

# РОЗРОБКА ПРОГРАМНОГО МОДУЛЮ АВТОМАТИЗОВАНОГО АНАЛІЗУ ТЕКСТОВОЇ ІНФОРМАЦІЇ СИСТЕМИ ФОРМУВАННЯ КАДРОВИХ РЕЗЕРВІВ

Вінницький національний технічний університет

## *Анотація*

*Реалізовано паралельний алгоритм обробки текстових даних з метою підвищення швидкості обробки текстових даних.*

**Ключові слова:** алгоритм, потік, швидкодія, синтаксичний аналіз інформації, хеш-сума.

## *Abstract*

*A parallel algorithm for processing text data is implemented in order to increase the processing speed of the data.*

**Keywords:** algorithm, thread, speed, syntactic analysis of information, hash-sum.

## **Вступ**

Завдання створення бази кадрових резервів завжди актуальним. Особливо великі можливості підбору кадрів виникли з розвитком інформаційних технологій, які практично знищили обмеження передачі даних на відстані і сприяли виникненню великої кількості доступних баз даних з інформацією як про вакансії для працевлаштування, так і про тих, хто шукає роботу. Це викликало появу великої кількості агентств з працевлаштування, кадрових та рекрутингових агентств. З'явився і такий специфічний напрям кадрових організацій, як хедхантинг (headhunting - «полювання за головами»), який забезпечує пошук справжніх майстрів своєї справи, професіоналів з великим досвідом, висококваліфікованих фахівців. Саме тому, для полегшення роботи обробки даних працівників та створення системи прийняття рішень щодо відбору найкращих кандидатів на певні вакансії було вирішено створити спеціальну систему формування кадрових резервів, важливою складовою якої є модуль автоматизованого аналізу текстової інформації, що зберігає та обробляє особисті дані працівників.

## **Постановка задачі**

Метою даної роботи є розробка програмного додатку, що допоможе автоматизувати обробку текстових даних та значно підвищити її швидкість відносно відомих аналогів.

Для досягнення мети необхідно забезпечити вирішення таких задач:

- читання файлів анкет;
- обробка анкет;
- запис до файлу результатів роботи програми;
- формування хеш-суми;
- обробка помилок;
- тестування програмного додатку.

## **Результати розробки**

Результатом розробки став програмний додаток, що забезпечує обробку анкети кандидатів та працівників, формує унікальні ідентифікатори для кожного з них, створює базу даних, що вміщує результуючий список, та забезпечує паралельну обробку текстових даних (рис.1).

Реалізація алгоритму полягає в пошуку по ключових словах, а саме назвах полів в тексті, а також відокремленні назв полів від їх значень, створенні структури даних кандидата, формуванні хеш-суми сукупності даних та створенні відповідного кортежу в результуючій базі даних.

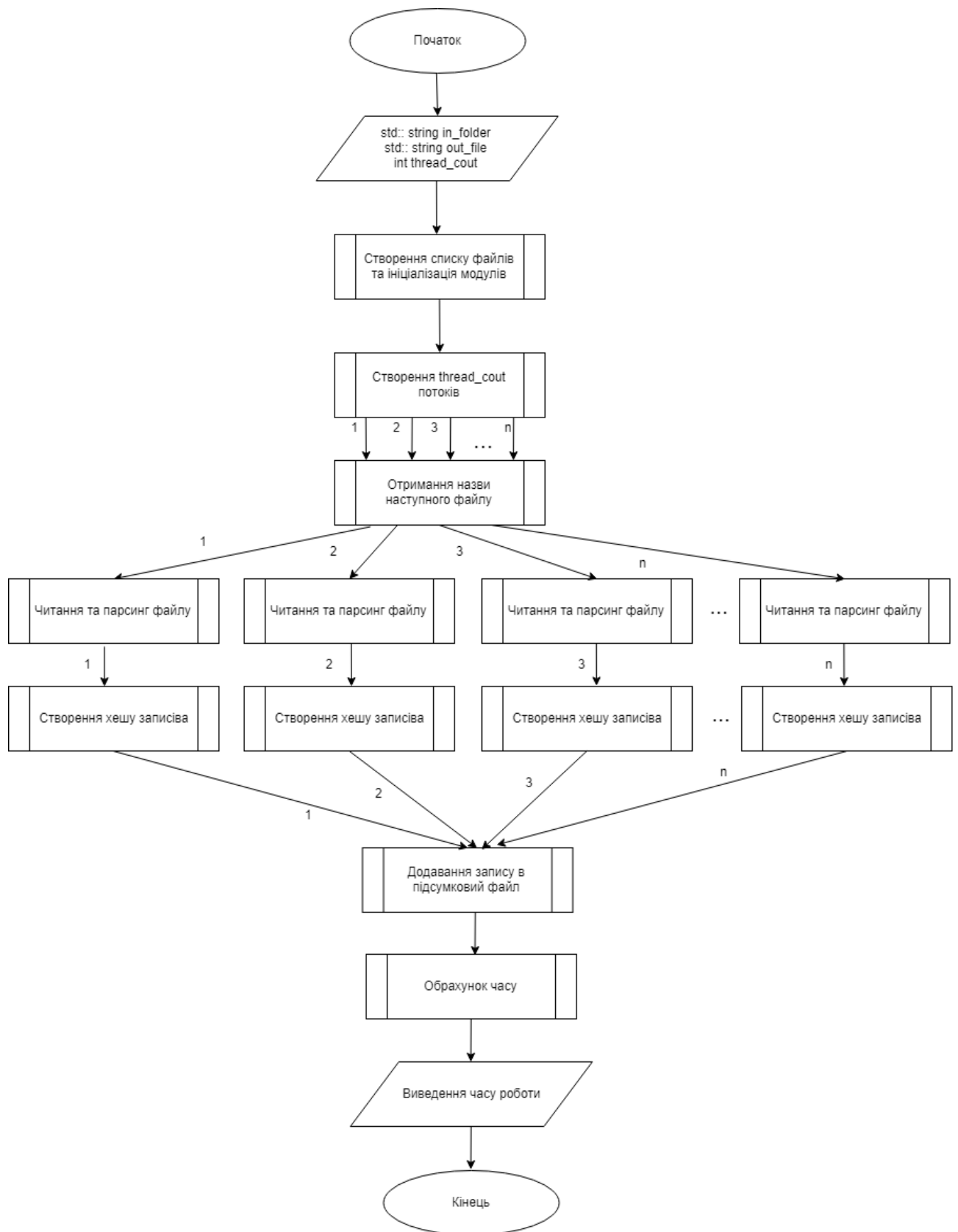


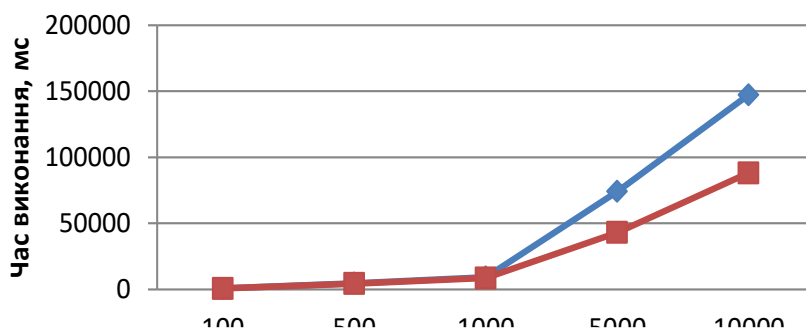
Рисунок 1– Схема паралельного виконання алгоритму обробки текстових даних

Алгоритм пошуку працює таким чином: обирається одне ключове слово – назва поля – і наступне ключове слово – назва наступного поля. Їх порядок визначається заздалегідь при першому проході, тобто з’ясовується порядок подання полів в тексті даної форми. На наступному кроці ведеться пошук

входження першого ключового слова та другого відповідно. Також відома символна довжина назви першого поля, значення якого власне і шукається. Значення поля визначається як текст, що лежить між початком назви першого поля плюс його довжина та початком другого ключового слова [1].

Програмний додаток реалізований таким чином, що вищезгаданий алгоритм пошуку входжень шуканих значень, а також формування хеш-суми та запису кортежу в базу даних можна виконувати в декількох потоках, що надає суттєву перевагу в швидкості обробки. Також програмна реалізація є достатньо гнучкою, щоб забезпечити найбільшу швидкодію на різних системах, адже користувач може самостійно задавати кількість потоків для обробки [2].

В процесі тестування розробленого додатку було проведено дослідження його швидкодії при обробці 100, 500, 1000 і 2000 анкет з використанням однакової кількості потоків в програмі. На діаграмі 1 зображено графіки залежності часу виконання обробки вхідних даних, від кількості анкет, на ній можна спостерігати переваги, що надає використання паралельної обробки.



Діаграма 1 – Порівняння часу виконання паралельної та послідовної програм

Збільшення кількості потоків програми є доцільним в залежності від технічних характеристик комп'ютера, на якому виконується програмний додаток. Тобто використання великої кількості потоків на машині з низькими технічними показниками не дасть переваги в швидкості обробки даних.

### Висновки

Шляхом тестових запусків програмного додатку було встановлено, що створений програмний додаток дозволяє автоматизувати процес обробки текстових даних, а також збільшити швидкодію шляхом використання декількох потоків для обробки.

В подальшому планується розширення функціоналу програмного додатку для вирішення задач підбору персоналу з використанням методів штучного інтелекту [3,4] та нечіткої логіки.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ахо А. Теория синтаксического анализа, перевода и компиляции / А. Ахо, У. Джеффри. – Москва: Мир, 1978. – 243 с.
2. Антонов А. С. Паралельне програмування з використанням технології MPI / Андрій Сергійович Антонов. – Москва: Московський державний університет, 2004. – 71 с.
3. Месюра В. І. Основи проектування систем штучного інтелекту. Навчальний посібник / В. І. Месюра, Л. М. Ваховська. – В.: ВДТУ, 2000. – 96 с.
4. Яровий А. А. Експертні системи. Частина 2 : навчальний посібник / Яровий А. А., Арсенюк І. Р., Месюра В. І. – Вінниця : ВНТУ, 2017. – 106 с.

*Херсоненко Олесь Едуардівна* — студентка групи ІКН-156 факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: 1kn15b.hersonenko@gmail.com

*Месюра Володимир Іванович* — канд. техн. наук, доцент, професор кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця

Науковий керівник: *Месюра Володимир Іванович* — канд. техн. наук, доцент, професор кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця

***Khersonenko Olesya E.*** — Department of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: [1kn15b.hersonenko@gmail.com](mailto:1kn15b.hersonenko@gmail.com)

***Mesyura Volodymyr I.***, Cand. Sc. (Eng.), Professor of the Department of Computer Science, Vinnytsia National Technical University, Vinnytsia, e-mail: [mesyura@vntu.edu.ua](mailto:mesyura@vntu.edu.ua)

Supervisor: ***Mesyura Volodymyr I.*** — Cand. Sc. (Eng.), Professor of the Department of Computer Science, Vinnitsa National Technical University, Vinnytsia