

## АЛГОРИТМИ І МОДЕЛІ АВТОМАТИЧНОЇ ІДЕНТИФІКАЦІЇ ТА КОРЕКЦІЇ ТИПОВИХ ПОМИЛОК КОРИСТУВАЧА НА ОСНОВІ ПРИРОДНОЇ НАДМІРНОСТІ

**Abstract:** The algorithms and models of automatic identification and corrections of errors of the user are investigated on the basis of the reference glossary. The quantitative estimations of efficiency of algorithms are received. It is shown, that the considered method and the concrete algorithms of its realization can be used for decreasing of common labour of input of the information in the COMPUTER and increasing the resulting reliability. In particular, the application of a method allows automatically or with the minimal participation of the user to correct from 62 % up to 93 % of typical errors, depending on the chosen ensemble of corrected errors. The recommendations for program realization and practical application of a method are given.

**Key words:** user errors, automatic correction, data reliability.

**Анотація:** Досліджено алгоритми і моделі автоматичної ідентифікації та корекції помилок користувача на основі словника-еталону. Отримані кількісні оцінки ефективності алгоритмів. Показано, що розглянутий метод і конкретні алгоритми його реалізації можуть бути використані для зниження загальної трудомісткості введення інформації в ЕОМ і підвищення результатної вірогідності. Зокрема, застосування методу дозволяє автоматично чи з мінімальною участю користувача виправляти від 62% до 93% типових помилок, у залежності від обраного ансамблю помилок, що коригуються. Дано рекомендації з програмної реалізації і практичного застосування методу.

**Ключові слова:** помилки користувача, автоматична корекція, вірогідність даних.

**Аннотация:** Исследованы алгоритмы и модели автоматической идентификации и коррекции ошибок пользователя на основе словаря-эталона. Получены количественные оценки эффективности алгоритмов. Показано, что рассмотренный метод и конкретные алгоритмы его реализации могут быть использованы для снижения общей трудоемкости ввода информации в ЭВМ и повышения результатной достоверности. В частности, применение метода позволяет автоматически или с минимальным участием пользователя исправлять от 62% до 93% типовых ошибок, в зависимости от выбранного ансамбля корректируемых ошибок. Даны рекомендации по программной реализации и практическому применению метода.

**Ключевые слова:** ошибка пользователя, автоматическая коррекция, достоверность данных.

### 1. Вступ

Одним із загальних напрямків підвищення якості інформаційної бази систем є забезпечення погодженості значень її елементів на всіх етапах життєвого циклу, починаючи з вводу даних. В рамках цього напрямку істотну роль відіграє автоматична ідентифікація і коректування (АІК) типових помилок користувача при вводі інформації (2-й рівень системи контролю погодженості інформації, запропонованої в [1]).

Огляд методів протиперешкодного кодування для захисту інформації від помилок оператора вводу – методів, заснованих на „цілеспрямованій” надмірності, що спеціально вводиться в інформацію, наведено в [2]. Розглянуті в [2] методи потребують попереднього спеціального кодування, порівняно високої додаткової надмірності і не кращим чином пристосовані до більшості специфічних помилок людини, зокрема, до транспозицій, пропусків, доповнень символів (частковим виключенням можна вважати метод Сетхі [3]). Це пояснюється значною мірою тим, що відзначені методи більшою частиною призначені для автокорекції помилок в електронних пристроях та телекомунікаційних каналах зв'язку (і саме в цьому призначенні вони широко і успішно використовуються).

В [4] і [5] розглянуті два часткових варіанти методу АІК помилок користувача на основі так званої „природної” надмірності, що у даному випадку описується словником дозволених слів. Цей

метод, не дивлячись на свою явну концептуальну простоту, уявляється цілком перспективним для ідентифікації і корекції специфічних помилок користувача. Потенціальна перспективність методу обумовлена:

- нульовою додатковою інформаційною надмірністю;
- відсутністю необхідності у попередньому розрахунку кодів;
- можливістю ідентифікації і корекції всіх основних видів специфічних помилок людини.

Мета даної роботи полягає в істотному доповненні і узагальненні часткових результатів [4] і [5] у напрямку порівняльного аналізу всіх можливих стратегій (алгоритмів) вибору слова словника, що є „найближчим” до спотвореного, а також розширенні набору помилок, що коригуються, у деякому сенсі, до повного.

## 2. Загальні положення

### 2.1. Вирішуючі правила АІК за словником

Загальна сутність методу АІК, що розглядається, полягає у варіюванні „зворотних” спотворень помилкового слова передбачуваними помилками, пошуку відповідних найближчих слів у словнику, прийнятті рішення відносно помилки, що відбулася, та у виборі слова для корекції. Уточнимо для даного методу поняття „найближче” слово та пов’язане з ним поняття „відстань”, що загалом є основою введення метрики у просторі помилок.

Як відомо, метрика являє собою правило визначення відстані  $d(AB)$  між точками простору  $A$  і  $B$  з координатами (признаками) відповідно  $a_1 \dots a_n$  та  $b_1 \dots b_n$ , що задовольняють відомим аксіомам. З багатьох різних відомих метрик для рішення традиційних задач автокорекції найбільше значення має метрика з відстанню Хеммінга, яка визначає сумарну кількість розрядів (символів), що не збігаються. Зокрема, для двійкових признаков

$$d_2^x(AB) = \sum_i a_i \oplus b_i,$$

де  $\oplus$  означає додавання по модулю 2.

Для довільних  $q$ -ічних слів ця відстань може бути записана в узагальненому вигляді:

$$d_q^x(AB) = \sum_i \frac{|a_i - b_i|}{|b_i - a_i|},$$

де  $\frac{0}{0} = 0$ .

Як відомо, практичний сенс ефективного вибору міри відстані для контролю-корекції помилок полягає в тому, щоб помилки, які найчастіше зустрічаються, приводили до переходів  $A \rightarrow B$  з меншою відстанню, ніж більш рідкі помилки. Це унеможлиблює появу різних за типом помилок з однаковою відстанню та забезпечує упевнену корекцію з мінімальними хибними „виправленнями” (а вони принципово можливі практично для будь-яких методів автокорекції). У зв’язку з цим відстань типа Хеммінгівської доцільно використовувати для пошуку найближчого слова, коли кратність помилок зворотно залежить від імовірності їх появи, та різниця в значеннях

імовірності помилок різної кратності суттєво велика. Саме ці умови мають місце для статистики випадкових помилок у каналах зв'язку.

У нашому ж випадку для специфічних помилок людини (транскрипцій, транспозицій, доповнень, випадінь тощо) згадані закономірності відсутні. Так, помилки типу пропусків і доповнень символів мають кратність від „1” до „n” в залежності від позиції хибного символу, а імовірність появи такої помилки близька до імовірності появи транспозиції (подвійна помилка) і не на багато менше імовірності одиночної транскрипції [6]. З цієї причини для вирішуючого правила відносно найближчого слова в методі АІК, що розглядається, класичне поняття „відстань” неприйнятне і практично може використовуватися лише одне „крайнє” значення  $d_q^x(AB) = 0$ , тобто повний збіг двох слів.

## 2.2. Основні поняття та визначення

Приймемо такі позначення, визначення та припущення:

$A_x, A_y, B_x$  – відповідно правильне слово вхідного повідомлення, спотворене слово  $A$ , в якому викрита помилка, і деяке з еталонних слів словника, для якого  $d_q^x(A_x B_x) = 0$ . Будемо вважати, що значення  $N$  слів словника в алфавіті  $q$  довжиною  $n_x$  символів випадково-рівномірно розподілені в інтервалі  $0 \div q^{n_{\max}}$ . Це припущення дозволяє при оцінці імовірності збігу довільного слова із словником, слова змінної довжини  $n_x$  інтерпретувати як слова постійної довжини, у яких відкинуті старі незначущі символи (зокрема, для цифрових значень – нулі), доформовуються до максимальної довжини  $n_{\max} = n$ ;

$w_k$  – ранг помилки класу  $E_k$ , що відповідає імовірності появи помилки. Значення імовірностей [6] та рангів для основних класів помилок користувача наведено в табл. 1.

Таблиця 1. Класи помилок

Клас помилки	Характер помилки	Кратність	Імовірність	Ранг $w_k$
$E_1$	Однократна транскрипція	1	$P_1 = 0,5557$	1
$E_2$	Транспозиція	2	$P_2 = 0,0664$	4
$E_3$	Вставка символу	1...n	$P_3 = 0,1567$	2
$E_4$	Пропуск символу	1...n	$P_4 = 0,1204$	3
$E_5$	Двократна транскрипція	2	$P_5 = 0,0322$	5
$E_6$	Багатократна транскрипція	3...n	$P_6 = 0,0686$	6

Варіація  $\tilde{A}^k$  – деяка зміна слова  $A$  в межах  $k$ -го класу помилок. Наприклад, в класі  $E_1$  варіаціями слова 5731 є слова 0731, 1731, ..., 5031, 5131 ... і т.д. Кількість варіацій  $V_k^y$  слова  $A_y$  з довжиною  $n_y$  символів визначається, як легко показати, таким чином:

$$\begin{aligned}
V_1^y &= (q-1)n_y; & V_4^y &= q(n_y+1); \\
V_2^y &= n_y-1; & V_5^y &= (q-1)^2 C_{n_y}^2 - n_y + 1; \\
V_3^y &= qn_y; & V_6^y &= \sum_{t=3}^{n_y} (q-1)^t C_{n_y}^t.
\end{aligned}$$

Для довільного слова з довжиною в діапазоні  $n_{\min} \div n$  кількість варіацій  $V_k$  визначимо як  $V_k = V_k^y$  при умові  $n_y = \text{ent}[n_{cp}]$ , де  $\text{ent}[\cdot]$  означає ціле, найближче до середнього значення  $n_{cp}$ .

Ансамбль  $K(k_j)$  – сукупність класів помилок, що коригуються і для яких генеруються варіації  $\tilde{A}^k$ .

Значність помилок  $Z$  – кількість слів словника, для яких виконується  $d_q^x(\tilde{A}^k B_x) = 0$ .

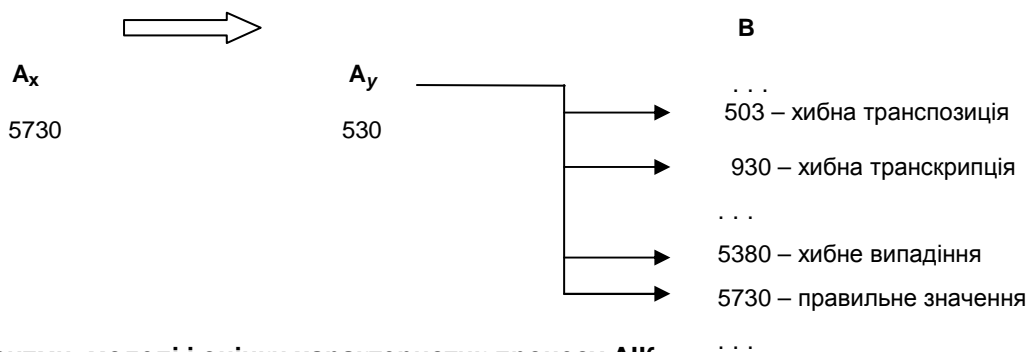
Справедливі такі твердження.

Твердження 1. Для будь-якої помилки  $A_x \rightarrow A_y$ , що належить ансамблю  $K(k_j)$ , значність  $z > 0$ .

Це означає, що для такої помилки існує по меншій мірі одне слово  $B_x$ , для якого  $d_q^x(A^k B_x) = 0$ .

Для помилки, що не належить  $K(k_j)$ , цього може і не бути. Твердження цілком очевидне: якщо відбувся помилковий перехід  $A_x \rightarrow A_y$  з класу  $k$ , для якого генерується  $\tilde{A}^k$ , то одне із зворотних спотворень  $A_y$  обов'язково збігається з істинним  $A_x$ , що, природно, має бути в словнику.

Твердження 2. Для будь-якої помилки  $A_x \rightarrow A_y$  можливі значення  $z > 1$ , і, відповідно, хибна ідентифікація. Твердження легко доводиться від супротивного шляхом побудови контрприкладів для кожної пари класів помилок – істинної та хибної. Як ілюстрацію наведено такий контрприклад для  $k = (1,2,3)$



### 3. Алгоритми, моделі і оцінки характеристик процесу АІК

Загальний алгоритм ідентифікації – корекції помилки, що була виявлена на етапі контролю, полягає в послідовній генерації варіацій  $\tilde{A}^k$  ансамблю  $K(k_j)$  у порядку збільшення рангу помилок,

перевірки виконання для кожної з них умови  $d_q^x(A^k B_x) = 0$  для кожного із слів словника  $i$ , у випадку позитивного результату, прийняття відповідного рішення щодо подальших дій. Конкретні алгоритми визначаються саме цими рішеннями, проте незалежно від них у кінцевому підсумку ми маємо один із наступних загальних результатів – подій:

$R_{AK}$ ,  $R_{IAK}$  – помилка правильно ідентифікована і автоматично (імовірність  $P_{AK}$ ) або напівавтоматично (імовірність  $P_{IAK}$ ) виправлена;

$R_{PK}$  – ідентифікувати помилку не вдається і вона виправляється „вручну” (імовірність  $P_{PK}$ );

$R_{XK}$  – помилка ідентифікована невірно і виправлена хибно (імовірність  $P_{XK}$ ).

Повна група всіляких подій включає ще випадок, коли на етапі контролю слова  $A_x$  по словнику, що передує етапу ідентифікації - корекції, помилка не виявляється (імовірність  $P_{HB}$ ).

Розглянемо рішення, що можуть прийматися при збігу чергової варіації з деяким словом словника.

### 3.1. Алгоритм 1

При першому ж збігу однієї з  $V = \sum_k V_k$  варіацій  $\tilde{A}^k$  із словом  $B_x$  процес генерації і порівнянь припиняється, здійснюється автоматичне виправлення слова  $A_y$  по слову  $B_x$ . Граф зв'язку варіантів часткових подій з загальними результатами наведено на рис. 1.

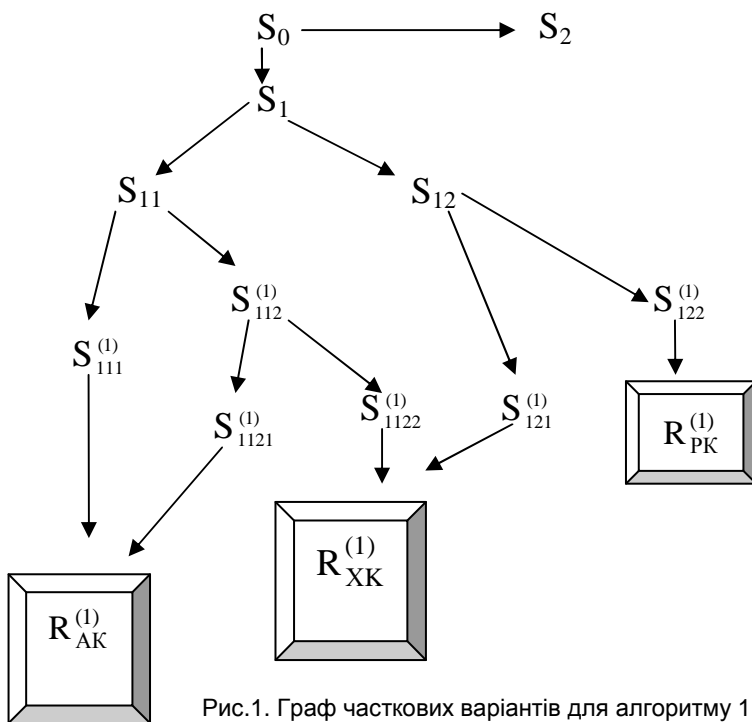


Рис.1. Граф часткових варіантів для алгоритму 1

Прийняті такі позначення подій:

$S_0$  – помилка відбулася;

$S_1$  – помилка виявлена;

$S_2 = \bar{S}_1$  – помилка не виявлена;

$S_{11}$  – помилка належить ансамблю  $K(k_j)$ ;

$S_{12} = \bar{S}_{11}$  – помилка не належить ансамблю  $K(k_j)$ ;

$S_{111}^{(1)}$  – помилка однозначна ( $z = 1$ );

$S_{112}^{(1)} = \bar{S}_{111}^{(1)}$  – помилка

багатозначна ( $z > 1$ );

$S_{122}^{(1)}$  – для помилки, що не належить ансамблю  $K(k_j)$ , не відбулося жодного випадкового збігу варіацій із словом словника ( $z = 0$ );

$S_{121}^{(1)} = \bar{S}_{122}$  – для помилки, що не належить ансамблю  $K(k_j)$ , значність  $z \geq 1$ ;

$S_{1121}^{(1)}$  – фактичній багатозначній помилці відповідає перше з  $g$  додаткових випадкових збігів ( $g = z - 1$ );

$S_{1122}^{(1)} = \bar{S}_{1121}$ .

З урахуванням наведених означень і вигляду графу рис. 1 можна записати такі логічні вирази:

$$R_{AK}^{(1)} = (S_1 \wedge S_{11} \wedge S_{111}^{(1)}) \vee (S_1 \wedge S_{11} \wedge \bar{S}_{111}^{(1)} \wedge S_{1121}^{(1)}) = [S_1 \wedge S_{11}] \wedge [S_{111}^{(1)} \vee (\bar{S}_{111}^{(1)} \wedge S_{1121}^{(1)})];$$

$$R_{PK}^{(1)} = S_1 \wedge \bar{S}_{11} \wedge S_{122}^{(1)};$$

$$R_{XK}^{(1)} = S_1 \wedge [(\bar{S}_{11} \wedge \bar{S}_{122}^{(1)}) \vee (S_{11} \wedge \bar{S}_{111}^{(1)} \wedge \bar{S}_{1121}^{(1)})].$$

Вважаючи чисто випадковим характер процесу збігу варіацій із словами словника і враховуючи у зв'язку з цим, що імовірність  $P(g, r, V)$  в точності  $g$  випадкових збігів  $V$  варіацій визначається відомою формулою біноміального розподілу

$$P(g, r, V) = C_V^g r^g (1-r)^{V-g},$$

де  $r$  – імовірність збігу довільної варіації, для значень  $P_{AK}^{(1)}$ ,  $P_{PK}^{(1)}$ ,  $P_{XK}^{(1)}$ , можемо записати такі остаточні вирази:

$$P_{AK} = \left[ (1-r) \cdot \sum_j P_j \right] \cdot \left\{ (1-r)^{V-1} + [1 - (1-r)^{V-1}] \cdot \frac{\sum_{g=1}^{V-1} \frac{1}{g+1} P(g, r, V-1)}{\sum_{g=1}^{V-1} P(g, r, V-1)} \right\}, \quad (1)$$

$$P_{PK} = (1-r) \cdot \left( 1 - \sum_j P_j \right) (1-r)^V \quad (2)$$

$$P_{XK} = (1-r) \cdot \left\{ \left( 1 - \sum_j P_j \right) [1 - (1-r)^V] + \sum_j P_j [1 - (1-r)^{V-1}] \cdot \left[ 1 - \frac{\sum_{g=1}^{V-1} \frac{1}{g+1} P(g, r, V-1)}{\sum_{g=1}^{V-1} P(g, r, V-1)} \right] \right\}, \quad (3)$$

де  $\sum_j P_j$  – сумарна імовірність появи помилок ансамблю  $K(k_j)$ .

При одержанні останніх рівнянь враховано, що, якщо помилка належить до ансамблю  $K(k_j)$ , то, у відповідності з Твердженням 1, одна з  $V$  варіацій цілком визначено збігається із словником. При цьому ще  $g$  варіацій ( $g = 1, \dots, V-1$ ) можуть дати хибні збіги. Якщо помилка не належить до ансамблю  $K(k_j)$ , то, у відповідності з Твердженням 2, можливі  $1, \dots, V$  випадкових хибних збігів.

З урахуванням припущення п. 2.2 відносно рівномірного розподілу  $N$  слів словника в інтервалі  $0 \div q^n$  величина  $r$  дорівнює  $r = \frac{N}{q^n}$ .

В табл. 2 наведено розрахункові значення  $P_{AK}, P_{PK}, P_{XK}$  для ансамблів  $K_1(k_j) = (1,4)$ ,  $K_2(k_j) = (1,2,3,4)$ ,  $K_3(k_j) = (1,2,3,4,5)$ . Значення розраховано для  $q = 10$ ,  $n_{cp} = 6$ ; при цьому  $V(1,4) = V_1 + V_4 = 59$ ,  $V(1,2,3,4) = 135$ ,  $V(1,2,3,4,5) = 1345$ . Відповідно  $P_{\Sigma} = 0,6221$ ,  $0,8992$  та  $0,9314$ .

Таблиця 2. Результати розрахунків для алгоритму 1

$K(k_j)$	$r$	$P_{AK}$	$P_{PK}$	$P_{XK}$	$\delta$	$\frac{P_{XK}}{r}$
$K_1$	$10^{-2}$	$4,6694 \cdot 10^{-1}$	$2,0677 \cdot 10^{-1}$	$3,1629 \cdot 10^{-1}$	$3,6228 \cdot 10^{+1}$	$3,1629E+1$
	$10^{-3}$	$6,0379 \cdot 10^{-1}$	$3,5588 \cdot 10^{-1}$	$3,9325 \cdot 10^{-2}$	$4,7942 \cdot 10^{+0}$	$3,9325 \cdot 10^{+1}$
	$10^{-4}$	$6,2024 \cdot 10^{-1}$	$3,7564 \cdot 10^{-1}$	$4,0234 \cdot 10^{-3}$	$8,2932 \cdot 10^{-1}$	$4,0234 \cdot 10^{+1}$
	$10^{-5}$	$6,2191 \cdot 10^{-1}$	$3,7767 \cdot 10^{-1}$	$4,0327 \cdot 10^{-4}$	$4,2314 \cdot 10^{-1}$	$4,0327 \cdot 10^{+1}$
	$10^{-6}$	$6,2208 \cdot 10^{-1}$	$3,7788 \cdot 10^{-1}$	$4,0336 \cdot 10^{-5}$	$3,8243 \cdot 10^{-1}$	$4,0336 \cdot 10^{+1}$
$K_2$	$10^{-2}$	$3,8680 \cdot 10^{-1}$	$1,3505 \cdot 10^{-2}$	$5,8969 \cdot 10^{-1}$	$6,7167 \cdot 10^{+1}$	$5,8969 \cdot 10^{+1}$
	$10^{-3}$	$8,1493 \cdot 10^{-1}$	$8,2520 \cdot 10^{-2}$	$1,0155 \cdot 10^{-1}$	$1,1542 \cdot 10^{+1}$	$1,0155 \cdot 10^{+2}$
	$10^{-4}$	$8,9027 \cdot 10^{-1}$	$9,8804 \cdot 10^{-2}$	$1,0829 \cdot 10^{-2}$	$1,3198 \cdot 10^{+0}$	$1,0829 \cdot 10^{+2}$
	$10^{-5}$	$8,9830 \cdot 10^{-1}$	$1,0060 \cdot 10^{-1}$	$1,0900 \cdot 10^{-3}$	$2,2349 \cdot 10^{-1}$	$1,0900 \cdot 10^{+2}$
	$10^{-6}$	$8,9911 \cdot 10^{-1}$	$1,0078 \cdot 10^{-1}$	$1,0907 \cdot 10^{-4}$	$1,1308 \cdot 10^{-1}$	$1,0907 \cdot 10^{+2}$
$K_3$	$10^{-2}$	$5,4919 \cdot 10^{-2}$	$3,1875 \cdot 10^{-9}$	$9,3508 \cdot 10^{-1}$	$1,0649 \cdot 10^{+2}$	$9,3508 \cdot 10^{+1}$
	$10^{-3}$	$4,5088 \cdot 10^{-1}$	$1,2775 \cdot 10^{-2}$	$5,3535 \cdot 10^{-1}$	$6,0428 \cdot 10^{+1}$	$5,3535 \cdot 10^{+2}$
	$10^{-4}$	$8,5736 \cdot 10^{-1}$	$5,7991 \cdot 10^{-2}$	$8,4548 \cdot 10^{-2}$	$9,5909 \cdot 10^{+0}$	$8,4548 \cdot 10^{+2}$
	$10^{-5}$	$9,2362 \cdot 10^{-1}$	$6,7457 \cdot 10^{-2}$	$8,9130 \cdot 10^{-3}$	$1,0723 \cdot 10^{+0}$	$8,9130 \cdot 10^{+2}$
	$10^{-6}$	$9,3062 \cdot 10^{-1}$	$6,8485 \cdot 10^{-2}$	$8,9609 \cdot 10^{-4}$	$1,6951 \cdot 10^{-1}$	$8,9609 \cdot 10^{+2}$

Оцінюючи характеристики алгоритмів (того, що розглядається і наступних) додатково приймемо до уваги, що окрім імовірності хибної корекції  $P_{XK}$ , притаманної власне алгоритму, існує також імовірність хибного виправлення користувачем помилки і при „ручній” корекції. Позначимо відносну кількість хибних ручних корекцій через  $\pi_{XPK}$ . Для  $K_1(k_j) = 1,4$  при  $r = 10^{-5}$  відносна кількість помилок, що правильно автоматично виправляється, складає, як видно з табл. 2,  $P_{AK} = 0,6220$ . Відповідна кількість помилок, що виправлялися б вручну без АІК, дали б внесок у загальну кількість невиправлених помилок, який дорівнює  $\pi_{XPK} \cdot 0,6220$ . Якщо припустити, що при ручному коректуванні виправляється 1 символ і імовірність викривити цей символ дорівнює статистичній імовірності викривлення символу при вводі інформації  $\pi_c = 8,87 \cdot 10^{-3}$  [6], то при відмові від застосування АІК додаткова імовірність проникнення помилок через „бар’єр” контролю складала б  $8,87 \cdot 10^{-3} \cdot 0,6220 = 55 \cdot 10^{-4}$ . В той же час імовірність хибної корекції, що супроводить згадане значення  $P_{AK}$ , дорівнює всього  $4 \cdot 10^{-4}$ , тобто на порядок менше.

Узагальнюючи ці міркування, введемо до розгляду такий показник:

$$\delta = \frac{P_{XK} + \pi_{XPK} \cdot P_{PK}}{\pi_{XPK} \cdot (1 - r)}. \quad (4)$$

Чисельник відношення (4) являє собою апостеріорну імовірність проникнення помилки через систему AIK, а знаменник – імовірність такої ж події при звичайному контролі і ручній корекції. Значення  $r_0$ , що відповідає рівності  $\delta = 1$ , назвемо точкою рівноваги. З наведених міркувань можна зробити висновок, що для  $r < r_0$  і відповідно  $\delta < 1$  застосування методу AIK не тільки зменшує трудомісткість вводу, а й підвищує підсумкову достовірність.

В табл. 2 наведено значення  $\delta$ , отримані в припущенні  $\pi_{XPK} = 8,87 \cdot 10^{-3}$ . Значення точки рівноваги для ансамблів  $K_1$ ,  $K_2$ ,  $K_3$ , одержані сумісним чисельним рішенням (2) – (4), складають відповідно  $1,36 \cdot 10^{-4}$ ;  $7,4 \cdot 10^{-5}$ ;  $9,25 \cdot 10^{-6}$ .

З формул (1), (2), (3), що проілюстровані даними табл. 2, видно, що всі залежності  $P(r)$  монотонні, причому для  $r \rightarrow 0$   $P_{AK} \rightarrow P_{\Sigma}$ ,  $P_{PR} \rightarrow (1 - P_{\Sigma})$ ,  $P_{HB} \rightarrow 0$ .

Величина  $P_{XK}$  також  $\rightarrow 0$ , але повільніше, ніж  $P_{HB}$ . Так,  $\lim_{r \rightarrow 0} \frac{P_{XK}}{r}$  для ансамблю  $K_1 \approx 40,5$ , для ансамблю  $K_2 \approx 109$ , для ансамблю  $K_3 \approx 8,96 \cdot 10^2$ . Аналізуючи сукупність часткових подій  $S$  для розглянутого алгоритму, можна відзначити, що сумарна частка помилок, що хибно коригуються, містить дві складові:

- частина помилок, що не коригуються і не належать ансамблю  $K_1(k_j)$  і які дають випадкові збіги варіацій із словником;
- частина помилок ансамблю  $K_1(k_j)$  при наявності власної багатозначності ( $z > 1$ ), що приводить до хибних збігів по визначенню.

Позбутися від першої складової принципово неможливо. Зменшення ж частки помилок, що не коригуються, (перехід до ансамблів  $K_2$  і  $K_3$ ) дає у підсумку, як видно з табл. 2, всупереч можливим очікуванням, зворотний результат за рахунок ще більшого підвищення другої складової.

На повне виключення другої складової спрямований наступний алгоритм (для часткового випадку  $K_1(k_j) = K_1$  і  $n_x = \text{const}$  розглянутий в [4] – на жаль, з деякими допущеними неточностями).

### 3.2. Алгоритм 2

Процес генерації варіацій ансамблю  $K_1(k_j)$  та сканування словника не зупиняється при першому ж збігу, а продовжується до встановлення факту наявності або ж відсутності однозначності. У випадку  $z = 1$  здійснюється автоматична корекція слова  $A$  за відповідною варіацією, якщо  $z = 0$  або  $z > 1$ , то автоматична корекція не виконується і спотворене слово виправляється вручну.

Граф варіантів часткових подій подано на рис. 2.



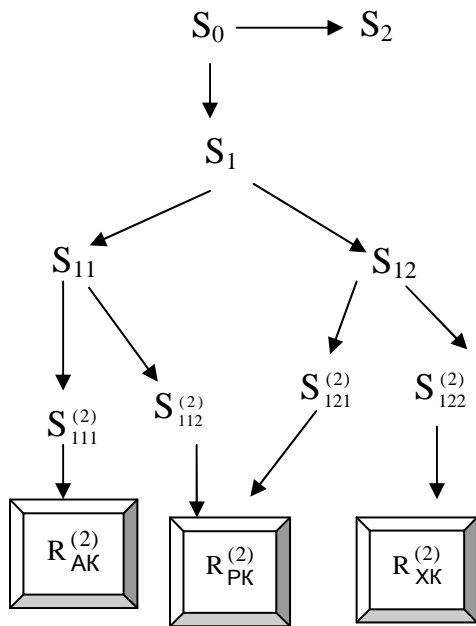


Рис. 2. Граф часткових варіантів для алгоритму 2

У доповнення до загальних для всіх алгоритмів позначень подій  $S_1, S_2 = \bar{S}_1, S_{11}, S_{12} = \bar{S}_{11}$  прийняті такі позначення:

$S_{111}^{(2)}$  – помилка з ансамблю  $K_1(k_j)$  однозначна;

$S_{112}^{(2)} = \bar{S}_{111}^{(2)}$  ;

$\bar{S}_{122}^{(2)}$  – помилка, що не належить ансамблю

$K(k_j)$ , однозначна;

$S_{121}^{(2)} = \bar{S}_{122}^{(2)}$  .

Як впливає з рис. 2,

$R_{AK}^{(2)} = S_1 \wedge S_{11} \wedge S_{111}^{(2)}$ ;

$R_{PK}^{(2)} = S_1 \wedge [(S_{11} \wedge \bar{S}_{111}^{(2)}) \vee (\bar{S}_{11} \wedge \bar{S}_{122}^{(2)})]$ ;

$R_{XK}^{(2)} = S_1 \wedge \bar{S}_{11} \wedge \bar{S}_{122}^{(2)}$  .

Звідси, з урахуванням колишніх припущень,

$$P_{AK}^{(2)} = (1-r) \cdot P_{\Sigma} \cdot (1-r)^{V-1}; \quad (5)$$

$$P_{PK}^{(2)} = (1-r) \cdot \left\{ P_{\Sigma} \left[ 1 - (1-r)^{V-1} + (1-P_{\Sigma}) \cdot \left[ 1 - V r (1-r)^V \right] \right] \right\}; \quad (6)$$

$$P_{XK}^{(2)} = (1-r)(1-P_{\Sigma}) \cdot V_r (1-r)^V. \quad (7)$$

Таблиця 3. Результати розрахунків для алгоритму 2

$K(k_j)$	$r$	$P_{AK}$	$P_{PK}$	$P_{XK}$	$\delta$	$P_{XK}/r$
$K_1$	$10^{-2}$	$3,4382 \cdot 10^{-1}$	$5,2418 \cdot 10^{-1}$	$1,2199 \cdot 10^{-1}$	$1,4422 \cdot 10^{+1}$	$1,2199 \cdot 10^{+1}$
	$10^{-3}$	$5,8644 \cdot 10^{-1}$	$3,9156 \cdot 10^{-1}$	$2,0997 \cdot 10^{-2}$	$2,7615 \cdot 10^{+0}$	$2,0997 \cdot 10^{+1}$
	$10^{-4}$	$6,1844 \cdot 10^{-1}$	$3,7924 \cdot 10^{-1}$	$2,2163 \cdot 10^{-3}$	$6,2917 \cdot 10^{-1}$	$2,2163 \cdot 10^{+1}$
	$10^{-5}$	$6,2173 \cdot 10^{-1}$	$3,7803 \cdot 10^{-1}$	$2,2283 \cdot 10^{-4}$	$4,0316 \cdot 10^{-1}$	$2,2283 \cdot 10^{+1}$
	$10^{-6}$	$6,2206 \cdot 10^{-1}$	$3,7791 \cdot 10^{-1}$	$2,2295 \cdot 10^{-5}$	$3,8043 \cdot 10^{-1}$	$2,2295 \cdot 10^{+1}$
$K_2$	$10^{-2}$	$1,2169 \cdot 10^{-1}$	$8,4143 \cdot 10^{-1}$	$2,6875 \cdot 10^{-2}$	$3,9104 \cdot 10^{+0}$	$2,6875 \cdot 10^{+0}$
	$10^{-3}$	$7,3687 \cdot 10^{-1}$	$2,4571 \cdot 10^{-1}$	$1,6421 \cdot 10^{-2}$	$2,0992 \cdot 10^{+0}$	$1,6421 \cdot 10^{+1}$
	$10^{-4}$	$8,8148 \cdot 10^{-1}$	$1,1645 \cdot 10^{-1}$	$1,9662 \cdot 10^{-3}$	$3,3815 \cdot 10^{-1}$	$1,9662 \cdot 10^{+1}$
	$10^{-5}$	$8,9741 \cdot 10^{-1}$	$1,0238 \cdot 10^{-1}$	$2,0019 \cdot 10^{-4}$	$1,2495 \cdot 10^{-1}$	$2,0019 \cdot 10^{+1}$
	$10^{-6}$	$8,9902 \cdot 10^{-1}$	$1,0096 \cdot 10^{-1}$	$2,0055 \cdot 10^{-5}$	$1,0322 \cdot 10^{-1}$	$2,0055 \cdot 10^{+1}$
$K_3$	$10^{-2}$	$4,3715 \cdot 10^{-8}$	$9,9000 \cdot 10^{-1}$	$5,3518 \cdot 10^{-8}$	$1,0000 \cdot 10^{+0}$	$5,3518 \cdot 10^{-6}$
	$10^{-3}$	$1,7362 \cdot 10^{-1}$	$8,0394 \cdot 10^{-1}$	$2,1448 \cdot 10^{-2}$	$3,2252 \cdot 10^{+0}$	$2,1448 \cdot 10^{+1}$
	$10^{-4}$	$7,8743 \cdot 10^{-1}$	$2,0273 \cdot 10^{-1}$	$9,7367 \cdot 10^{-3}$	$1,3006 \cdot 10^{+0}$	$9,7367 \cdot 10^{+1}$
	$10^{-5}$	$9,1589 \cdot 10^{-1}$	$8,2965 \cdot 10^{-2}$	$1,1326 \cdot 10^{-3}$	$2,1066 \cdot 10^{-1}$	$1,1326 \cdot 10^{+2}$
	$10^{-6}$	$9,2984 \cdot 10^{-1}$	$7,0047 \cdot 10^{-2}$	$1,1499 \cdot 10^{-4}$	$8,3010 \cdot 10^{-2}$	$1,1499 \cdot 10^{+2}$

Як можна бачити з даних табл. 3, що відбиває результати розрахунків для умов, визначених в п. 3.1, значення  $R_{\text{ХК}}^{(2)}$  менше, а  $R_{\text{АК}}^{(2)}$  більше, ніж відповідні значення табл. 2 (як і потрібно було чекати). Зокрема, для  $r \rightarrow 0$  відношення  $\frac{P_{\text{ХК}}}{r} \rightarrow 22,3$  для  $K_1$ ,  $\rightarrow 20$  для  $K_2$  та  $\rightarrow 114$  для  $K_3$  (має місце деякий оптимум, що, втім, не має суттєвого значення). В загальному випадку

$$\lim_{r \rightarrow 0} \frac{P_{\text{ХК}}}{r} = \frac{(1-r)(1-P_{\Sigma})Vr(1-r)^V}{r} = V(1-P_{\Sigma}).$$

Меншими для зіставних значень  $r \in$  і відповідні значення  $\delta$ . Що стосується значень  $r_0$ , то для  $K_1$   $r_0 = 2,49 \cdot 10^{-4}$ , для  $K_2$   $r_0 = 4,01 \cdot 10^{-4}$ , для  $K_3$   $r_0 = 7,2 \cdot 10^{-5}$ , тобто точки рівноваги досягаються при більших значеннях  $r$  (і це є, безумовно, позитивним моментом, бо більшим значенням  $r$  відповідає менша потрібна інформаційна надмірність). Разом з тим слід відзначити, що переваги, які одержуються у порівнянні з алгоритмом 1, дістаються ціною приблизно подвійного (у середньому) збільшення кількості варіацій, що генеруються і аналізуються.

### 3.3. Алгоритм 3

Даний алгоритм спрямований на використання можливостей діалогового режиму вводу, а саме можливості напівавтоматичної корекції, в процесі якої користувач підтверджує або не підтверджує корекцію, що йому пропонується (характеристики алгоритму для часткового випадку  $K(k_j) = K_1$ ,  $n_x = \text{const}$  розглянуті в [5]). В цьому випадку, при умові цілком безпомилкових дій користувача, можливість появи випадків хибної корекції була б виключена зовсім, – хоча і ціною додаткового навантаження на користувача у порівнянні з алгоритмами 1, 2. Граф варіантів часткових подій подано на рис. 3.

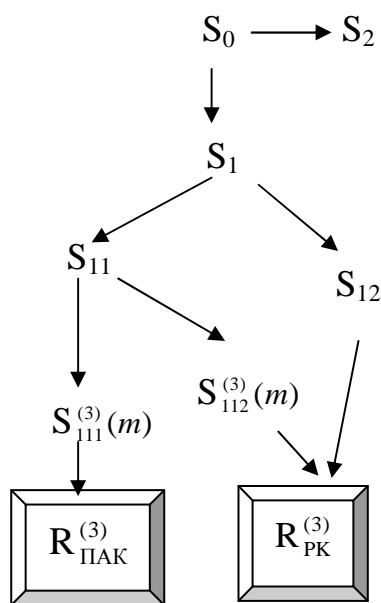


Рис. 3. Граф часткових варіантів для алгоритму 3

Прийняті такі позначення:

$S_{111}^{(3)}$  – помилку ідентифіковано користувачем за  $m$  спроб;

$$S_{112}^{(3)} = \bar{S}_{111}^{(3)}(m).$$

Для підсумкових результатів маємо:

$$R_{\text{ПАК}}^{(3)} = S_1 \wedge S_{11} \wedge S_{111}^{(3)}(m),$$

$$R_{\text{ПК}}^{(3)} = S_1 \wedge \bar{S}_{11} \wedge \bar{S}_{111}^{(3)}(m).$$

Звідси

$$P_{\text{ПАК}}^{(3)} = (1-r) \cdot P_{\Sigma} \cdot \pi(m), \tag{8}$$

$$P_{\text{ПК}}^3 = (1-r)(1-P_{\Sigma})[1-\pi(m)], \tag{9}$$

де  $\pi(m)$  дорівнює [5]:

$$\pi(m) = \sum_{g=0}^{m-1} C_{V-1}^g r^g (1-r)^{V-g-1} + \sum_{g=m}^{V-1} \frac{m}{g+1} C_{V-1}^g r^g (1-r)^{V-g-1}. \quad (10)$$

Результати відповідних розрахунків для  $m = 2$  наведено в табл. 4.

Таблиця 4. Результати розрахунків для алгоритму 3

$K(k_j)$	$r$	$P_{ПАК}$	$P_{ПК}$	$\delta(0,5)$	$\delta(0,167)$
$K_1$	$10^{-2}$	$5,900 \cdot 10^{-1}$	$3,9995 \cdot 10^{-1}$	$7,0200 \cdot 10^{-1}$	$5,0352 \cdot 10^{-1}$
	$10^{-3}$	$6,2114 \cdot 10^{-1}$	$3,7786 \cdot 10^{-1}$	$6,8912 \cdot 10^{-1}$	$4,8207 \cdot 10^{-1}$
	$10^{-4}$	$6,2203 \cdot 10^{-1}$	$3,7787 \cdot 10^{-1}$	$6,8895 \cdot 10^{-1}$	$4,8179 \cdot 10^{-1}$
	$10^{-5}$	$6,2209 \cdot 10^{-1}$	$3,7790 \cdot 10^{-1}$	$6,8895 \cdot 10^{-1}$	$4,8179 \cdot 10^{-1}$
	$10^{-6}$	$6,2210 \cdot 10^{-1}$	$3,7790 \cdot 10^{-1}$	$6,8895 \cdot 10^{-1}$	$4,8179 \cdot 10^{-1}$
$K_2$	$10^{-2}$	$6,5191 \cdot 10^{-1}$	$3,3809 \cdot 10^{-1}$	$6,7075 \cdot 10^{-1}$	$4,5147 \cdot 10^{-1}$
	$10^{-3}$	$8,9300 \cdot 10^{-1}$	$1,0600 \cdot 10^{-1}$	$5,5305 \cdot 10^{-1}$	$2,5539 \cdot 10^{-1}$
	$10^{-4}$	$8,9905 \cdot 10^{-1}$	$1,0085 \cdot 10^{-1}$	$5,5043 \cdot 10^{-1}$	$2,5101 \cdot 10^{-1}$
	$10^{-5}$	$8,9919 \cdot 10^{-1}$	$1,0080 \cdot 10^{-1}$	$5,5040 \cdot 10^{-1}$	$2,5097 \cdot 10^{-1}$
	$10^{-6}$	$8,9920 \cdot 10^{-1}$	$1,0080 \cdot 10^{-1}$	$5,5040 \cdot 10^{-1}$	$2,5097 \cdot 10^{-1}$
$K_3$	$10^{-2}$	$1,0984 \cdot 10^{-1}$	$8,8016 \cdot 10^{-1}$	$9,4453 \cdot 10^{-1}$	$9,0758 \cdot 10^{-1}$
	$10^{-3}$	$7,2814 \cdot 10^{-1}$	$2,7086 \cdot 10^{-1}$	$6,3556 \cdot 10^{-1}$	$3,9285 \cdot 10^{-1}$
	$10^{-4}$	$9,2729 \cdot 10^{-1}$	$7,2613 \cdot 10^{-2}$	$5,3631 \cdot 10^{-1}$	$2,2749 \cdot 10^{-1}$
	$10^{-5}$	$9,3135 \cdot 10^{-1}$	$6,8643 \cdot 10^{-2}$	$5,343 \cdot 10^{-1}$	$2,2418 \cdot 10^{-1}$
	$10^{-6}$	$9,3140 \cdot 10^{-1}$	$6,8600 \cdot 10^{-2}$	$5,3430 \cdot 10^{-1}$	$2,2414 \cdot 10^{-1}$

При визначенні  $\delta$  враховано, що виправлення ідентифікованих помилок підтверджується користувачем, і поява „вторинних” помилок принципово можлива, як і при чисто ручному виправленні, хоча і з меншою імовірністю. У зв'язку з цим для  $\delta$  можна записати вираз, що узагальнює (4):

$$\delta = \frac{P_{ЛК} + \pi_{ХПК} (P_{тк} + \alpha \cdot P_{ПАК})}{\pi_{ХПК} (1-r)}. \quad (11)$$

Коефіцієнт  $\alpha$  визначає, наскільки напівавтоматична корекція менше схильна до вторинних помилок, ніж чисто ручна. У зв'язку з відсутністю в літературі прямих даних для  $\delta$  у розрахунках прийнято два гіпотетичних значення: 0,5 та 0,167. Перше значення враховує міркування відносно механізмів виникнення транскрипційних помилок людини [6], а друге впливає з гіпотетичного припущення, що імовірність вторинної помилки в режимі підтвердження в  $n_x$  разів менше, ніж в режимі чисто ручного коригування. Отже, в табл. 4 наведено дві групи значень  $\delta$ , зокрема,  $\delta(0,5)$  та  $\delta(0,167)$ .

Що стосується значень точок рівноваги, то навіть для найбільших значень  $r$ , що розглядаються ( $r = 10^{-2}$ ), ми маємо, як видно з табл. 4,  $\delta < 1$ , тобто  $r_0 < 10^{-2}$ . Теоретично тут

значення  $r_0$  близьке до 1, а практично це означає, що алгоритм 3 завжди забезпечує більш достовірні результати, ніж чисто ручна корекція без АК. Вважати цей алгоритм, безумовно, кращим, ніж алгоритми 1, 2, заважає та обставина, що  $\delta_3$  по мірі зменшення  $r$  залишається практично незмінним, в той час, як  $\delta_1, \delta_2$  зменшуються також. Тобто існують (і це видно з табл. 2, 3, 4) такі значення  $r$ , для яких алгоритми 1, 2 дають кращі результати з точки зору результативної достовірності.

### 3.4. Алгоритм 4

Даний алгоритм є комбінацією алгоритмів 2, 3, що спрямована на підвищення значень  $P_{ак}$  і зменшення  $P_{Пак}$ ,  $P_{рк}$  у порівнянні з алгоритмом 3 за рахунок перебору **всіх** варіацій ансамблю, що коректується. Якщо в результаті цього перебору з'ясується, що помилка однозначна, здійснюється її **автоматичне** виправлення. При багатозначній помилці користувачу пропонується, як і в алгоритмі 3, до  $m$  варіантів корекції, після чого помилка виправляється напівавтоматично або цілком вручну.

Граф варіантів подій наведено на рис. 4.

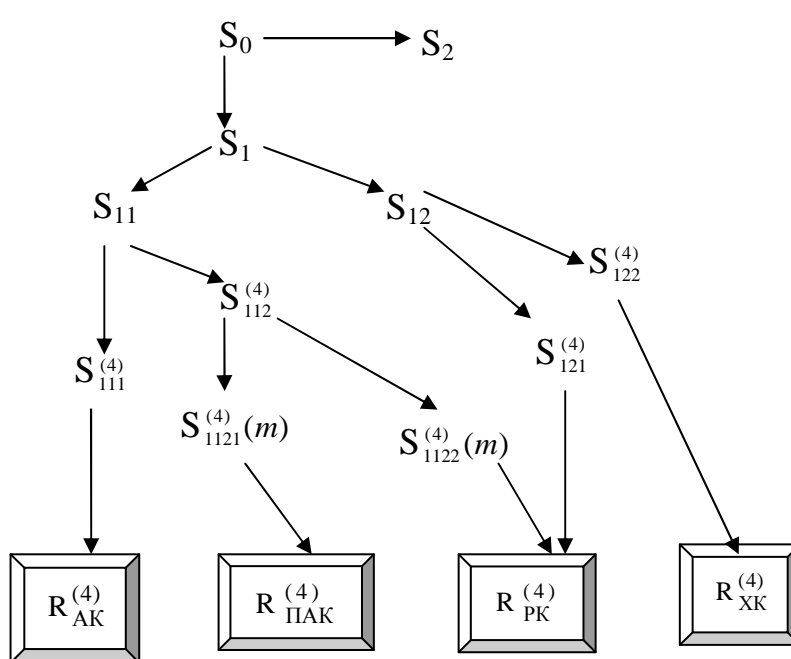


Рис. 4. Граф часткових варіантів для алгоритму 4

$S_{111}^{(4)}$  – помилка ансамблю  $K(k_j)$  однозначна;  
 $S_{112}^{(4)} = \overline{S}_{111}^{(4)}$ ;  
 $S_{122}^{(4)}$  – помилка, що не належить ансамблю  $K(k_j)$ , однозначна;  
 $S_{121}^{(4)} = \overline{S}_{122}^{(4)}$ ;  
 $S_{1121}^{(4)}(m)$  – помилка ідентифікована користувачем за  $m$  спроб;  
 $S_{1122}^{(4)} = \overline{S}_{1121}^{(4)}(m)$ .

Пропускаючи запис загальних результатів, досить ясний з рис. 4 і опису алгоритмів 2, 3, наведемо остаточні вирази для підсумкових імовірностей:

$$P_{АК}^{(4)} = (1-r) \cdot P_{\Sigma} \cdot (1-r)^{V-1}; \quad (12)$$

$$P_{ПАК}^{(4)} = (1-r) \cdot P_{\Sigma} \cdot [1 - (1-r)^{V-1}] \cdot \pi(m); \quad (13)$$

$$P_{РК}^{(4)} = (1-r) \cdot \left\{ (1-P_{\Sigma}) \cdot [1 - V r (1-r)^V] + P_{\Sigma} [1 - (1-r)^{V-1}] [1 - \pi(m)] \right\}; \quad (14)$$

$$P_{XK}^{(4)} = (1-r) \cdot (1-P_{\Sigma}) \cdot V_r(1-r)^V. \quad (15)$$

Результати розрахунків для  $m = 2$  наведено у табл. 5.

Таблиця 5. Результати розрахунків для алгоритму 4

$K(k_j)$	$r$	$P_{AK}$	$P_{IIAK}$	$P_{PK}$	$P_{XK}$	$\delta(0.5)$	$\delta(0.167)$
$K_1$	$10^{-2}$	$3,4382 \cdot 10^{-1}$	$2,6064 \cdot 10^{-1}$	$2,6354 \cdot 10^{-1}$	$1,2199 \cdot 10^{-1}$	$1,4290 \cdot 10^{+1}$	$1,4203 \cdot 10^{+1}$
	$10^{-3}$	$5,8644 \cdot 10^{-1}$	$3,5019 \cdot 10^{-2}$	$3,5654 \cdot 10^{-1}$	$2,0997 \cdot 10^{-2}$	$2,7440 \cdot 10^{+0}$	$2,7323 \cdot 10^{+0}$
	$10^{-4}$	$6,1844 \cdot 10^{-1}$	$3,5975 \cdot 10^{-3}$	$3,7565 \cdot 10^{-1}$	$2,2163 \cdot 10^{-3}$	$6,2737 \cdot 10^{-1}$	$6,2617 \cdot 10^{-1}$
	$10^{-5}$	$6,2173 \cdot 10^{-1}$	$3,6071 \cdot 10^{-4}$	$3,7767 \cdot 10^{-1}$	$2,2283 \cdot 10^{-4}$	$4,0298 \cdot 10^{-1}$	$4,0286 \cdot 10^{-1}$
	$10^{-6}$	$6,2206 \cdot 10^{-1}$	$3,6081 \cdot 10^{-5}$	$3,7788 \cdot 10^{-1}$	$2,2295 \cdot 10^{-5}$	$3,8041 \cdot 10^{-1}$	$3,8040 \cdot 10^{-1}$
$K_2$	$10^{-2}$	$1,2169 \cdot 10^{-1}$	$5,6279 \cdot 10^{-1}$	$2,7864 \cdot 10^{-1}$	$2,6875 \cdot 10^{-2}$	$3,6262 \cdot 10^{+0}$	$3,4369 \cdot 10^{+0}$
	$10^{-3}$	$7,3687 \cdot 10^{-1}$	$1,6048 \cdot 10^{-1}$	$8,5230 \cdot 10^{-2}$	$1,6421 \cdot 10^{-2}$	$2,0188 \cdot 10^{+0}$	$1,9653 \cdot 10^{+0}$
	$10^{-4}$	$8,8148 \cdot 10^{-1}$	$1,7627 \cdot 10^{-2}$	$9,8825 \cdot 10^{-2}$	$1,9662 \cdot 10^{-3}$	$3,2934 \cdot 10^{-1}$	$3,2347 \cdot 10^{-1}$
	$10^{-5}$	$8,9741 \cdot 10^{-1}$	$1,7786 \cdot 10^{-3}$	$1,0060 \cdot 10^{-1}$	$2,0019 \cdot 10^{-4}$	$1,2406 \cdot 10^{-1}$	$1,2347 \cdot 10^{-1}$
	$10^{-6}$	$8,9902 \cdot 10^{-1}$	$1,7802 \cdot 10^{-4}$	$1,0078 \cdot 10^{-1}$	$2,0055 \cdot 10^{-5}$	$1,0313 \cdot 10^{-1}$	$1,0307 \cdot 10^{-1}$
$K_3$	$10^{-2}$	$4,3715 \cdot 10^{-1}$	$1,0984 \cdot 10^{-1}$	$8,8016 \cdot 10^{-1}$	$5,3518 \cdot 10^{-8}$	$9,4453 \cdot 10^{-1}$	$9,0759 \cdot 10^{-1}$
	$10^{-3}$	$1,7362 \cdot 10^{-1}$	$5,9228 \cdot 10^{-1}$	$2,1166 \cdot 10^{-1}$	$2,1448 \cdot 10^{-2}$	$2,9288 \cdot 10^{+0}$	$2,7314 \cdot 10^{+0}$
	$10^{-4}$	$7,8743 \cdot 10^{-1}$	$1,4325 \cdot 10^{-1}$	$5,9477 \cdot 10^{-2}$	$9,7367 \cdot 10^{-3}$	$1,2289 \cdot 10^{+0}$	$1,1812 \cdot 10^{+0}$
	$10^{-5}$	$9,1589 \cdot 10^{-1}$	$1,5498 \cdot 10^{-2}$	$6,7467 \cdot 10^{-2}$	$1,1326 \cdot 10^{-3}$	$2,0291 \cdot 10^{-1}$	$1,9775 \cdot 10^{-1}$
	$10^{-6}$	$9,2984 \cdot 10^{-1}$	$1,5616 \cdot 10^{-3}$	$6,8485 \cdot 10^{-2}$	$1,1499 \cdot 10^{-4}$	$8,2229 \cdot 10^{-2}$	$8,1709 \cdot 10^{-2}$

З порівняння даних табл. 3, 4, 5 видно, що алгоритм 4 об'єднує позитивні риси алгоритмів, у певній мірі зменшуючи визначені недоліки. Значення точок рівноваги досягаються при  $r_0 = 2,5162 \cdot 10^{-4}$  для  $K_1$ ,  $r_0 = 4,1868 \cdot 10^{-4}$  для  $K_2$ ,  $r_0 = 7,7380 \cdot 10^{-4}$  для  $K_3$ , а значення  $\delta$  зменшується при подальшому зменшенні  $r$ , досягаючи менших співставних значень, ніж для алгоритмів 1, 2, 3.

#### 4. Загальні висновки і заключні зауваження

Таким чином, нами розглянуті всі модифікації загального методу автоматичної ідентифікації та корекції помилок (АІК), що представляють практичний інтерес. Одержані вирази (1) – (15) та ілюстративні дані табл. 2 – 5 дають інформацію до прийняття рішення про доцільність використання АІК у конкретному випадку, про вибір ансамблю помилок, що коригуються, та алгоритму.

На основі викладеного зробимо такі висновки.

1. Розглянутий метод і конкретні алгоритми його реалізації можуть бути використані для зниження загальної трудомісткості підготовки та вводу інформації в ЕОМ. Використання методу дозволяє автоматично або з мінімальною участю користувача виправляти від 62% до 93% помилок, в залежності від обраного ансамблю помилок, що коригуються. Більше того, для  $r < r_0$  метод забезпечує і підвищення результативної достовірності. Зокрема, наприклад, для алгоритму 2 і

ансамблю  $K_1$ , що містить помилки з сумарною імовірністю більше 0,62,  $r_0 = 2,49 \cdot 10^{-4}$ . Якщо врахувати, що величина  $r$  для багатьох інформаційних об'єктів сягає значень  $10^{-5} - 10^{-6}$ , а часто і менше, то підвищення достовірності АІК можна вважати скоріше правилом, ніж виключенням.

2. Всі розглянуті алгоритми за відповідних умов мають право на існування і застосування, тобто вони є прийнятними у сенсі Парето. Так, алгоритми 3 і 4 не застосовні у режимі off-line. Алгоритм 1 має найгіршу коригуючу спроможність, проте потребує перебору приблизно вдвічі меншої кількості варіацій, ніж алгоритм 2. Відзначимо також, що при досить малих  $r$  ( $10^{-5} \div 10^{-6}$ ) різниця між значеннями  $P_{AK}$  для обох алгоритмів взагалі майже відсутня.

3. Алгоритм 4 краще ніж 3 (обидва потребують режиму on-line) по коректуючій спроможності і обсягу ручних коректувань, проте, як і алгоритм 2 у порівнянні з алгоритмом 1, потребує приблизно вдвічі більшого обсягу перебору. А в режимі on-line це особливо суттєво, тим більше, якщо  $N$  і  $V$  мають великі значення.

При малих значеннях розширення ансамблю для всіх алгоритмів підвищує коректуючу спроможність та результативну достовірність методу. Так, неважко показати, що для всіх алгоритмів

$$\lim_{r \rightarrow 0} \delta = \delta(0) \approx P_{PK} + \alpha P_{ПАК}.$$

Для алгоритмів 1, 2, 4  $\delta(0) \approx 0,378; 0,1; 0,068$  відповідно для  $K_1, K_2, K_3$ . Для алгоритму 3 і  $\alpha = 0,167$   $\delta(0)$  відповідно дорівнює близько 0,48; 0,25; 0,22.

4. Як впливає з попередніх зауважень, практичне застосування методу АІК, і зокрема, доцільний вибір алгоритму і ансамблю  $K(k_j)$  залежить від багатьох факторів. Серед них характеристики інформації, тобто словника (значення  $N, r, n_x, n$ ), обчислювальні потужності ЕОМ, режим вводу і контролю – корекції (on-line, off-line) тощо. У зв'язку з цим, доцільним може бути таке рішення по програмній реалізації АІК. Програма АІК має включати:

- ядро з функцією генерації варіацій та сканування словника (з можливістю налаштування на ансамбль помилок, що коригуються);
- блок реалізації алгоритму прийняття рішення при співпадині варіацій із словником;
- інтерфейсний блок, що забезпечує взаємодію з користувачем і комплексування з базовими засобами вводу СУБД.

5. Наприкінці відзначимо, що при наявних обмеженнях з боку обчислювальних потужностей суттєве значення може мати оптимізація організації і обробки словника (генерації варіацій і сканування). Це задача, що потребує окремого дослідження в рамках подальшого розвитку даного напрямку роботи.

## СПИСОК ЛІТЕРАТУРИ

1. Задача повышения качества информационной базы агентно-ориентированных интеллектуализованных СППР / Г.Е. Кузьменко, В.А. Литвинов, Ю.Г. Пилипенко, В.И. Ходак // Математические машины системы. – 2002. – № 3. – С. 53 – 61.

2. Использование помехоустойчивого кодирования для защиты информации от ошибок оператора / И.М. Бояринов, А.А. Давыдов, Э.М. Мамедли, Ю.Б. Смеркис. – М.: АТ, 1983. – № 3. – С. 5 – 49.
3. Sethi A.S., Rajaraman V., Kenjaie P.S. An error-correcting coding scheme for alphanumeric data // Information Processing Letters. – 1988. – Vol. 7. – P. 72 – 77.
4. Дремов И.В., Литвинов В.А. Автоматическая коррекция ошибок на основе словаря-эталона // УСиМ. – 1989. – № 3. – С. 77 – 80.
5. Литвинов В.А., Майстренко С.Я., Ступак Н.Б. Некоторые оценки вероятностных характеристик процесса автоматической идентификации ошибок пользователя на основе эталонного словаря // УСиМ. – 2001. – № 2. – С. 21 – 24.
6. Литвинов В.А., Крамаренко В.В. Контроль достоверности и восстановление информации в человеко-машинных системах. – Киев: Техніка, 1988. – 200 с.