:  
13.06.2024 14:34:15 EEST

Тип перевірки:  
Doc vs Internet + Library

Дата звіту:  
13.06.2024 14:50:36 EEST

ID користувача:  
100013622

Назва документа: 2024\_М\_ПІ\_ІПЗм\_22\_2\_Мичка\_С\_О\_скорочений

Кількість сторінок: 70 Кількість слів: 11042 Кількість символів: 78472 Розмір файлу: 3.95 MB ID файлу: 1016161216

Виявлено модифікації тексту (можуть впливати на відсоток схожості)

**0.87%**

**Схожість**

Найбільша схожість: 0.16% з Інтернет-джерелом (<https://doc.sitecore.com/xp/ja/developers/101/sitecore-experience-pla>).

0.79% Джерела з Інтернету

109

Сторінка 72

0.35% Джерела з Бібліотеки

41

Сторінка 72

**0% Цитат**

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

**0%**

**Вилучень**

Немає вилучених джерел

**Модифікації**

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи

5

Підозріле форматування

30  
сторінок

Рисунок Г.1 – Звіт з результатами перевірки роботи на академічну доброчесність

## ДОДАТОК Д

Експертний висновок результатів перевірки кваліфікаційної роботи на  
відповідність оформлення вимогам ДСТУ 3008:2015

## Експертний висновок результатів перевірки кваліфікаційної роботи

студент  
(посада)

програмної інженерії  
(кафедра)

ІПЗМ-22-2  
(група)

**Мичка Святослав Олегович**

(прізвище, ім'я, по батькові)

## Зауваження

Пункт ДСТУ 3008-2015	Зміст пункту	Сторінка кваліфікаційної роботи
1	2	3
	7.1 Загальні положення	
	7.3 Нумерація сторінок звіту	
	7.4 Нумерація розділів, підрозділів, пунктів, підпунктів	
	7.5 Рисунки	
	7.6 Таблиці	
	7.7 Переліки	
	7.8 Примітки	
	7.9 Виноски	
	7.10 Формули та рівняння	
	7.11 Посилання	
	7.13 Список авторів	
	7.14 Скорочення та умовні позначки	
	7.15 Додатки	

зауважень немає

Експерт

\_\_\_\_\_

(відпис)

**Олена ОЛІЙНИК**

\_\_\_\_\_

(прізвище, ініціали)

16.06.2024

Рисунок Д.1 – Експертний висновок результатів перевірки кваліфікаційної роботи