

## Опис алгоритмів вирішення завдання

*Алгоритм рішення задачі виключення грубих помилок  
у спостереженнях*

Нехай  $y = \{y_1, y_2, \dots, y_n\}$  – множина тимчасових інтервалів між натисканням клавіш,  $n$  – кількість тимчасових інтервалів<sup>2</sup>. Необхідно розв'язати задачу перевірки значимості елементів  $y_i, i = \overline{1, n}$  множини  $y$ . Для цього сформуємо множину  $y' = y \setminus y_i$ , тобто всі математичні параметри розраховуються без урахування елемента  $y_i$  у вихідній множині  $y'$ ,  $n'$  – кількість елементів у множині  $y'$ ,  $n' = n - 1$ . Розраховуємо математичне сподівання  $M_i$ :

$$M_i = \frac{\sum_{k=1}^{n'} y'_k}{n'};$$

розраховуємо дисперсію:

$$S_i^2 = \frac{\sum_{k=1}^{n'} (y'_k - M_i)^2}{n' - 1};$$

розраховуємо середньоквадратичне відхилення:

$$S_i = \sqrt{S_i^2}.$$

Розраховуємо коефіцієнт Стюдента<sup>3</sup>  $t_p$ :

$$t_p = \left| \frac{y_i - M_i}{S_i / \sqrt{n'}} \right|.$$

Для числа ступенів свободи  $n' - 1$  і рівня значимості  $\alpha = 0,05$  визначаємо табличний коефіцієнт Стюдента  $t_T$ . Якщо  $t_p > t_T$ , то елемент  $y_i$  – відкидається і розрахунок починається знову до тих пір, поки  $i \neq n$ . В іншому випадку елемент  $y_i$  оголошується значущим.

<sup>2</sup> <https://scienceforum.ru/2016/article/2016029537>

<sup>3</sup> <https://docs.microsoft.com/ru-ru/dotnet/api/system.web.ui.datavisualization.charting.statisticformula?redirectedfrom=MSDN&view=netframework-4.8>

*Алгоритм рішення задачі перевірки гіпотези про рівність центрів розподілу  
двох нормальних генеральних сукупностей*

Вирішуємо задачу перевірки однорідності двох вибірових дисперсій. Нехай  $n$  – число елементів у множині тимчасових інтервалів<sup>2</sup>.

– Вибіркова дисперсія  $S_1$  для еталонних параметрів.

– Вибіркова дисперсія  $S_2$  для множини тимчасових інтервалів, отримана в режимі автентифікації, тоді

$$S_{max}^2 = \max(S_1^2, S_2^2), S_{min}^2 = \min(S_1^2, S_2^2).$$

Розраховуємо коефіцієнт Фішера<sup>4</sup> – розрахунковий  $F_p$  :

$$F_p = \frac{S_{max}^2}{S_{min}^2}.$$

Для кількості ступенів свободи  $n-1$  і рівня значимості  $\alpha=0,05$  визначаємо табличний коефіцієнт Фішера  $F_T$ . Якщо  $F_p > F_T$ , то дисперсії неоднорідні, в іншому випадку дисперсії рівні.

Вирішуємо задачу перевірки гіпотези про рівність центрів розподілу двох нормальних генеральних сукупностей.

Нехай  $y = \{y_1, y_2, \dots, y_n\}$  – множина тимчасових інтервалів між натисканням клавіш отриманих у режимі автентифікації,  $n$  – кількість тимчасових інтервалів. Нехай  $k_e$  – кількість еталонних спроб набору ключової фрази,  $x_\lambda = \{x_1, x_2, \dots, x_n\}$  – множина еталонних часових інтервалів ( $\lambda=1 \dots k_e$ ), тоді

$$S_{x_\lambda}^2 = \frac{\sum_{i=1}^n (x_{\lambda_i} - M_{x_\lambda})^2}{n-1}; S_y^2 = \frac{\sum_{i=1}^n (y_i - M_y)^2}{n-1};$$
$$S = \sqrt{\frac{(S_{x_\lambda}^2 + S_y^2)(n-1)}{2n-1}}; t_p = \frac{|M_{x_\lambda} - M_y|}{S \cdot \sqrt{\frac{2}{n}}}.$$

Тепер необхідно перевірити, чи можна з певною ймовірністю  $p$  вважати, що розбіжність між  $M_{x_\lambda}$  і  $M_y$  викликані випадковими причинами. Для цього по таблиці розподілу Стюдента (рівень значимості  $\alpha = 1-p$ , число ступенів свободи

<sup>4</sup> <https://docs.microsoft.com/ru-ru/archive/msdn-magazine/2015/november/test-run-the-t-test-using-csharp>

$n-1$ ) визначимо табличне значення  $t_T$ , після чого розрахуємо величину  $t_p$ . Якщо  $t_p > t_T$ , то розбіжність не випадкова.

Нехай  $r$  – кількість позитивних рішень задачі, тобто що розбіжність випадкова, тоді  $P = r / k_e$  – оцінка вірогідності, що користувач є автором еталонних характеристик. Необхідно пам'ятати, що з плином часу клавіатурний почерк користувача буде зазнавати змін. У цьому випадку рекомендується після кожної успішної спроби автентифікації зберігати розраховану величину в списку еталонних характеристик користувача.

#### *Розрахунок помилок 1-го та 2-го роду*

На підставі отриманих результатів з перевірки гіпотез, необхідно розрахувати помилки 1-го та 2-го роду.

Помилки 1-го роду визначаються як вірогідність того, що легітимний користувач не буде вірно ідентифікований:

$$P_1 = \frac{N_1}{N_0},$$

де  $N_0$  – загальна кількість спроб пройти процедуру автентифікації легітимним користувачем,  $N_1$  – кількість невдалих спроб.

Помилки 2-го роду визначаються як вірогідність того, що нелегітимний користувач буде ідентифікований як легітимний:

$$P_2 = \frac{N_2}{N_0},$$

де  $N_0$  – загальна кількість спроб пройти процедуру автентифікації нелегітимним користувачем,  $N_2$  – кількість вдалих спроб.

### **3. Блок-схема алгоритму рішення завдання**

У ході проектування та реалізації програмного забезпечення (згідно з завданням даної практичної роботи) розроблено дві блок-схеми, які відповідають режимам роботи цього програмного забезпечення. На рис. 1. Показана блок-схема алгоритму роботи програмного засобу у режимі навчання, а на рис. 2 – блок-схема алгоритму роботи програмного засобу у режимі автентифікації користувачів на основі кодової фрази (клавіатурний почерк).

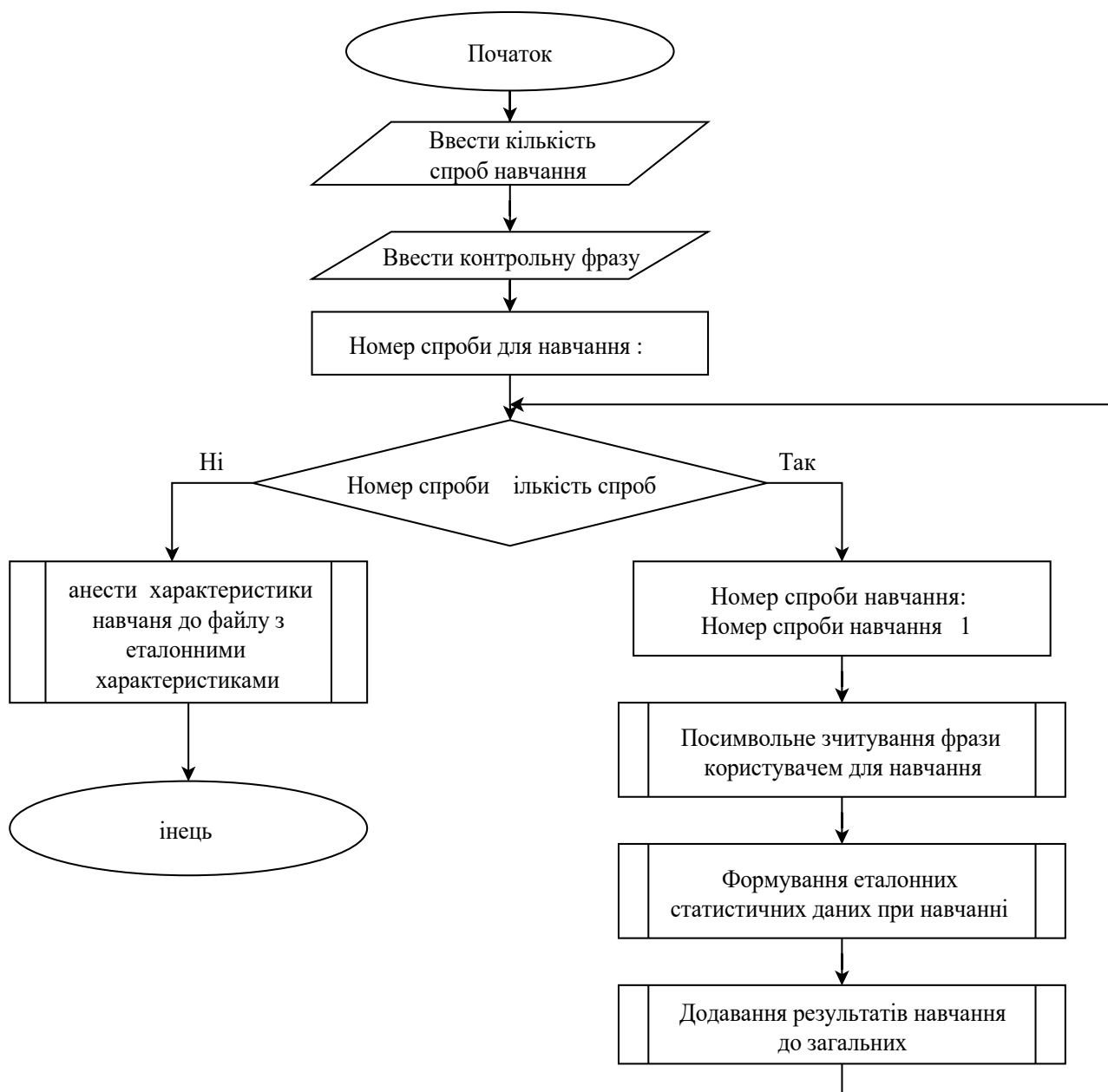


Рис. 1 – Блок-схема алгоритму роботи програмного засобу в режимі навчання

Слід зазначити, що кожна з процедур, які показані блок-схемах на рис. 1 та рис. 2 відповідає за відповідну обробку даних. Процедура посимвольного зчитування формує інтервали затримки між натисканням кнопок на клавіатурі. Процедура формування статистичних даних виконує формування даних про математичне сподівання, дисперсію, критерії Стюдента і Фішера тощо.

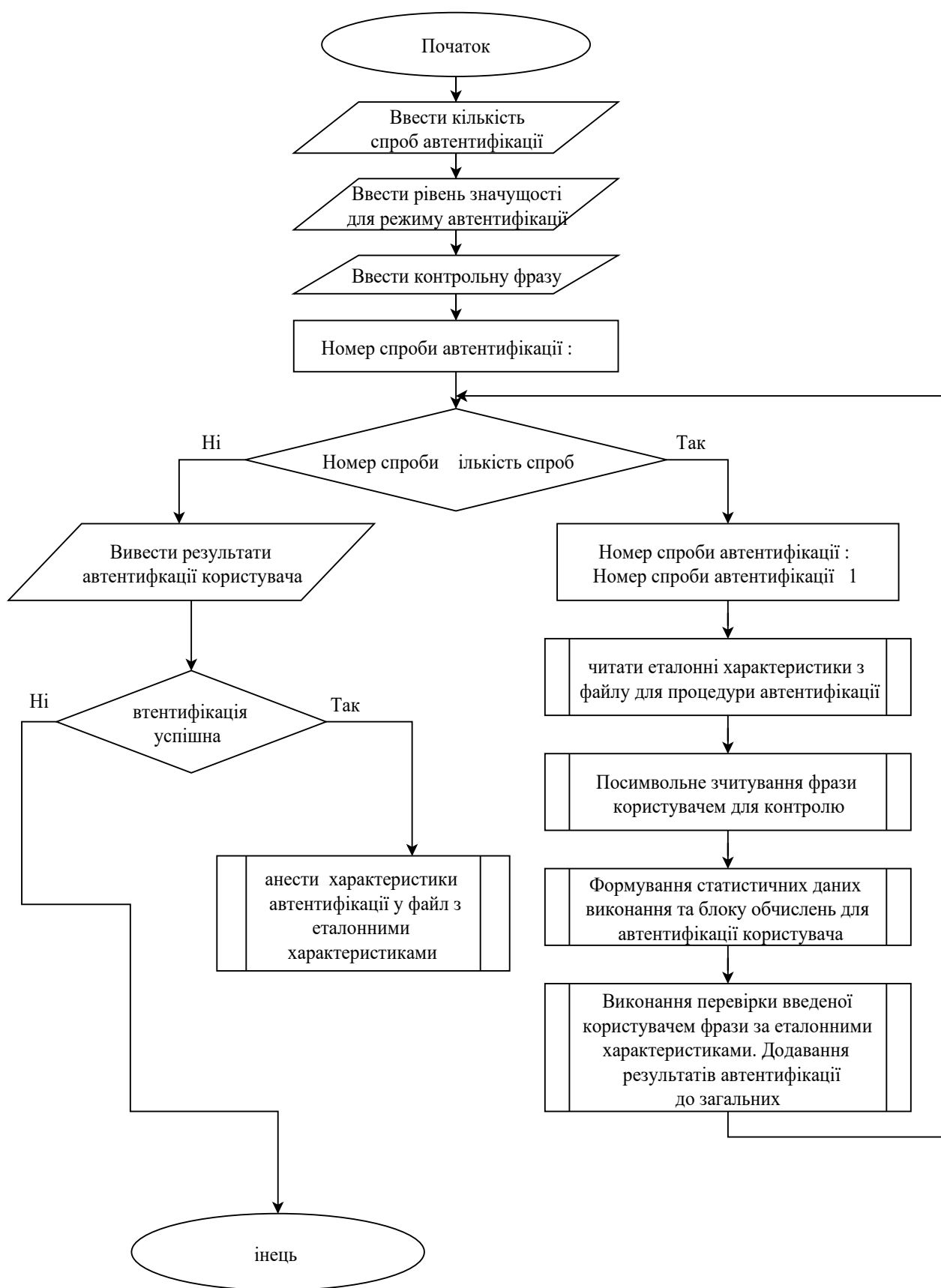


Рис. 2 – Блок-схема алгоритму роботи програмного засобу в режимі перевірки (автентифікації користувача)