**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ   
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Факультет комп’ютерних наук та кібернетики

Кафедра математичної інформатики

«До захисту допущено»

Завідувач кафедри

Терещенко В.М

(підпис)

«\_\_\_\_» \_\_\_\_\_\_\_\_\_ 20\_\_р.

**Дипломна робота**

**На здобуття ступеня бакалавра**

за спеціальністю 122 Комп’ютерні науки

на тему:

**ПЕРЕТВОРЕННЯ БАРРОУЗА-УІЛЛЕРА ТА ЙОГО ЗАСТОСУВАННЯ**

Виконав студент 4-го курсу  
Волохович Ігор Ігорович

(підпис)

Науковий керівник:

професор, доктор фіз.-мат. наук

Анісімов Анатолій Васильович

(підпис)

Засвідчую, що в цій курсовій роботі немає запозичень з праць інших авторів без відповідних посилань.

Студент

(підпис)

Київ – 2021

**Реферат**

Обсяг роботи ХХ сторінки, ХХ ілюстрацій, ХХ таблиць, ХХ джерел посилань.

ПЕРЕТВОРЕННЯ БАРРОУЗА-УІЛЛЕРА, МЕТОДИ СТИСНЕННЯ ДАНИХ, АЛГОРИТМ ЛЕМПЕЛЯ–ЗІВА–ВЕЛЧА, АЛГОРИТМ ХАФФМАНА

Об’єктом роботи є дослідження методів стиснення даних, перетворення Барроуза – Уіллера та доцільність його використання у методах стиснення.

Метою роботи є розробка алгоритмів стиснення даних та перетворення Барроуза–Уіллера. Модифікація одного з алгоритму та порівняння якість та час стиснення даних.

Методи розроблення: аналіз алгоритмів стискання, кодування методів стиснення, аналіз якості та часу стиснення алгоритмів та їх модифікацій, розробка загальної бібліотеки алгоритмів, розробка прикладів використання алгоритмів, розробка веб–застосунку для демонстрації можливостей бібліотеки.

Інструменти розроблення. Для розробки реалізації було використано інтерактивне середовище розробки JebBrains Rider[1]. Було використано мову C#. Створено .NET Library та використовуючи ASP.NET Framework створено веб – застосунок. Розроблено систему логування часу та якості стиснення.

Результат роботи. Було досліджено методи стиснення даних та вже існуючі продукти стиснення на ринку. Розроблено декілька демонстраційних консольних програм та веб-застосунок. Також було запроваджено стандартну загальну бібліотеку методів, яка було викладена в мережу на правах open-source та доступна всім розробникам. Було показано використання бібліотеки на різних платформах.

**Зміст**

[ВСТУП 4](#_Toc70961041)

[РОЗДІЛ 1. ЗАГАЛЬНІ ВІДОМОСТІ ПРО МЕТОДИ СТИСНЕННЯ ДАНИХ 6](#_Toc70961042)

[1.1 Перше застосування терміну «База Даних» 6](#_Toc70961043)

[РОЗДІЛ 2. ЗАГАЛЬНІ ВІДОМОСТІ ПРО ПЕРЕТВОРЕННЯ БАРРОУЗА-УІЛЛЕРА (BWT) 6](#_Toc70961044)

[2.1 Відомості про Entity Framework 6](#_Toc70961045)

[РОЗДІЛ 3. ІМПЛЕМЕНТАЦІЯ ПЕРЕТВОРЕННЯ BWT 6](#_Toc70961046)

[3.1 Відомості про Entity Framework[7] 6](#_Toc70961047)

[РОЗДІЛ 4. ІМПЛЕМЕНТАЦІЯ МЕТОДІВ СТИСНЕННЯ 6](#_Toc70961048)

[4.1 Стандартний метод LZW 6](#_Toc70961049)

[4.2 Метод LZW з перетворенням Барроуза-Уіллера 6](#_Toc70961050)

[4.3 Метод Хаффмана (Huffman coding) 6](#_Toc70961051)

[РОЗДІЛ 5. ПОРІВНЯННЯ МЕТОДІВ СТИСНЕННЯ 6](#_Toc70961052)

[ВИСНОВКИ 6](#_Toc70961053)

[ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ 6](#_Toc70961054)

[ДОДАТКИ 6](#_Toc70961055)

# ВСТУП

**Оцінка сучасного стану об’єкта дослідження.** На сьогоднішній день, зі зростанням кількості користувачів та кількості даних в мережі інтернет гостро постає питання зберігання даних. Так як будь-яка людина розуміє, що просто зберігати файли не дуже правильно, люди замислюються щодо використання алгоритмів стиснення задля збереження дорогого та цінного місця на серверах, комп’ютерах, мережевих системах. Через це були придумані та розроблені різні алгоритми стиснення даних та допоміжні перетворення для збільшення якості стиснення. Наприклад великій продуктовій компанії Amazon не вигідно зберігати сотні мільйонів даних про користувачів, товари та історію у первісному вигляді. Це займає багато місця. Тому вони використовують алгоритми, які допомагають в такому же об’ємі пам’яті зберігати набагато більше файлів.

**Актуальність роботи та підстави для її виконання.** Застосування, розробка нових та модифікація існуючих алгоритмів стиснення має важливе значення зі збільшенням генерації інформації у сьогоднішній час. Задля раціонального використання пам’яті та попередження надлишкової інформаційної катастрофи, коли не вистачатиме пам’яті. Методи стиснення та модифіковані (наприклад з використанням алгоритму Барроуза–Уіллера) допомагають зменшити розмір файлу тим самим зменшуючи витрати та попередження настання інформаційної катастрофи.

**Мета й завдання роботи.** Метою дипломної роботи є огляд алгоритму перетворення Барроуза–Уіллера, модифікація алгоритму Лемпелля – Зівела – Велча використовуючи цей алгоритм та порівняльний аналіз результатів стиснення, які приведуть до висновку модифікування алгоритмів у такий спосіб. Також завданням є реалізація власне перетворення BWT(Барроуза – Уіллера), алгоритму стиснення LZW(Лемпелля – Зівела – Велча), модифікація даного алгоритма використовуючи перетворення та реалізація алгоритму Huffman coding(алгоритм стиснення Хаффмана). Порівняння даних результатів у розрізі часу виконання та результуючому розмірі файла. Для досягнення цієї поставлено такі завдання:

* Огляд методу BWT на текстових та байтових даних.
* Огляд основних методів стиснення та їх алгоритмів.
* Провести роботу у дослідженні розробки бібліотек
* Розробити програмну реалізацію алгоритму BWT, LZW, LZW + BWT, Huffman coding з можливістю використовувати у будь-якому проекті на базі мови C#.

**Об'єкт, методи й засоби розроблення**. Об’єктом порівняльного аналізу є стандартні та модифіковані алгоритми стиснення даних. Об’єктом розробки є декілька програмних реалізацій, які допомагають побачити та застосувати бажаний алгоритм стиснення та бібліотеку з алгоритмами, яка може бути використана та встановлена будь-яким розробником через NuGet пакет у середовищі мови програмування C#

# РОЗДІЛ 1. ЗАГАЛЬНІ ВІДОМОСТІ ПРО МЕТОДИ СТИСНЕННЯ ДАНИХ

## Що таке стиснення даних?

Стиснення даних – це зменшення кількості бітів, необхідних для репродукції даних. Такі алгоритми допомагають зберігати пам’ять, збільшити швидкість передачі даних та зменшити вартість накопичувачів та необхідну пропускну властивість інтернету для передачі. Іншими словами, це процес кодування, реструктуризації або модифікування даних для зменшення її розміру. Фундаментально, це включає повторне шифрування інформації, користуючись меншими бітами (шматками), ніж оригінальне представлення. Стиснення даних підлягає компромісу між кількістю пам’яті, яку займає та часом. Це широко застосовується в обчислювальних сервісах та рішеннях, особливо в комунікації даними, Стискання даних працює за допомогою декількох методів стиснення та програмних рішень, які використовують алгоритми стискання для зменшення обсягу даних.

## Історія створення методів стиснення

Теоретичну базу стискання даних було розроблено Клодом Шеноном, який опублікував роботи по алгоритмічній теорії інформації(AIT) для компресії без втрат та теорії швидкості спотворення для компресії з втратами. Вони стали фундаментом започаткування стискання даних, які на сьогоднішній день досягли величезних здібностей та дуже часто застосовуються у машинному навчанні.

З народженням інтернету в 70 – х роках, взаємозв’язок між розміром файлів та швидкістю передачі даних став більш значущим. Видатні розуми всього світу боролися з цією проблемою роками, але цього не було допоки універсальний алгоритм стиснення без втрат Лемпелля – Зівела – Велча (LZW) не з’явився у середині 80 – х роках, який показав реальні переваги стискання.

Метод LZW був першим широко вживаним алгоритмом стискання даних, який був імплементований на комп’ютерах та досі використовується сьогодні (у різноманітних варіаціях). Великий текст англійською мовою може бути стиснутий приблизно на 50% використовуючи LZW. Навіть код Морзе (Morse core), створений у 1838 році, вважається найранішим методом стискання даних оскільки найпоширеніші літери в англійській мові, такі як «е» і «t» мають коротші коди Морзе.

## Як працюють методи стиснення

## Типи стиснення

## Переваги та недоліки стискання даних

# РОЗДІЛ 2. ЗАГАЛЬНІ ВІДОМОСТІ ПРО ПЕРЕТВОРЕННЯ БАРРОУЗА-УІЛЛЕРА (BWT)

## Відомості про Entity Framework

# РОЗДІЛ 3. ІМПЛЕМЕНТАЦІЯ ПЕРЕТВОРЕННЯ BWT

## Відомості про Entity Framework[[7]](#_ПЕРЕЛІК_ВИКОРИСТАНИХ_ДЖЕРЕЛ)

# РОЗДІЛ 4. ІМПЛЕМЕНТАЦІЯ МЕТОДІВ СТИСНЕННЯ

## Стандартний метод LZW

## Метод LZW з перетворенням Барроуза-Уіллера

## Метод Хаффмана (Huffman coding)

## РОЗДІЛ 5. ПОРІВНЯННЯ МЕТОДІВ СТИСНЕННЯ

# ВИСНОВКИ

# ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Jeb Brains Rider [Електронний ресурс]. – Режим доступу до ресурсу: <https://www.jetbrains.com/rider/>.
2. C# programming langauge [Електронний ресурс] – Режим доступу до ресурсу: <https://en.wikipedia.org/wiki/C_Sharp_(programming_language)>.
3. .NET Library [Електронний ресурс] // Microsoft. – 2018. – Режим доступу до ресурсу: <https://docs.microsoft.com/en-us/dotnet/standard/library-guidance/>.
4. ASP.NET Framework [Електронний ресурс] // Microsoft. – 2021. – Режим доступу до ресурсу: https://dotnet.microsoft.com/apps/aspnet.

# ДОДАТКИ

ДОДАТОК А