

МЕТОД ТА КРОСПЛАТФОРМЕННИЙ ЗАСІБ АРХІВАЦІЇ ОДНОТИПНИХ ФАЙЛІВ

Виконав: студент групи 1КІ-18м


Чирва Павло

Керівник: к. т. н., доц. каф. ОТ

Савицька Л. А.


Актуальність дослідження

- ▣ Задача компактного зберігання, перетворення та передавання інформаційних даних завжди була актуальною в галузі інформаційних технологій.
- ▣ Інформаційні ресурси нині є продуктом інтелектуальної діяльності дійсно найбільш кваліфікованої й творчо активної частини молоді та працездатного населення світу.
- ▣ В якості доказу є об'єктивне збільшення інформаційного потоку з початку цього сторіччя більш ніж в 30 разів!
- ▣ Отже, **актуальною** є наукова задача розробки та застосування принципово нових методів і засобів сприйняття, передачі, обробки, зберігання і розповсюдження інформаційних даних, таких, що здатних оперувати великими масивами інформації, причому, у реальному часі.




Мета, об'єкт, предмет дослідження

- ▮ **Метою** дослідження магістерської кваліфікаційної роботи є збільшення середнього значення процесу архівації для великої кількості однотипних файлів
- ▮ **Об'єкт дослідження** – це сучасні процеси архівування та стиснення інформаційних даних
- ▮ **Предмет дослідження** – це методи та програмні засоби словникового методу архівування та стиснення інформаційних даних



Наукова новизна одержаних результатів

- вперше запропоновано метод архівації однотипних файлів, який дозволяє збільшити середнє значення процесу архівації однотипних файлів;
- вдосконалено процес формування основного словника за рахунок застосування методу «ковзного вікна» до його формування;



Практичне значення одержаних результатів

- Розроблено новий метод архівації однотипних файлів.
- Вдосконалено процес формування основного словника за рахунок застосування до його формування методу «ковзного вікна».
- Розроблено алгоритм роботи формування основного словника однотипних файлів, який дозволяє ефективніше стискати велику кількість однотипних файлів.
- Розроблено програмний засіб для архівації однотипних файлів.

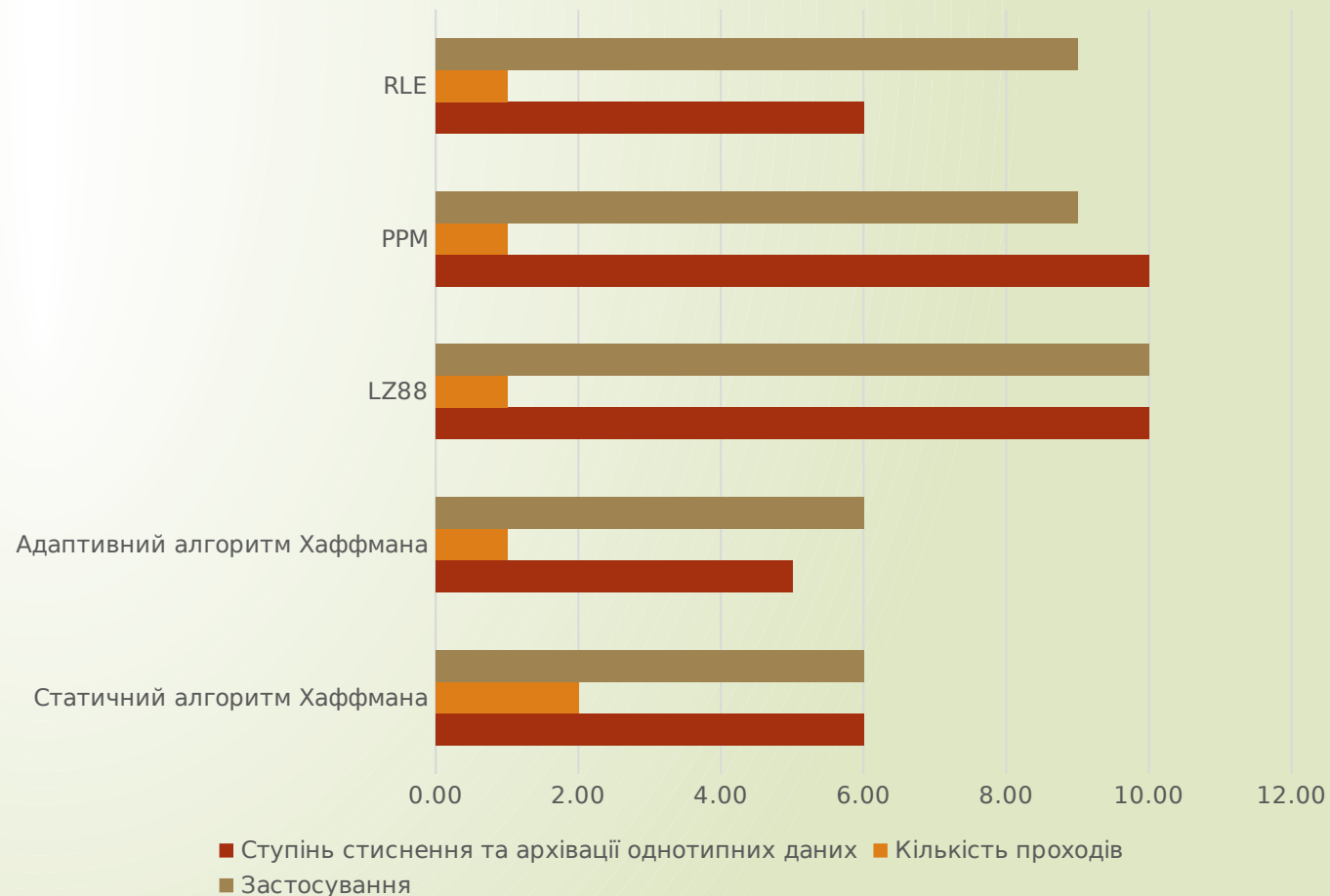
Аналіз методу словникового стиснення та архівації

- ▮ Суть процесу словникового стиснення та архівації полягає в заміні послідовностей деяких елементів вихідних даних на ID
- ▮ Серед словникових методів найбільше поширення мають методи Зіва-Лемпеля і їх можна розділити на 2 сімейства: LZ77 (LZ1) і LZ78 (LZ2).
- ▮ Словник, що застосовується у таких словникових методах для стиснення та архівації даних, можна розглядати як аналог статистичної моделі, що і застосовується у статистичних методах.

Інші методи стиснення та архівації даних


- Кодування серій, RLE
- Алгоритм LZ77
- Алгоритм Шеннона-Фано
- Алгоритм Хаффмана
- Алгоритм стиснення та архівації PPM

Порівняння алгоритмів архівації та стиснення



Порівняння алгоритмів

Алгоритм	Ступінь стиснення та архівації	Стиснення без втрат?	Кількість проходів	Застосування	Належність
Статичний алгоритм Хаффмана	5-6	так	2	Стиснення текстів бінарної інформації, ін.	Неблоковий, статистичний
Адаптивний алгоритм Хаффмана	4-5	так	1	Стиснення текстів бінарної інформації, ін.	Неблоковий, статистичний
LZ88	9-10	так	1	Універсальний	Блоковий
PPM	10-12	так	1	Універсальний (краще - текст і зображення)	Адаптивний, статистичний
RLE	2-6	так	1	Стиснення аудіо та зображень	Блоковий, статистичний

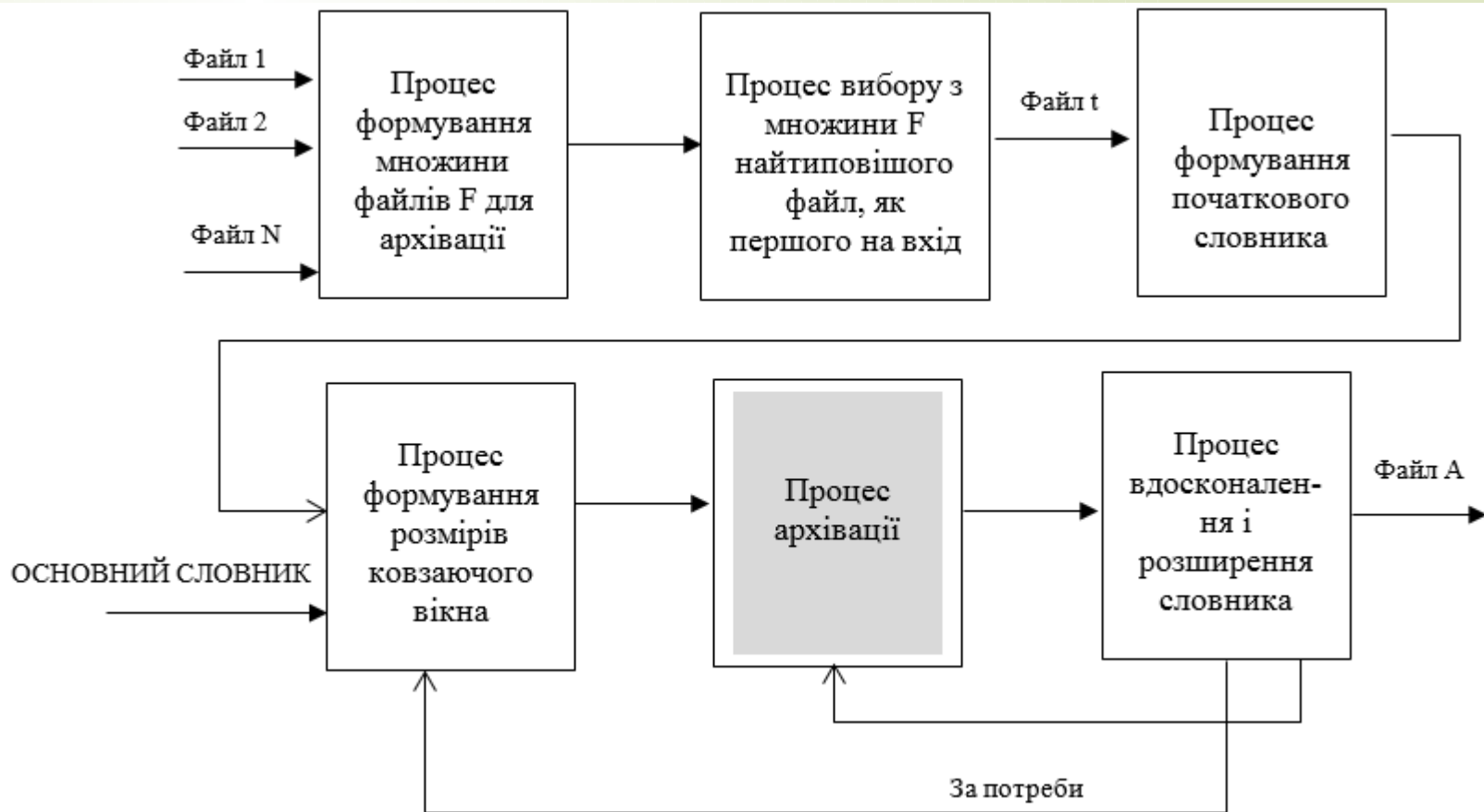


Порівняння алгоритмів

- Алгоритм LZ був обраний, оскільки він виконує процеси стиснення та архівації даних без втрат, є досить простим в практичній (програмній також) реалізації та актуальним нині, тому що LZ широко застосовується в інших методах стиснення та архівації в якості засобу для архівації однотипних файлів.

Метод архівації однотипних

ф

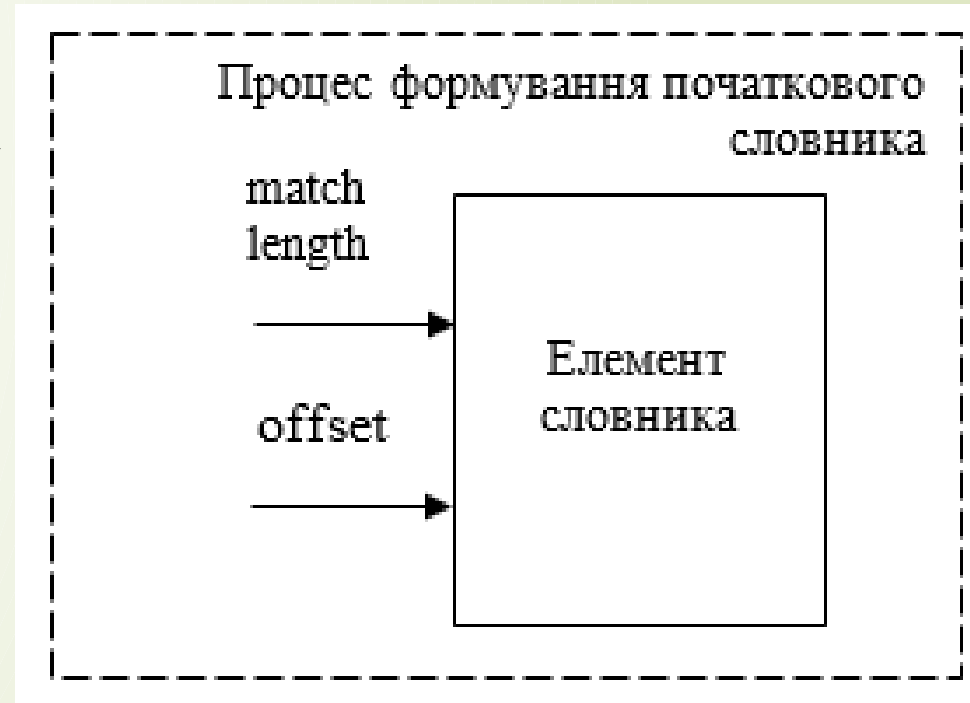


Розроблений метод призначений для процесів оптимізованої архівації великих кількостей маленьких однотипних файлів. В основі запропонованого методу є ідея заміни повторного входження цілого блока даних посиланням на попередню позицію його входження.

Процес формування початкового словника

Процес формування словника у стандартному використанні алгоритму LZ це формування комбінації, що повторюється і кодуються парою:

- довжина збігу (match length)
- зсув (offset) або дистанція (distance).

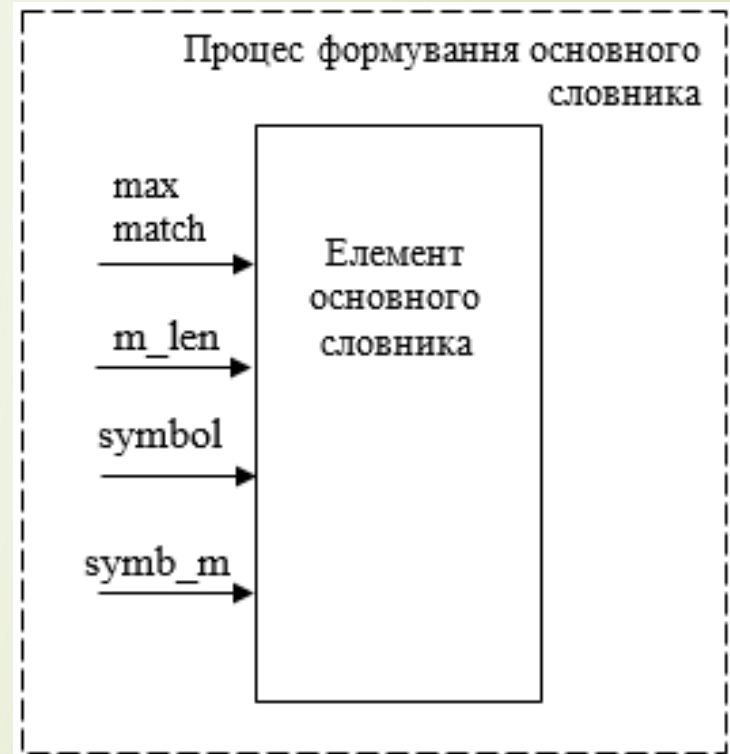


Процес формування ОСНОВНОГО СЛОВНИКА

Кодується такими величинами:

- код словникового блоку, який дав максимальний збіг – `max match`;
- довжина максимального збігу `m_len`;
- символ, що порушив співпадіння `symbol`;
- сам символ `symb_m`.

Процес формування основного словника вдосконалено **саме** за рахунок застосування до його формування методу «**КОВЗНОГО ВІКНА**».



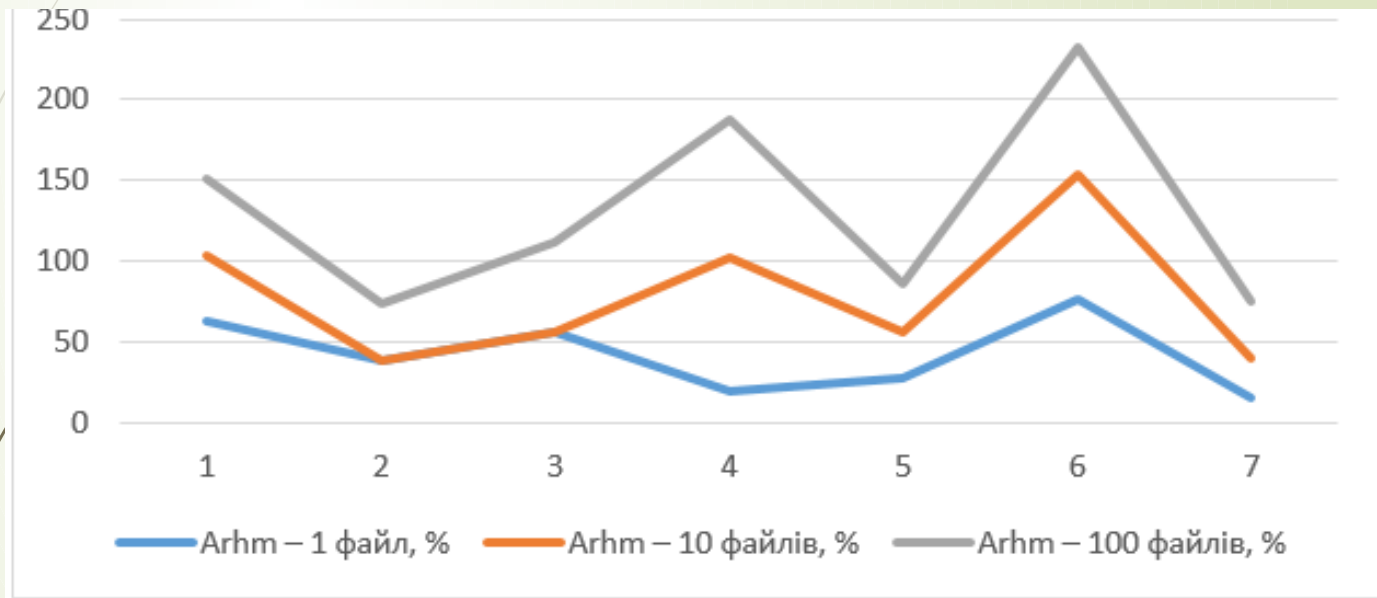
Процес архівації

Особливість розглянутого в даному методі архівації однотипних файлів є сам процесу архівації, якій полягає в тому, що використання комбінації кодової пари «довжина-зміщення» є не тільки прийнятним, але й ефективним у тих випадках, коли значення довжини збігу *match length* перевищує значення зсуву *offset*.

Середнє значення процесу архівації

Тип файлів	$Arch_m - 1$ файл, %	$Arch_m - 10$ файлів, %	$Arch_m - 100$ файлів, %
Документ (*.docx)	62,875	40,33	47,4
Документ (*.doc)	37,7825	32,567	35,7
Текстовий (*.txt)	55,305	47,376	56,78
Бази даних (*.accdb)	18,89	83,02	85,189
Графічні файли (*.jpg)	27,31	28,5	30,01
Звукові файли (*.mp3)	76,4	77,08	78,5
HTML (*.html)	14,74	24,129	35,67
Усереднене значення показника архівування	41,9003571	50,612	52,75

Середнє значення процесу архівації

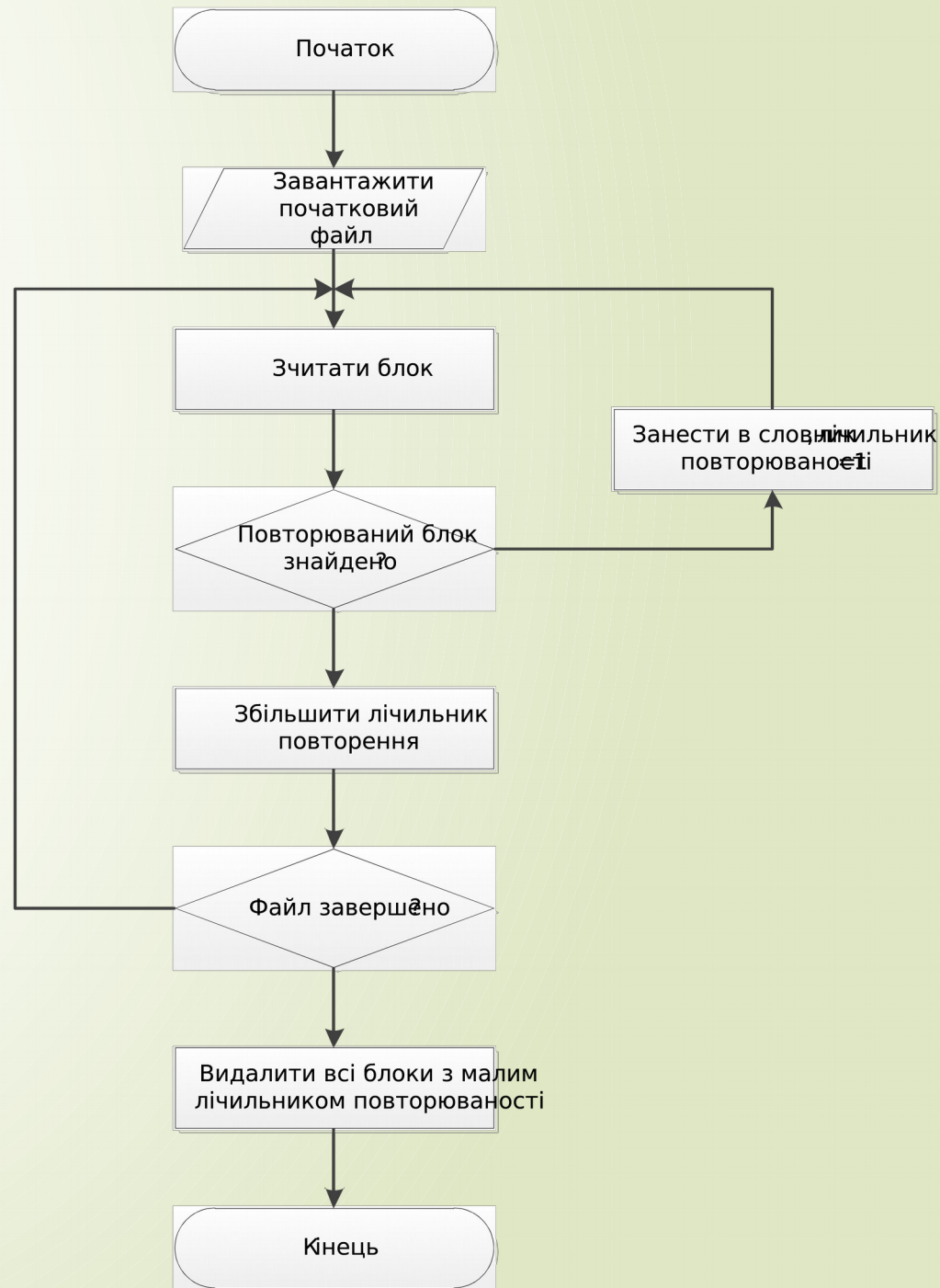


Запропонований метод архівації однотипних файлів дає приріст середнього значення процесу архівації на **10,85%**. А зі збільшенням кількості однотипних файлів для процесу архівації цей показник може все більше зростати.

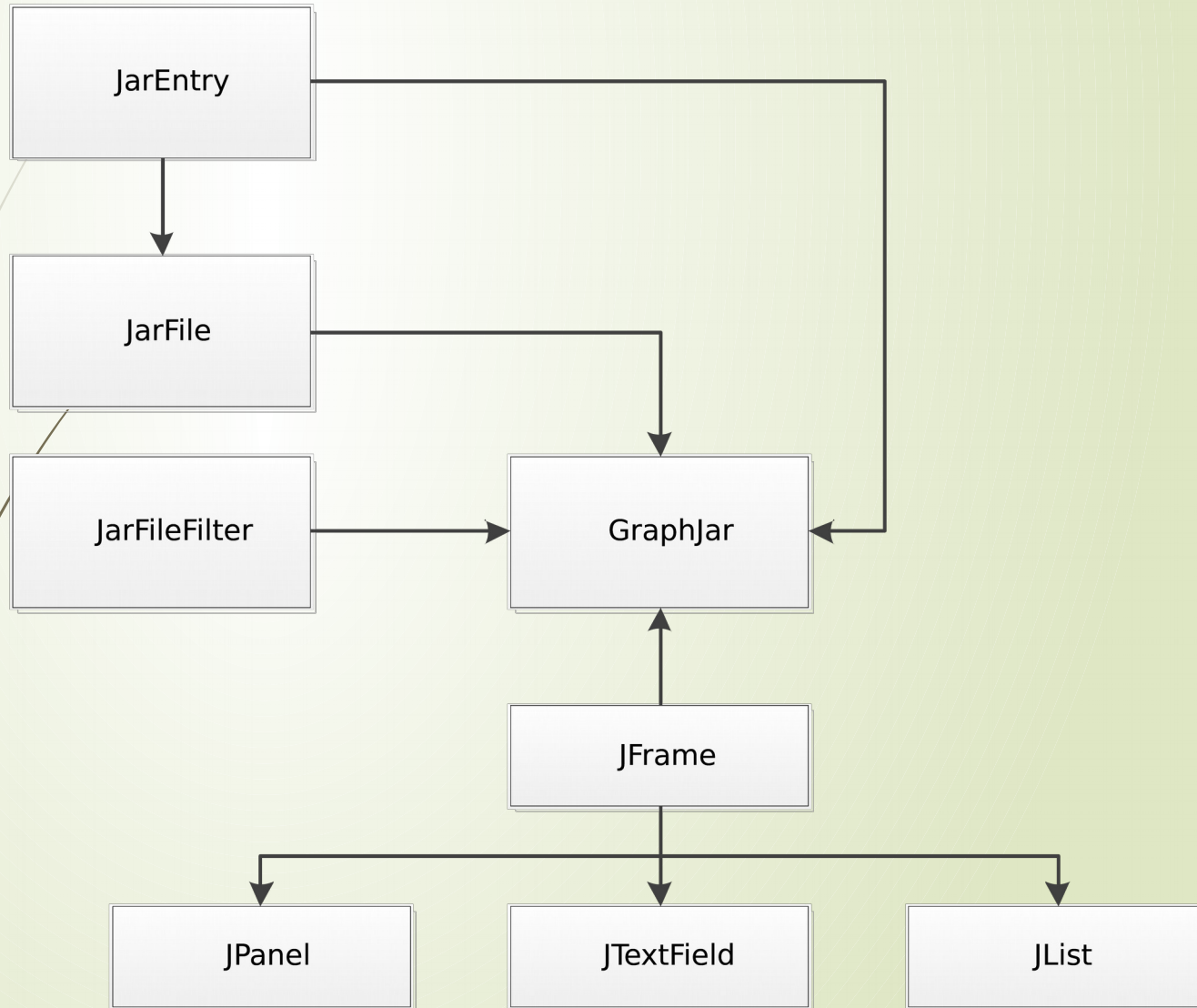
Алгоритм роботи кросплат-форменого засобу архівації однотипних файлів



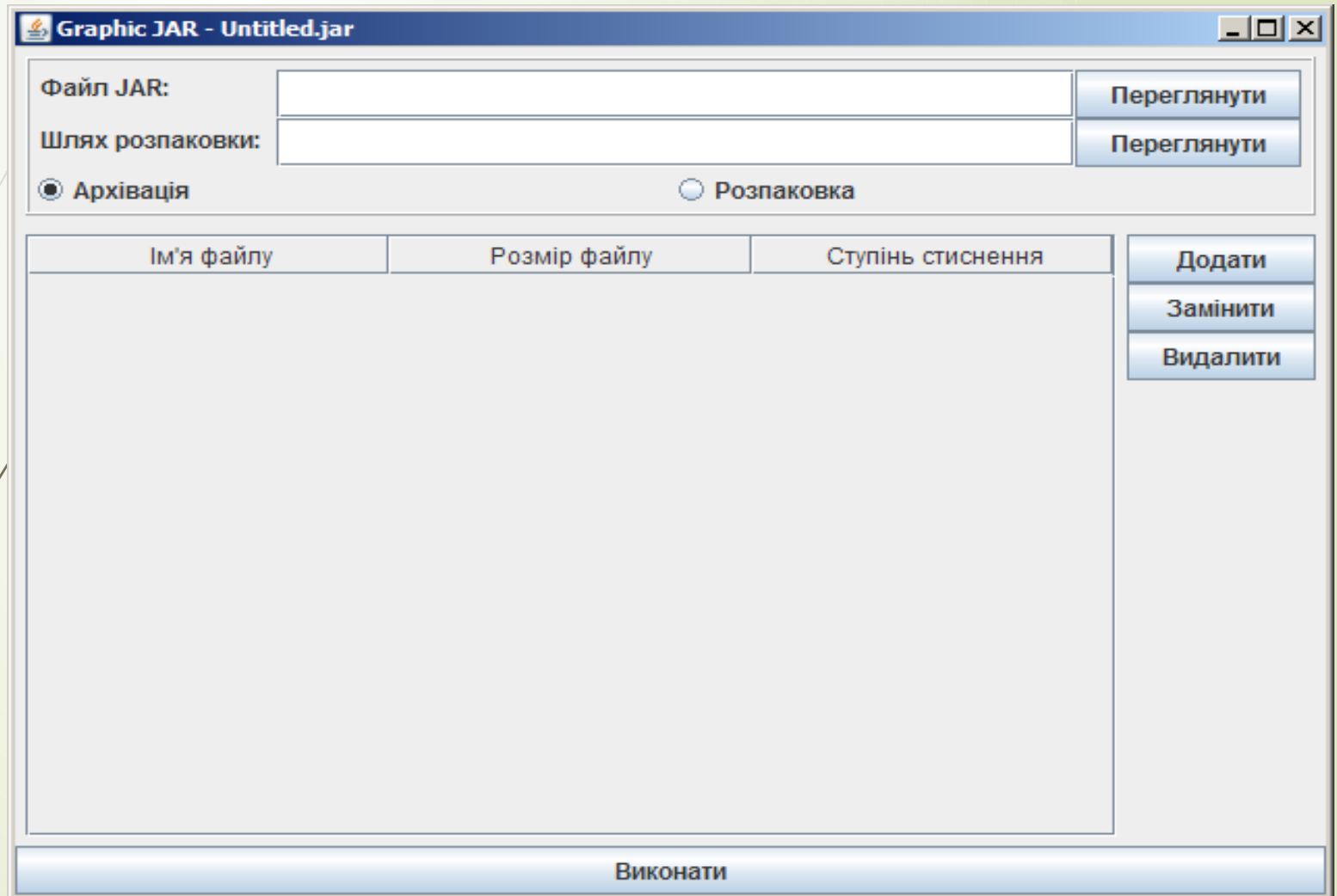
Алгоритм створення основного словника



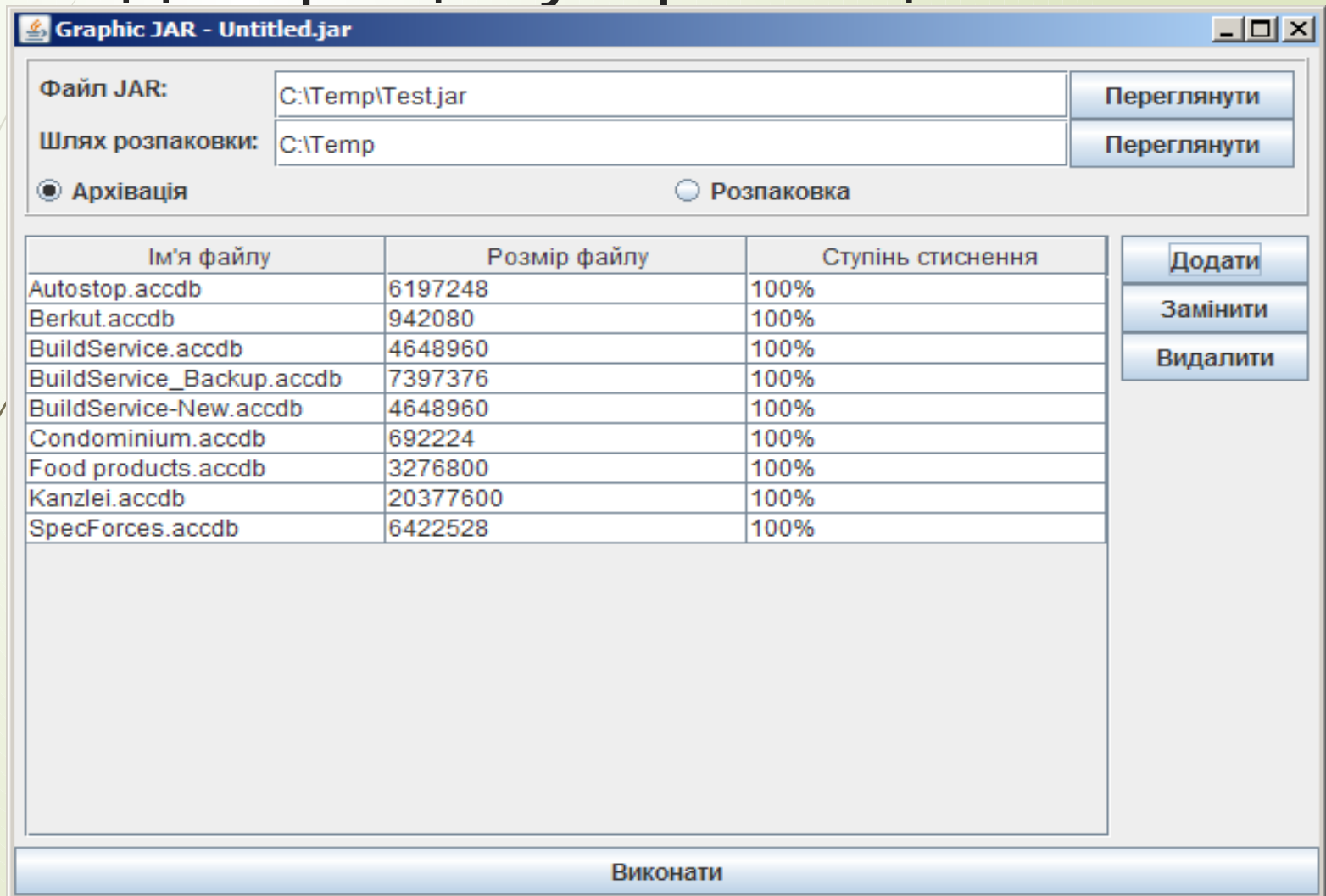
Діаграма класів



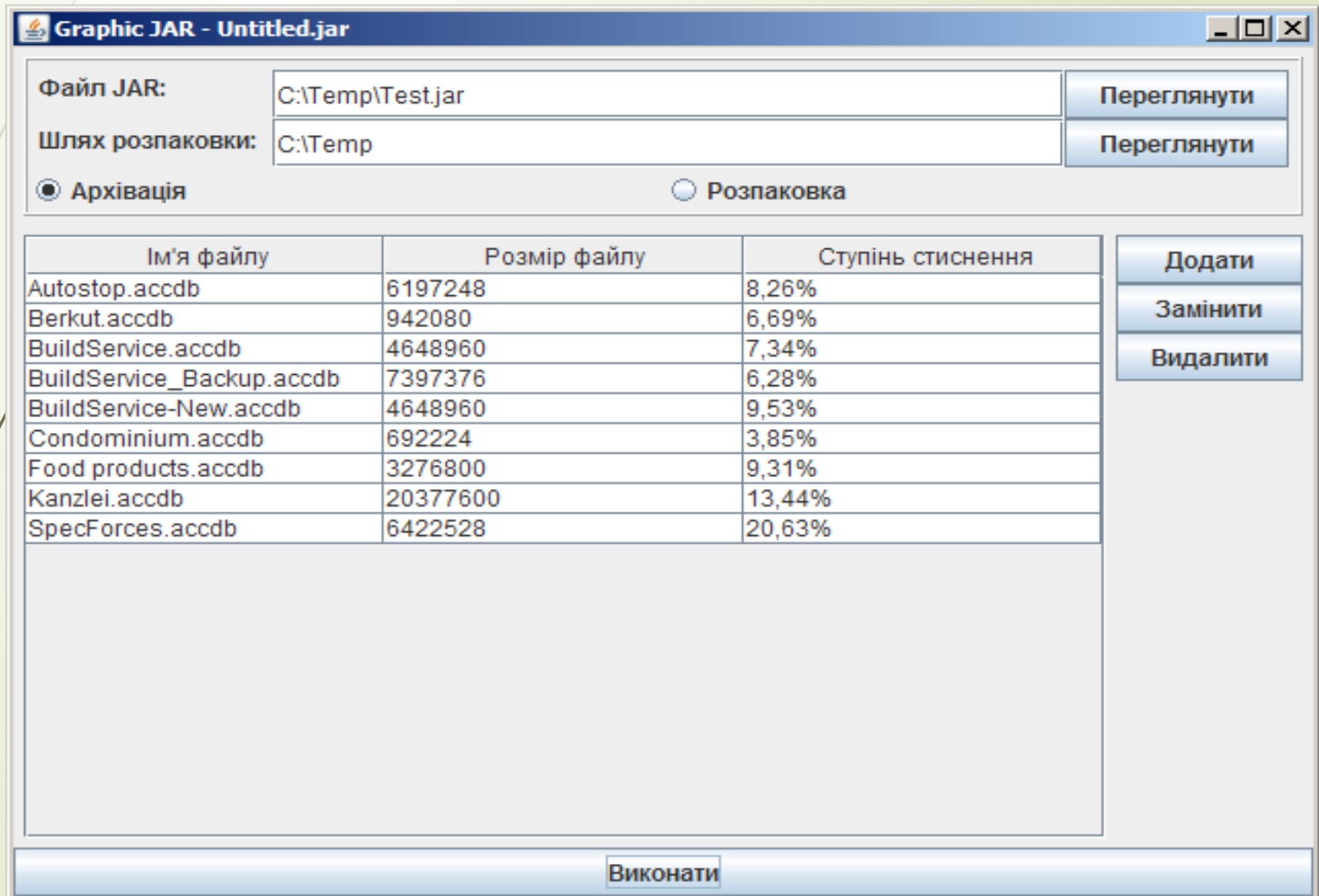
Дизайн користувацького інтерфейсу



Формування списку файлів до процесу архівації



Стан програми архівації однотипних файлів після виконання процесу архівації




ВИСНОВОК

- ▣ Розроблений метод призначений для процесів оптимізованої архівації великих кількостей маленьких однотипних файлів. В основі запропонованого роботи методу архівації однотипних файлів є ідея заміни повторного входження цілого блока даних посиланням на попередню позицію його входження.
- ▣ Зокрема, у роботі отримані такі наукові результати:
 - ▣ вперше запропоновано метод архівації однотипних файлів, який дозволяє збільшити середнє значення процесу архівації однотипних файлів на 10, 85%;
 - ▣ вдосконалено процес формування основного словника за рахунок застосування методу «ковзного вікна» до його формування;
 - ▣ вдосконалено процес архівації за рахунок застосування вдосконаленого процесу формування основного словника.

ВИСНОВОК

Практичне значення одержаних результатів полягає у такому: розроблено новий метод архівації однотипних файлів, вдосконалено процес формування основного словника за рахунок застосування до його формування методу «ковзного вікна»; розроблено алгоритм роботи формування основного словника однотипних файлів, який дозволяє ефективніше стискати велику кількість однотипних файлів; розроблено програмний засіб для архівації однотипних файлів.

В рамках даної роботи виявлено, що розроблений метод архівації однотипних файлів гарно працює в умовах наявності не менше 10 файлів на вході методу, а при наявності більшої кількості (100 і більше) однотипних файлів значення Arh_m росте. Запропонований метод дає приріст середнього значення процесу архівації на 10,85%, а зі збільшенням кількості однотипних файлів для процесу архівації цей показник може все більше



Дякую за увагу!
Доповідь закінчено.

