

## МЕТОДЫ ОПРЕДЕЛЕНИЯ ЛИЧНОСТИ ПОЛЬЗОВАТЕЛЯ НА ОСНОВЕ ИНДИВИДУАЛЬНЫХ ОСОБЕННОСТЕЙ КОМПЬЮТЕРНОГО ПОЧЕРКА

© 2021 Л. С. Крыжевич✉

*Курский государственный университет  
ул. Радищева, 33, 305000 Курск, Российская Федерация*

**Аннотация.** В данной статье выдвигается гипотеза о том, что набор текста на клавиатуре каждым конкретным человеком имеет индивидуальные особенности. Развитие этой технологии в дальнейшем поможет предотвратить попытки несанкционированного доступа к персональным данным, банковским счетам и коммерческой тайне. Среди существующих методов биометрической аутентификации предложенный подход относится к категории динамических методов, подвергающихся изменениям со временем. Эта особенность не позволяет злоумышленнику через доступ к сети украсть шаблон почерка пользователя, его скопировать или подделать. Во время работы на клавиатуре человеком используется более 20 различных мышц, что делает стиль набора уникальным. В качестве основных характеристик для проведения аутентификации личности были взяты скорость набора, время удержания клавиши в нажатом положении, время поиска следующей клавиши, периодические опечатки в процессе набора текста и многое другое. Компьютерный почерк может быть зафиксирован в виде различных метрик и проанализирован статистическими методами. Автором статьи раскрывается методика и условия проведения эксперимента. Система ведет подсчет количества нажатий за единицу времени, устанавливает временные метки, производит сбор статистических данных для построения гистограмм. Эксперимент проводится в разное время суток с использованием различных типов клавиатур. Закладывается гипотеза о том, что данные выборки подчиняются нормальному распределению, что подтверждается анализом полученных результатов по критерию согласия Пирсона. Для определения отличий стилей набора испытуемого от своего собственного находится процент совпадений показателей по интегральной формуле Муавра — Лапласа для нормальных распределений, значение которого около 90 %. Аналогичным образом проводится сравнительный анализ полученных результатов с разными пользователями. В данном случае этот показатель будет гораздо ниже и не превышает 60 %. Сравнительный анализ дает возможность аутентифицировать личность и является достаточным информационным показателем для предотвращения попыток несанкционированного доступа.

**Ключевые слова:** компьютерный почерк, аутентификация пользователя, нормальное распределение, интегральная теорема Муавра — Лапласа, критерий согласия Пирсона.

### ВВЕДЕНИЕ

В современном обществе почти все данные хранятся в цифровом виде, в базах данных, или на облачных хранилищах, к которым осуществляется доступ в режиме онлайн. Храниться могут важные документы, договоры,

банковские данные, пароли — при хищении этих данных может пропасть личная или коммерческая информация, могут опустошаться банковские счета. Соответственно, растет количество злоумышленников, желающих похитить информацию.

Существует очень много способов защиты информации, однако они постоянно устаревают. Чтобы выявить нарушителя, нужно определить, имеет ли данный человек права

✉ Крыжевич Леонид Святославович  
e-mail: [Leonid@programist.ru](mailto:Leonid@programist.ru)



Контент доступен под лицензией Creative Commons Attribution 4.0 License.  
The content is available under Creative Commons Attribution 4.0 License.

на доступ к системе. На этом основании возникла идея об аутентификации пользователя по электронному почерку.

Каждый человек обладает своими особенностями при наборе текста на клавиатуре, печатает с определенной скоростью, разное время удерживает клавиши. Было решено замерить эти характеристики и провести их анализ.

## 1. ОБЗОР ИЗВЕСТНЫХ РАБОТ В ОБЛАСТИ АУТЕНТИФИКАЦИИ ПО КОМПЬЮТЕРНОМУ ПОЧЕРКУ

В литературе представлено несколько подходов к определению личности пользователя по особенностям его набора на клавиатуре.

В статье [1] авторами была предложена схема проверки набора парольной фразы пользователя на соответствие эталонным наборам. Также перечислены условия, при которых полученный набор считался соответствующим эталонному и аутентификация пользователя по клавиатурному почерку подтверждается.

В работе [2] было показано, что минимальный объем набора в 30 слов обеспечивает достаточную устойчивость статистических характеристик и позволяет осуществлять текущий контроль набора пользователя даже для больших текстов. Аналогично подтверждается, что данный подход к авторизации пользователя применим для обработки и анализа выборок разных объемов.

В статье [3] применяется аппаратное решение задачи анализа компьютерного почерка. Новые результаты получены путем дооснащения клавиатуры тензорными датчиками вибрации и давления, что позволило снизить количество ошибок идентификации и выявить дополнительные признаки для аутентификации пользователя.

В работе [4] рассмотрен подход к организации процедуры анализа компьютерного почерка на основе принципов самоорганизации. Предлагаемая автором модель потока событий в системе «человек-клавиатура-компьютер» позволяет сформировать меру специального вида — внешнего критерия,

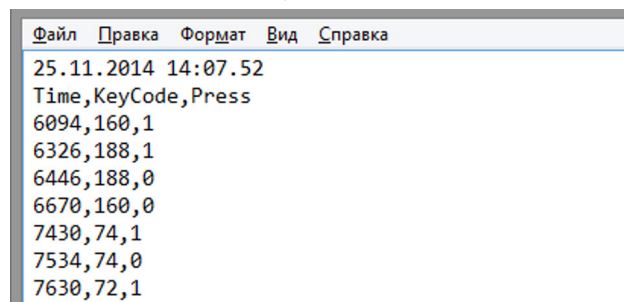
которая уменьшается с увеличением степени схожести получаемых групп индикаторов по структуре и рангу. Применение основанных на самоорганизации инструментов позволяет снизить субъективность результатов.

## 2. УСЛОВИЯ ЭКСПЕРИМЕНТА

Для получения тестовых результатов проводился эксперимент примерно на ста студентах факультета ФМИ [5]. Они должны были набрать текст, состоящий не менее, чем из 4 предложений, при этом программа измеряла следующие характеристики для каждого символа: время нажатия с момента запуска программы (в миллисекундах), ASCII-код нажатой клавиши, статус нажатия — 1(нажата) или 0(отпущена).

На рис. 1 представлен внешний вид файла со статистическими данными, подлежащими обработке.

Цель исследования — определение индивидуальных характеристик набора пользователя для выявления отличий с тестовыми образцами почерка других пользователей.



Time	KeyCode	Press
6094	160	1
6326	188	1
6446	188	0
6670	160	0
7430	74	1
7534	74	0
7630	72	1

Рис. 1. Файл с данными  
[Fig. 1. The file which includes data]

## 3. АНАЛИЗ ДАННЫХ

Подробно рассмотрен анализ статистики по первому характерному признаку — время удержания клавиш. Если из тестового образца выявить попарно все соседние замеры для одного и того же символа (когда был нажат и когда отпущен) и вычесть из времени отпускания время нажатия, то будет получено время длительности нажатия каждого символа. Тестовые длительности расположены для всех символов на двумерном графике, гори-

горизонтальная ось которого будет временем нажатия в миллисекундах, а вертикальная — частотой нажатия (т. е. отношением количества нажатий конкретной длительности к общему количеству нажатий). После сортировки по времени нажатия график будет иметь следующий вид (рис. 2).

### 3.1. Проверка на нормальное распределение

Выдвигается гипотеза, что это распределение является нормальным. Чтобы его проверить, проводится анализ полученных данных по непараметрическому критерию  $\chi^2$  Пирсона.

Выборка разбивается на 14 непересекающихся интервалов. Для каждого подсчитывается количество тестовых значений, попавших в него. Обязательное условие при этом — в каждый интервал должно попасть не менее 5 результатов нажатия одной клавиши [6].

Соблюдая это условие, по формуле среднеарифметического усредняются значения по этим интервалам и строится новый график.

Для того, чтобы определить, является ли распределение нормальным, необходимо воспользоваться критерием  $\chi^2$  Пирсона [7]. Необходимы следующие величины:

$x_i$  — ось абсцисс или время;  
 $f$  — частота;  
 $(x_i \cdot f)$  — величина для вычисления средней взвешенной;

$S$  — накопленная частота, получаемая путем сложения каждой предыдущей частоты с последующей.

$(|x_i - x_{cp}| \cdot f_i)$  — разность текущего  $x_i$  и средней взвешенной, умноженной на текущую частоту.

$((x_i - x_{cp})^2 \cdot f_i)$  — разность текущего  $x_i$  и средней взвешенной, возведенной в квадрат и умноженной на текущую частоту и  $(f_i / f)$  — отношение относительной частоты к общей сумме.

Необходимо рассчитать среднюю взвешенную:

$$x_{cp} = \frac{\sum_i^n x_i \cdot f_i}{\sum_i^n f_i} = \frac{47656}{479} = 99,49.$$

Данные величины для расчетов будут занесены в табл. 1.

Определяется мода для данного распределения. Для нашего случая выбирается мода при  $x_i = 96$  (при этом значение частоты = 59).

Медианой в этом случае так же будет значение  $x_i = 96$ , т. к. это первое значение, при

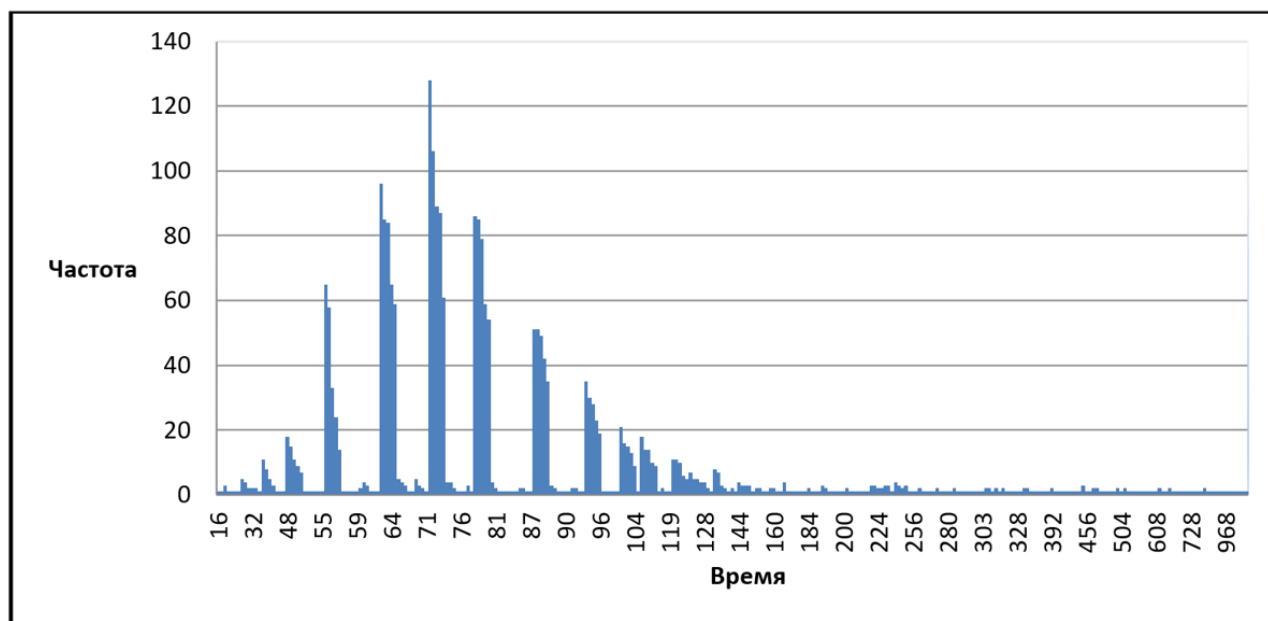


Рис. 2. Гистограмма время/частота для первого набора тестируемого  
 [Fig. 2. The time/frequency bar chart for the first typing session of a test person]

Таблица 1. Расчетная таблица для эмпирических частот первого набора  
 [Table 1. The calculation table for empirical frequencies of the first typing session]

$x_i$	Кол-во, $f_i$	Отн. частота, $p_i = f_i / f$	$x_i * p_i$	Накопленная частота, $S$	$ x - x_{cp}  * p_i$	$(x - x_{cp})^2 * p_i$	Накопленная частота, $S$
48	10	0.0209	480	0.0209	514.906	26512.824	10
56	14	0.0292	784	0,0501	608.868	26480.059	24
64	30	0.0626	1920	0,1127	1064.718	37787.492	54
72	35	0.0731	2520	0,1858	962.171	26450.669	89
80	51	0.106	4080	0,2918	994.021	19374.069	140
88	52	0.109	4576	0,4008	597.511	6865.769	192
96	59	0.123	5664	0,5238	205.946	718.875	251
104	50	0.104	5200	0,6278	225.47	1016.732	301
112	54	0.113	6048	0,7408	675.507	8450.187	355
120	37	0.0772	4440	0,818	758.848	15563.505	392
128	30	0.0626	3840	0,8806	855.282	24383.567	422
136	27	0.0564	3672	0,937	985.754	35989.269	449
144	16	0.0334	2304	0,9704	712.15	31697.379	465
152	14	0.0292	2128	0,9996	735.132	38601.311	479
Итого	479	1	47656		9896.284	299891.708	

котором значение накопленной частоты будет больше  $479 / 2 \approx 240$ .

В симметричных рядах распределения значение моды и медианы совпадают со средней величиной ( $x_{cp} = M_e = M_o$ ), а в умеренно асимметричных они соотносятся таким образом:  $3 \cdot (x_{cp} - M_e) \approx x_{cp} - M_o$ .

Размах вариации, т. е. разница между наибольшим и наименьшим значениями  $x$ , составляет  $R = 152 - 48 = 104$ .

Вычислим среднее линейное отклонение:

$$d = \frac{\sum_i (|x_{i0} - x_i| \cdot f_i)}{\sum_i f_i} = \frac{9896,284}{479} = 20,66.$$

Вычислим дисперсию

$$D = \frac{\sum_i (x_i - x_{cp})^2 f_i}{\sum_i f_i} = \frac{299891,708}{479} = 626,079.$$

Дисперсия показывает меру рассеяния всех значений выборки вокруг среднего.

Вычисляется среднее квадратическое отклонение:  $\sigma = \sqrt{D} = \sqrt{626,079} = 25,022$ .

Проверим по табл. 1 гипотезу о том, что  $X$  распределено по нормальному закону с помощью критерия согласия Пирсона

$K = \sum_i \frac{(n_i - n_i^*)^2}{n_i^*}$ , где  $n \cdot i$  — теоретические частоты, вычисляемые по формуле  $n_i = \frac{n \cdot h}{\sigma} \cdot \varphi_i$ .

В данной формуле используются следующие обозначения:  $n = 479$ ,  $h = 8$  (ширина интервала),  $\sigma = 25,022$ ,  $x_{cp} = 99,49$ ,  $\varphi_i$  — соответствующее значение из таблицы Лапласа. Теоретические частоты представлены в табл. 2.

Теперь необходимо сравнить теоретические и эмпирические частоты.

Составляется табл. 3, в которой представлены наблюдаемые значения критерия Пирсона:

$$\chi^2 = \sum_i \frac{(n_i - n_i^*)^2}{n_i^*}.$$

Для нее понадобятся следующие показатели:

$i$  — порядковый номер,

$n_i$  — наблюдаемые частоты,

$n_i^*$  — теоретические частоты,

$(n_i - n_i^*)$  — разница наблюдаемых и теоретических частот,

Таблица 2. Расчетная таблица для теоретических частот первого набора  
 [Table 2. The calculation table for theoretical frequencies of the first typing session]

$i$	$x_i$	$u_i$	$\varphi_i$	$n * i$
1	48	-2.0578	0,0478	7.32
2	56	-1.7381	0,0878	13.446
3	64	-1.4184	0,1456	22.298
4	72	-1.0987	0,2179	33.371
5	80	-0.779	0,2943	45.071
6	88	-0.4592	0,3589	54.965
7	96	-0.1395	0,3951	60.509
8	104	0.1802	0,3918	60.003
9	112	0.4999	0,3521	53.923
10	120	0.8197	0,285	43.647
11	128	1.1394	0,2083	31.901
12	136	1.4591	0,1374	21.043
13	144	1.7788	0,0818	12.527
14	152	2.0986	0,044	6.739

Таблица 3. Расчетная таблица для сравнения теоретических  
 и эмпирических частот первого набора  
 [Table 3. The calculation table for comparison of theoretical  
 and empirical frequencies of the first typing session]

$i$	$x_i$	$u_i$	$\varphi_i$	$n * i$
1	48	-2.0578	0,0478	7.32
2	56	-1.7381	0,0878	13.446
3	64	-1.4184	0,1456	22.298
4	72	-1.0987	0,2179	33.371
5	80	-0.779	0,2943	45.071
6	88	-0.4592	0,3589	54.965
7	96	-0.1395	0,3951	60.509
8	104	0.1802	0,3918	60.003
9	112	0.4999	0,3521	53.923
10	120	0.8197	0,285	43.647
11	128	1.1394	0,2083	31.901
12	136	1.4591	0,1374	21.043
13	144	1.7788	0,0818	12.527
14	152	2.0986	0,044	6.739

$\frac{(n_i - n_i^*)}{n_i^*}$  — разница, возведенная в квадрат, поделенная на текущее значение теоретической частоты.

Далее, необходимо рассчитать  $K_{набл}$  — наблюдаемое значение границы критической области и  $K_{кр}$  — теоретическое значение границы критической области. Чем больше будет значение  $K_{набл}$  отличаться от  $K_{кр}$ , тем сильнее довод против основной гипотезы [7].



Её границу  $K_{кр} = \chi^2(k-r-1; \alpha)$  находится по таблицам распределения  $\chi^2$  и заданным значениям  $x_{cp}$  и  $\sigma$  (оцененным по выборке),  $k=14$ ,  $r=2$ , уровень значимости  $\alpha$  определяем равным 0,05.

$$K_{кр} = (0.05; 11) = 19.67514;$$

$$K_{набл} = 17.99.$$

Наблюдаемое значение статистики Пирсона не попадает в критическую область: ( $K_{набл} < K_{кр}$ ). Справедливо предположение о том, что данные выборки подчиняются закону нормального распределения.

Рассуждая аналогичным образом, проверяется по критерию Пирсона вторая подборка данных, полученная для того же самого человека, но на другой текстовой последовательности (табл. 4).

Получаются следующие значения величин: средняя взвешенная (выборочная средняя)

$$x_{cp} = \frac{\sum_i x_i \cdot f_i}{\sum_i f_i} = \frac{42294}{430} = 98,36.$$

Таблица 4. Расчетная таблица для эмпирических частот второго набора  
[Table 4. The calculation table for empirical frequencies of the second typing session]

$x_i$	Кол-во, $f_i$	Отн. частота, $p_i = f_i / f$	$x_i * p_i$	Накопленная частота, $S$	$ x - x_{cp}  * p_i$	$(x - x_{cp})^2 * p_i$	Накопленная частота, $S$
56	5	0.0116	280	0.0116	211.791	8971.06	10
64	15	0.0349	960	0,0465	515.372	17707.226	24
72	35	0.0814	2520	0,1279	922.535	24316.303	54
80	38	0.0884	3040	0,2163	697.609	12806.809	89
86	45	0.105	3870	0,3213	556.116	6872.563	140
88	53	0.123	4664	0,4443	548.981	5686.426	192
96	56	0.13	5376	0,5743	132.056	311.406	251
104	51	0.119	5304	0,6933	287.735	1623.36	301
112	41	0.0953	4592	0,7886	559.316	7630.115	355
120	35	0.0814	4200	0,87	757.465	16392.954	392
128	30	0.0698	3840	0,9398	889.256	26359.197	422
136	15	0.0349	2040	0,9747	564.628	21253.645	449
144	8	0.0186	1152	0,9933	365.135	16665.435	465
152	3	0.00698	456	1,00028	160.926	8632.348	479
Итого	430	1	42294		7168.921	175228.847	

Максимальное значение повторений  $f = 56$  достигается при  $x = 96$ , тогда  $M_o = 96$ .

Половина суммы накопленной частоты составляет 216 при значении  $x_i = 96$ . Таким образом, медиана равна 96.

Размах вариации  $152 - 56 = 96$ .

Среднее линейное отклонение

$$d = \frac{\sum_i (|x_i - x_{cp}| \cdot f_i)}{\sum_i f_i} = \frac{7168,921}{430} = 16,67.$$

Каждое значение ряда отличается от другого в среднем на 16.67.

Рассчитаем дисперсию:

$$D = \frac{\sum_i (x_i - x_{cp})^2 f_i}{\sum_i f_i} = \frac{175228,847}{430} = 407,509.$$

Среднее квадратическое отклонение составит:  $\sigma = \sqrt{D} = \sqrt{407,509} = 20,187$ .

С помощью критерия согласия Пирсона проверяется гипотеза:  $X$  распределено по нормальному закону [7]. Вычисляются теоретические частоты, учитывая, что:  $n = 430$ ,

$h = 8$  (ширина интервала),  $\sigma = 20,187$ ,  
 $x_{cp} = 98,36$ .

$$n_i = \frac{n \cdot h}{\sigma} \cdot \varphi_i \Rightarrow n_i = \frac{430 \cdot 8}{20.187} \cdot \varphi_i = 170,41 \cdot \varphi_i$$

и соответствующие значения из таблицы Лапласа. Результаты представлены в табл. 5.

Сравниваются эмпирические и теоретические частоты. Составляется расчетная табл. 6 для второго набора с перечисленными выше значениями, из которой найдем наблюдаемое значение критерия Пирсона.

По принципу, описанному выше, получаем значения

$$K_{кр} = (0,05; 11) = 19,67514;$$

$$K_{набл} = 19,55.$$

Таким образом, ( $K_{набл} < K_{кр}$ ), а следовательно, распределение нормальное.

### 3.2. Сравнение различных наборов

Чтобы определить, насколько отличается стиль набора испытуемого от своего собственного, необходимо проанализировать площадь пересечения графиков [8]. Так как выборка 2 полностью пересекается выборкой 1, то площадь пересечения считают выборку 2, а площадь объединения — выборку 1.

Таблица 5. Расчетная таблица для теоретических частот второго набора  
 [Table 5. The calculation table for theoretical frequencies of the second typing session]

$i$	$x_i$	$u_i$	$\varphi_i$	$n_i^*$
1	56	-2.0983	0,044	7.498
2	64	-1.702	0,0925	15.763
3	72	-1.3057	0,1691	28.816
4	80	-0.9094	0,2637	44.937
5	86	-0.6122	0,3292	56.098
6	88	-0.5131	0,3485	59.387
7	96	-0.1168	0,3961	67.499
8	104	0.2795	0,3825	65.181
9	112	0.6758	0,3166	53.951
10	120	1.0721	0,2227	37.95
11	128	1.4684	0,1354	23.073
12	136	1.8647	0,0694	11.826
13	144	2.261	0,0303	5.163
14	152	2.6573	0,0116	1.977

Для этого необходимо воспользоваться следующей формулой:

$$\sum_i^n h \cdot l_i^{\max} - \sum_i^n h \cdot l_i^{\min},$$

где  $h$  — ширина столбиков;

$l_i^{\max} = \max(l_{1,i}; l_{2,i})$  — максимум из попарно сгруппированных высот столбиков двух графиков;

$l_i^{\min} = \min(l_{1,i}; l_{2,i})$  — соответственно, минимум.

Используя данную формулу, для первого набора получают следующие результаты

$$\sum_i^n h \cdot l_i^{\max} = 0,010438 + 0,029228 + 0,06263 + 0,073069 + 0,093946 + 0,108559 + 0,11691 + 0,104384 + 0,085595 + 0,073069 + 0,06263 + 0,031315 + 0,016701 + 0,006263 = 2,0459.$$

Для второго набора получаются следующие результаты

$$\sum_i^n h \cdot l_i^{\min} = 0,020876827 + 0,03131524 + 0,073068894 + 0,079331942 + 0,106471816 + 0,110647182 + 0,123173278 + 0,106471816 + 0,112734864 + 0,077244259 + 0,06263048 + 0,056367432 + 0,033402923 + 0,029227557 = 1,749.$$

Коэффициент совпадения эмпирических графиков:

Таблица 6. Расчетная таблица для сравнения теоретических и эмпирических частот первого набора

[Table 6. Calculated table for comparing theoretical and empirical frequencies of the first set]

$i$	$n_i$	$n_i^*$	$n_i - n_i^*$	$(n_i - n_i^*)^2$	$(n_i - n_i^*)^2 / n_i^*$
1	5	7.498	2.498	6.2398	0.832
2	15	15.7627	0.7627	0.5818	0.0369
3	35	28.816	-6.184	38.242	1.327
4	38	44.9366	6.9366	48.1161	1.071
5	45	56.0983	11.0983	123.1722	2.196
6	53	59.3872	6.3872	40.796	0.687
7	56	67.4986	11.4986	132.2176	1.959
8	51	65.181	14.181	201.102	3.085
9	41	53.9512	12.9512	167.7325	3.109
10	35	37.9499	2.9499	8.7016	0.229
11	30	23.0732	-6.9268	47.98	2.079
12	15	11.8263	-3.1737	10.0723	0.852
13	8	5.1634	-2.8366	8.0465	1.558
14	3	1.9767	-1.0233	1.0471	0.53
$\Sigma$	430	430			19.551

$$K_1 = \frac{1,749}{2,0459} = 0,85510 \approx 86 \%$$

совпадение у пользователя с самим собой.

Затем проверяют на совпадение нормальные распределения, соответствующие данным выборкам [9]. Пользуются интегральной формулой Муавра — Лапласа для нормально-го распределения:

$$\Phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{(t-m)^2}{2\sigma^2}} dt,$$

где  $\sigma$  — среднеквадратическое отклонение;

$t$  — время нажатия клавиши в миллисекундах;

$m$  — математическое ожидание.

Пусть  $S_1$  — площадь, ограниченная первым графиком,  $S_2$  — площадь, ограниченная вторым графиком.

Коэффициент совпадение теоретических графиков:

$$K_2 = \frac{S_1 \cap S_2}{S_1 \cup S_2} = \frac{53,082}{59,075} = 0,899582 \approx 90 \%$$

Даже учитывая высокую вероятность погрешности, полученный результат 86 % совпадения для эмпирической и 90 % для те-

оретической последовательности. Поэтому делается вывод, что для каждого человека существует индивидуальная закономерность в длительности нажатия кнопок при наборе, которая соблюдается.

#### 4. МАСШТАБИРОВАНИЕ НА НЕСКОЛЬКО НАБОРОВ

На рис. 3 приведены разброс математического ожидания для времени удержания различных клавиш клавиатуры [7] в наборах одного пользователя в разные дни (отмечены разным цветом).

В правом нижнем углу для этих же наборов по оси ординат отмечены среднее время удержания клавиш.

На рис. 4 аналогичные характеристики иллюстрируются для наборов разных пользователей.

Сравнительный анализ полученных результатов дает возможность сделать вывод о том, что время удержания различных клавиш клавиатуры является достаточно информативным показателем техники набора пользователя [10].



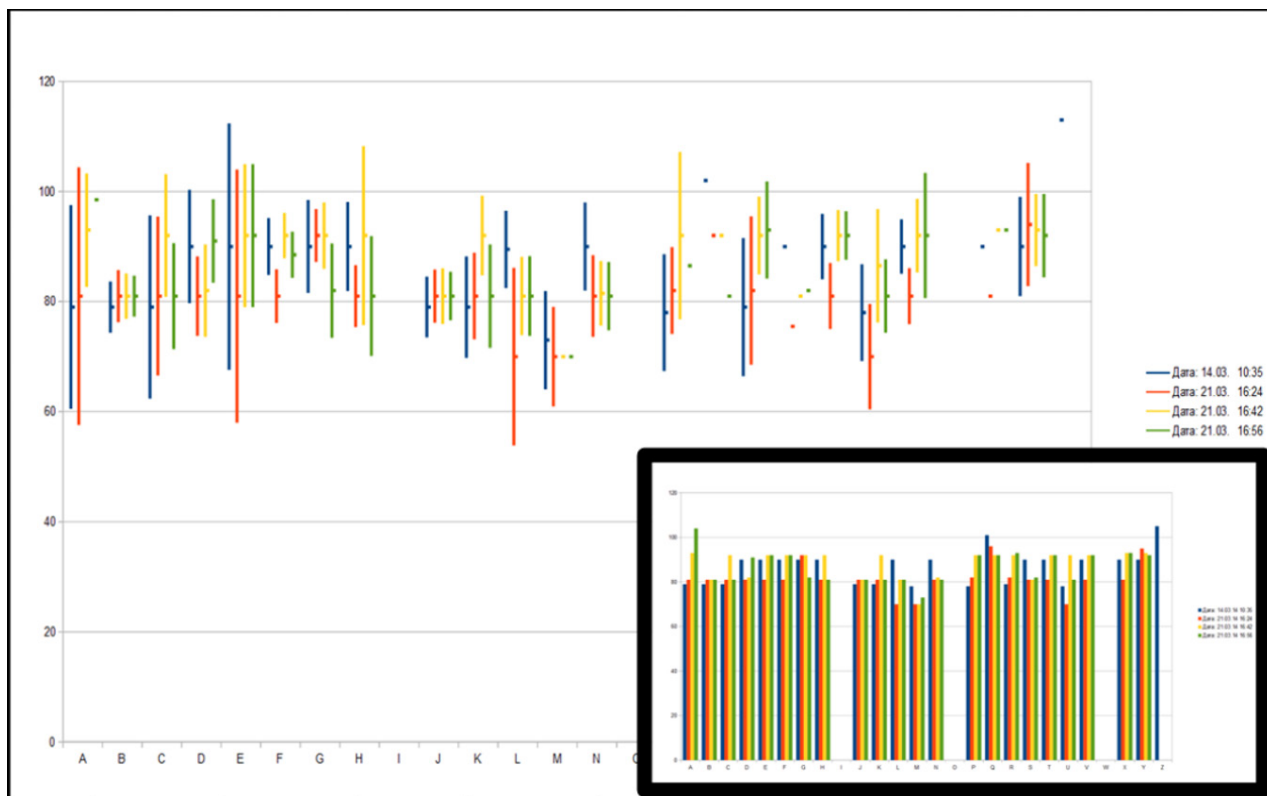


Рис. 3. График нормальных распределений различных наборов одного пользователя  
[Fig. 3. The histogram for the cases of normal distribution of various handwriting sessions of the same user]

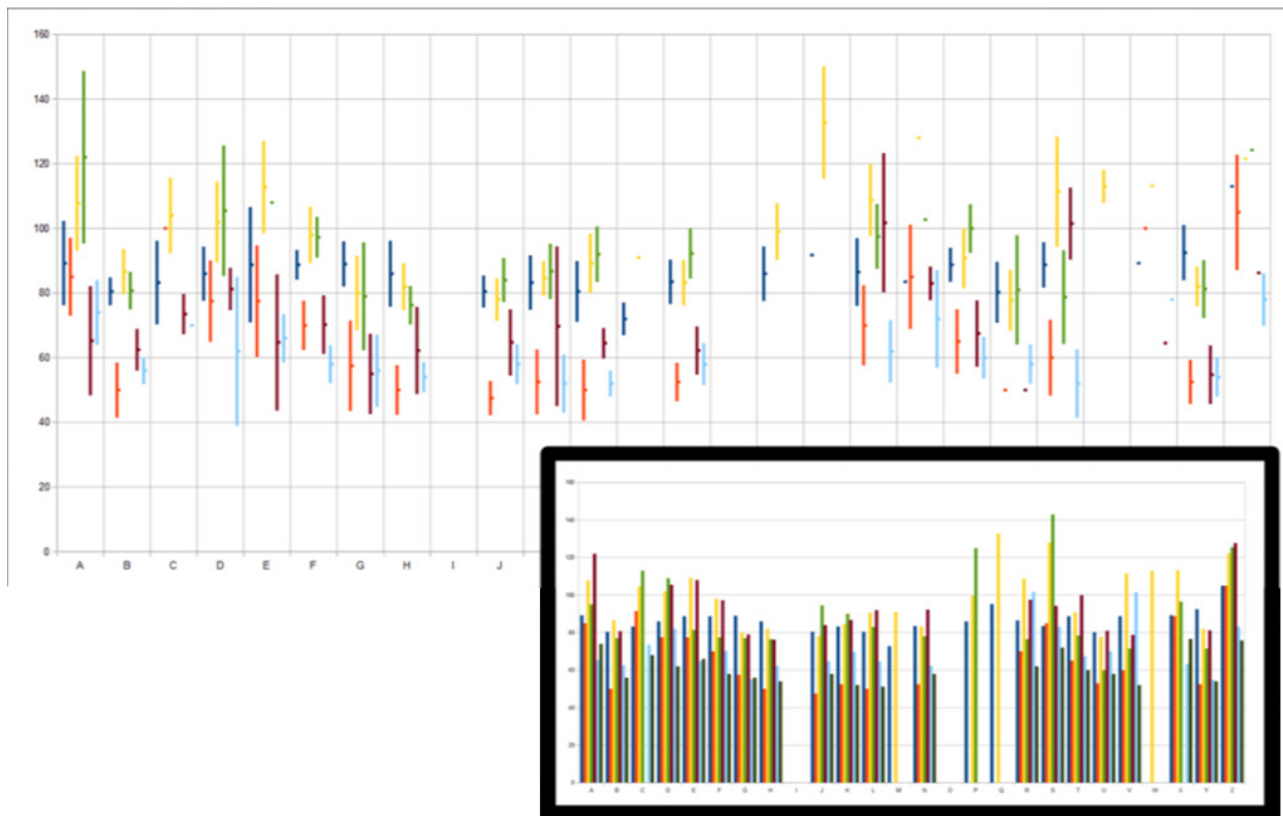


Рис. 4. График распределений наборов разных пользователей  
[Fig. 4. The histogram for the distribution of typing sessions of different users]

Несмотря на некоторый случайный разброс усредненных значений времени удержания клавиш, статистический анализ имеющихся различий позволит с достаточно большой вероятностью идентифицировать клавиатурные наборы одного пользователя и отличить наборы разных пользователей [11].

Результаты эксперимента показали, что в большинстве случаев интервалы времени для удержания клавиш являются случайной выборкой с нормальным законом распределения.

### ЗАКЛЮЧЕНИЕ

Данный метод проверки при авторизации пользователя может быть использован для выборок разных объемов. Значение  $K$  для каждого пользователя может несколько колебаться для различных наборов. Близость к единице  $K$  в основном зависит от степени сформированности клавиатурного почерка у пользователя. Для слабо набирающих пользователей критическое значение  $K$  для авторизации пользователя может быть установлено по результатам сравнительного анализа нескольких его наборов [12]. Дальнейший анализ наборов таких пользователей может быть несколько уточнен путем исключения из рассмотрения клавиш, время удержания которых имеет большое среднее квадратическое отклонение (например, существенно большее, чем среднее квадратическое по всему набору).

### КОНФЛИКТ ИНТЕРЕСОВ

Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

### СПИСОК ЛИТЕРАТУРЫ

1. Матюшина С. Н., Куликов С. С. Клавиатурный почерк в парольной аутентификации пользователей // Auditorium. – 2019. – № 4 (24). – С. 77–82.
2. Крыжевич Л. С., Матюшина С. Н. Обеспечение доступа к электронному оборудованию на основе распознавания компьютерного почерка // Сборник: «Актуальные исследования в области математики, информатики, физики и методики их изучения в современном образовательном пространстве». Результаты исследований в области математики, информатики, физики и методики их изучения в современном образовательном пространстве. Курск, 2016. – С. 83–88.
3. Сулавко А. Е., Федотов А. А., Еременко А. В. Распознавание пользователей компьютерных систем по клавиатурному почерку с учетом параметров вибрации и давления на клавиши // Динамика систем, механизмов и машин. – 2017. – Т. 5, № 4. – С. 95–105.
4. Саниев А. З. Организация процедуры анализа компьютерного почерка на основе принципов самоорганизации // Вестник Вологодского государственного университета. Серия: Технические науки. – 2021. – № 2 (12). – С. 40–42.
5. AragA'n-McndizAibal E., Delgado-Casas C., Romero-Oliva M. F. A comparative study of handwriting and computer typing in note-taking by university students, Comunicar. – 2016. DOI: 10.3916/C48-2016- 10
6. Summary and classification of statistics, 2018. – URL: <http://www.grandars.ru/student/statistika/gruppirovka-statistichskih-dannyh.html>
7. Shulenin V. P. Mathematical statistics. – Tomsk : NTL Publishing House, 2012.
8. Kryzhevich L. S., Rakov A. S. Kostenko I. V., Arkhipova V. V. Lukin D. E., Testing statistical hypotheses about the time parameters of keytyping, «Problems of cybersecurity, modeling and information processing in modern sociotechnical systems», KSU, Kursk, 2017.
9. Gmunnan V. E. Probability theory and mathematical statistics, 9th edition. – Moscow : Vysshaya shkola, 2003.
10. Fedorowich L. M., C6te J. N. Effects of standing on typing task performance and upper limb discomfort, vascular and muscular indicators, Applied Ergonomics. – 2018. DOI: 10.1016/j.apergo.2018.05. 009.
11. Kryzhevich L. S., Matyushina S. N., Kostenko I. V. Providing access to electronic equipment based on computer handwriting recognition,

«Current research in the field of exact sciences and their study in secondary and higher educational institutions», KSU, Kursk, 2015.

pressure of the fingertips during computer typing, *Journal of Physical Therapy Science*. – 2015. DOI: 10.1589/jpts.27.57.

12. Yoo W. G. Effects of different computer typing speeds on acceleration and peak contact

**Крыжевич Леонид Святославович** — канд. тех. наук, и.о. зав. кафедрой информационной безопасности, факультет физики, математики, информатики, Курский государственный университет.

E-mail: Leonid@programist.ru

ORCID iD: <https://orcid.org/0000-0002-6736-498X>

DOI: <https://doi.org/10.17308/sait.2021.3/3735>

ISSN 1995-5499

Received 05.08.2021

Accepted 20.11.2021

## METHODS FOR DETERMINING THE IDENTITY OF THE USER BASED ON THE INDIVIDUAL CHARACTERISTICS OF COMPUTER HANDWRITING

© 2021 L. S. Kryzhevich 

*Kursk State University  
33, Radisheva Street, 305000 Kursk, Russian Federation*

**Annotation.** In this article the hypothesis is considered that typing on the keyboard by each person has individual characteristics. In the future the development of this technology will help prevent attempts of unauthorized access to personal data, bank accounts and trade secrets. Among the existing methods of biometric authentication, the proposed approach belongs to the category of dynamic methods that undergo changes over time. This feature prevents an attacker from stealing, copying, or forging a user's handwriting template through network access. While working on the keyboard, a person uses more than 20 different muscles, which makes the typing style unique. The typing speed, key holding time, time to find the next key, periodic typos during typing, and much more were taken as the main characteristics for carrying out identity authentication. Computer handwriting can be recorded in the form of various metrics and analyzed by statistical methods. The author of the article reveals the methodology and conditions of the experiment. The system counts the number of clicks per unit of time, sets time stamps, and collects statistical data to build histograms. The experiment is carried out at different times of the day using different types of keyboards. It is hypothesized that the sample data obey a normal distribution, which is confirmed by the analysis of the results obtained using Pearson's nonparametric criterion. To determine the differences between the recruiting styles of the subject from his own, the percentage of coincidences of indicators is found according to the integral formula of Moivre — Laplace for normal distributions, the value of which is about 90 %. In a similar way, a comparative analysis of the results obtained with different users is carried out. In this case, this figure will be much lower and does not exceed 60 %. Comparative analysis makes it possible to authenticate a person and is a sufficient information indicator to prevent unauthorized access attempts.

**Keywords:** computer handwriting, user authentication, normal distribution, Moivre — Laplace integral theorem, nonparametric Pearson test.

---

 Kryzhevich Leonid S.  
e-mail: Leonid@programist.ru

## CONFLICT OF INTEREST

The author declare the absence of obvious and potential conflicts of interest related to the publication of this article.

## REFERENCES

1. *Matyushina S. N., Kulikov S. S.* (2019) Keyboard handwriting in password authentication of users // Auditorium. No. 4 (24). P. 77–82.
2. *Kryzhevich L. S., Matyushina S. N.* (2016) Providing access to electronic equipment based on computer handwriting recognition // Collection: “Actual research in mathematics, computer science, physics and methods of studying them in the educational space.” The results of research in the field of mathematics, computer science, physics and methods of their study in modern educational design. Kursk. P. 83–88.
3. *Sulavko A. E., Fedotov A. A., Eremenko A. V.* (2017) Recognition of users of computer systems by keyboard handwriting taking into account the parameters of vibration and pressure on the keys // Dynamics of systems, mechanisms and machines. Vol. 5. No. 4. P. 95–105.
4. *Sapiev A. Z.* (2021) Organization of the procedure for analyzing computer handwriting based on the principles of self-organization. Bulletin of the Vologda State University. Series: Engineering Sciences. No. 2 (12). P. 40–42.
5. *Aragón-McndizAibal E., Delgado-Casas C., Romero-Oliva M. F.* (2016) A comparative study of handwriting and computer typing in note-taking by university students, Comunicar. DOI: 10.3916/C48-2016-10.
6. Summary and classification of statistics (2018) URL: <http://www.grandars.ru/student/statistika/gruppirovka-statistichcskih-dannyh.html>
7. *Shulenin V. P.* (2012) Mathematical statistics. NTL Publishing House, Tomsk.
8. *Kryzhevich L. S., Rakov A. S., Kostenko I. V., Arkhipova V. V., Lukin D. E.* (2017) Testing statistical hypotheses about the time parameters of keytyping, “Problems of cybersecurity, modeling and information processing in modern sociotechnical systems”. KSU, Kursk.
9. *Gmunnan V. E.* (2003) Probability theory and mathematical statistics, 9th edition. Vysshaya shkola, Moscow.
10. *Fedorowich L. M., C6te J. N.* Effects of standing on typing task performance and upper limb discomfort, vascular and muscular indicators, Applied Ergonomics (2018). DOI: 10.1016/j.apergo.2018.05.009.
11. *Kryzhevich L. S., Matyushina S. N., Kostenko I. V.* (2015) Providing access to electronic equipment based on computer handwriting recognition, “Current research in the field of exact sciences and their study in secondary and higher educational institutions”, KSU, Kursk.
12. *Yoo W. G.* (2015) Effects of different computer typing speeds on acceleration and peak contact pressure of the fingertips during computer typing, Journal of Physical Therapy Science. DOI: 10.1589/jpts.27.57.

**Kryzhevich Leonid S.** — PhD in Technical Sciences, Chief of the Cybersecurity Department, Faculty of Physics, Mathematics, Computer science, Kursk State University.

E-mail: [Leonid@programist.ru](mailto:Leonid@programist.ru)

ORCID iD: <https://orcid.org/0000-0002-6736-498X>