

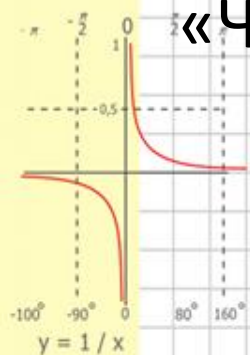
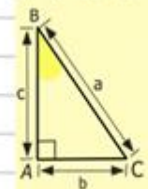
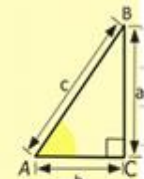
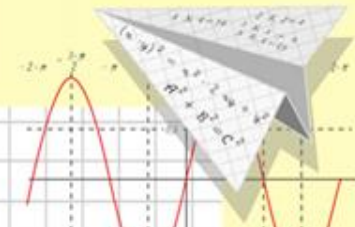
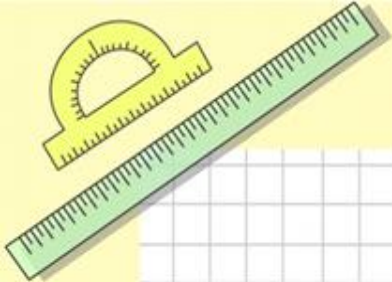
Математика

СТАТИСТИЧНЕ ДОСЛІДЖЕННЯ

«ЧАСТОТНИЙ АНАЛІЗ ВИКОРИСТАННЯ СИМВОЛІВ У ТЕКСТАХ ПО ЗАХИСТУ ІНФОРМАЦІЇ»

Лисенко Юлія Віталіївна

Кам'янський еколого-економічний ліцей



$$\begin{array}{r} 2500 \\ \times 42 \\ \hline 210 \\ + 84 \\ \hline 10500 \end{array}$$

$$y = \cos x$$

$$\begin{array}{l} 2 \times 2 = 4 \\ 3 \times 3 = 9 \\ 4 \times 4 = 16 \\ 5 \times 5 = 25 \\ 6 \times 6 = 36 \\ 7 \times 7 = 49 \\ 8 \times 8 = 64 \end{array}$$



$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

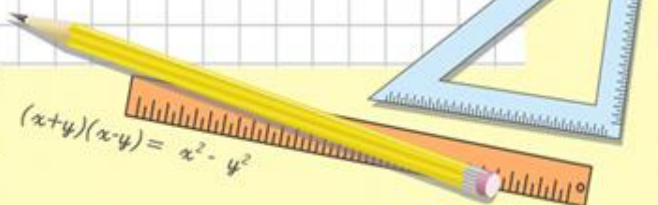
$$\frac{a}{c} + \frac{b}{c} = \frac{a+b}{c}$$



$$\sin 90^\circ = 1$$



$$\begin{cases} y = \sin 90 \\ x = 25y + 45 \\ y = 1 \\ x = 25 + 45 \\ \hline x = 70 \end{cases}$$



$$(x+y)(x-y) = x^2 - y^2$$

РОЗДІЛ 1. ТЕОРЕТИЧНА ЧАСТИНА

ЧАСТОТНИЙ АНАЛІЗ ВИКОРИСТАННЯ БУКВ РОСІЙСЬКОЇ ТА УКРАЇНСЬКОЇ МОВИ

Поняття частотного аналізу.

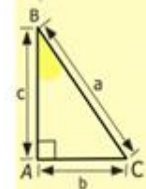
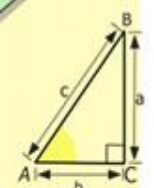
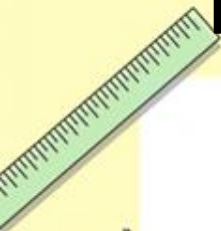
Аналіз використання букв російської мови.

Аналіз використання букв української мови.

О.О. Архипова, В.М. Журавльов

Частотний аналіз використання букв української мови

«Радіоелектроніка. Інформатика. Управління» № 2, 2009



$$\begin{array}{r} \frac{1}{2} 500 \\ \times 42 \\ \hline 210 \\ + 84 \\ \hline 105000 \end{array}$$

- $2 \times 2 = 4$
- $3 \times 3 = 9$
- $4 \times 4 = 16$
- $5 \times 5 = 25$
- $6 \times 6 = 36$
- $7 \times 7 = 49$
- $8 \times 8 = 64$
- $9 \times 9 = 81$



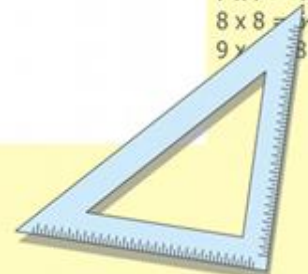
$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$
$$\frac{a}{c} + \frac{b}{c} = \frac{a+b}{c}$$

$\sin 90^\circ = 1$



$$\begin{cases} y = \sin 90 \\ x = 25y + 45 \end{cases}$$
$$\begin{cases} y = 1 \\ x = 25 + 45 \\ \hline x = 70 \end{cases}$$

$$(x+y)(x-y) = x^2 - y^2$$

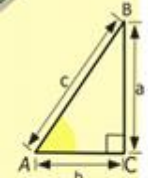


Поняття частотного аналізу

Частотний аналіз, частотний криптоаналіз – метод криптоаналізу, який ґрунтується на частоті появи знаків шифротексту. Власне – на припущенні про існування нетривіального статистичного розподілу окремих символів і їх послідовностей як у відкритому тексті, так і в шифротексті, який, з точністю до заміни символів, буде зберігатися в процесі шифрування і дешифрування.

Спрощено, частотний аналіз передбачає, що частота появи заданої літери алфавіту в досить довгих текстах одна і та ж для різних текстів однієї мови. При цьому у випадку моноалфавітного шифрування якщо в шифротексті буде символ з аналогічною ймовірністю появи, то можна припустити, що він і є зазначеною зашифрованою буквою. Аналогічні міркування застосовуються до біграм (двобуквених послідовностей), триграм і т.д. у разі поліалфавітного шифрів.

Метод частотного криптоаналізу відомий з IX-го століття (роботи Ал-Кінді), хоча найвідомішим випадком його застосування в реальному житті, можливо, є дешифровка єгипетських ієрогліфів Ж.-Ф. Шампольон у 1822 році.



$$\begin{array}{r} 1 \\ \times 2500 \\ \hline 2500 \\ \times 42 \\ \hline 210 \\ + 84 \\ \hline 10500 \end{array}$$



$$\frac{a}{A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

$$\frac{a}{c} + \frac{b}{c} = \frac{a+b}{c}$$

$$\sin 90^\circ = 1$$

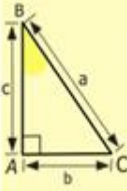
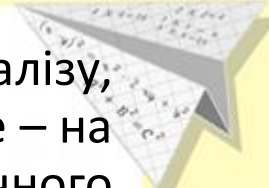


$$\begin{cases} y = \sin 90 \\ x = 25y + 45 \end{cases}$$

$$\begin{cases} y = 1 \\ x = 25 + 45 \end{cases}$$

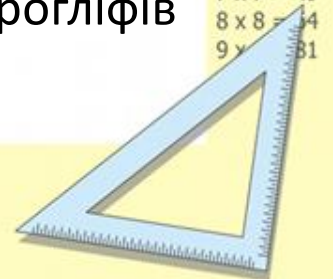
$$x = 70$$

$$(x+y)(x-y) = x^2 - y^2$$



$$y = \cos$$

$$\begin{array}{l} 2 \times 2 = 4 \\ 3 \times 3 = 9 \\ 4 \times 4 = 16 \\ 5 \times 5 = 25 \\ 6 \times 6 = 36 \\ 7 \times 7 = 49 \\ 8 \times 8 = 64 \\ 9 \times 9 = 81 \end{array}$$

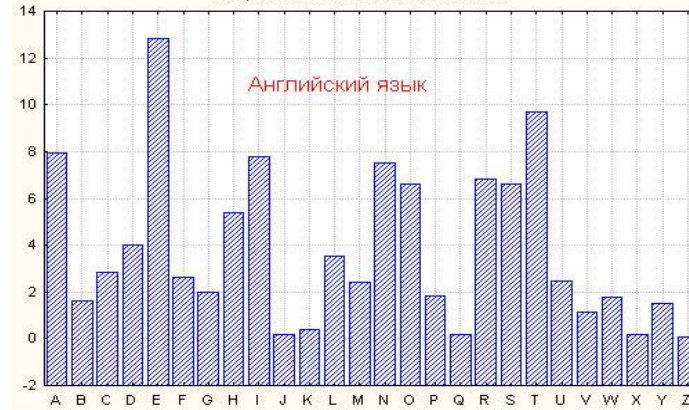


Гістограми частот

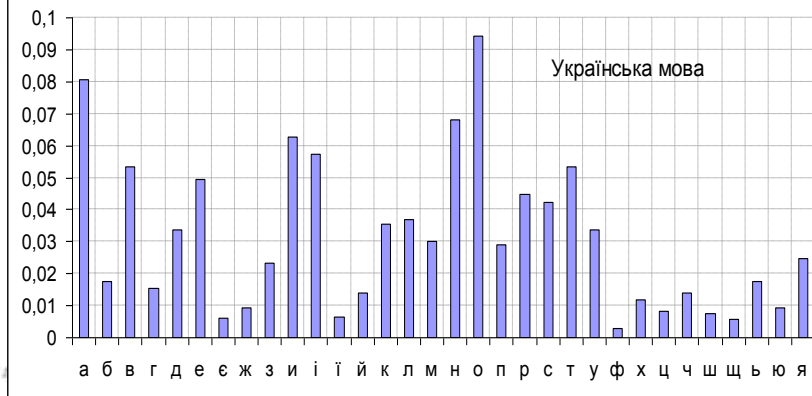
Гістограма частот використання букв алфавіту російської мови



Диаграмма частоты использования букв алфавита в Английском языке

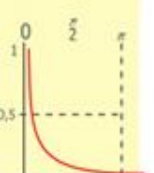
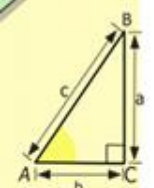
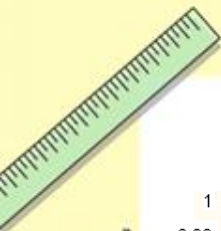


Гістограма частот використання букв алфавіту української мови



Порівнюючи гістограми частот використання букв російської, української та європейських мов, помітно різницю у їх розподілі, яка досягає десяти відсотків.

Даний факт свідчить про ймовірне виникнення методичної похибки визначення слогової та словесної розбірливості при використанні російськомовних артикуляційних таблиць для оцінки якості україномовних каналів мовленевої комунікації.



$y = 1/x$

$$\begin{array}{r} 2500 \\ \times 42 \\ \hline 2100 \\ + 8400 \\ \hline 105000 \end{array}$$



$$\frac{a}{A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

$$\frac{a}{c} + \frac{b}{c} = \frac{a+b}{c}$$

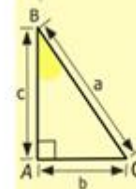
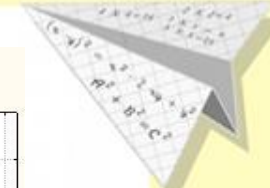
$\sin 90^\circ = 1$



$$\begin{cases} y = \sin 90 \\ x = 25y + 45 \end{cases}$$

$$\begin{cases} y = 1 \\ x = 25 + 45 \\ \hline x = 70 \end{cases}$$

$$(x+y)(x-y) = x^2 - y^2$$



$y = \cos$

- $2 \times 2 = 4$
- $3 \times 3 = 9$
- $4 \times 4 = 16$
- $5 \times 5 = 25$
- $6 \times 6 = 36$
- $7 \times 7 = 49$
- $8 \times 8 = 64$
- $9 \times 9 = 81$



РОЗДІЛ 2. ПРАКТИЧНА ЧАСТИНА

ЧАСТОТНИЙ АНАЛІЗ ВИКОРИСТАННЯ СИМВОЛІВ У ТЕКСТАХ ПО ЗАХИСТУ ІНФОРМАЦІЇ

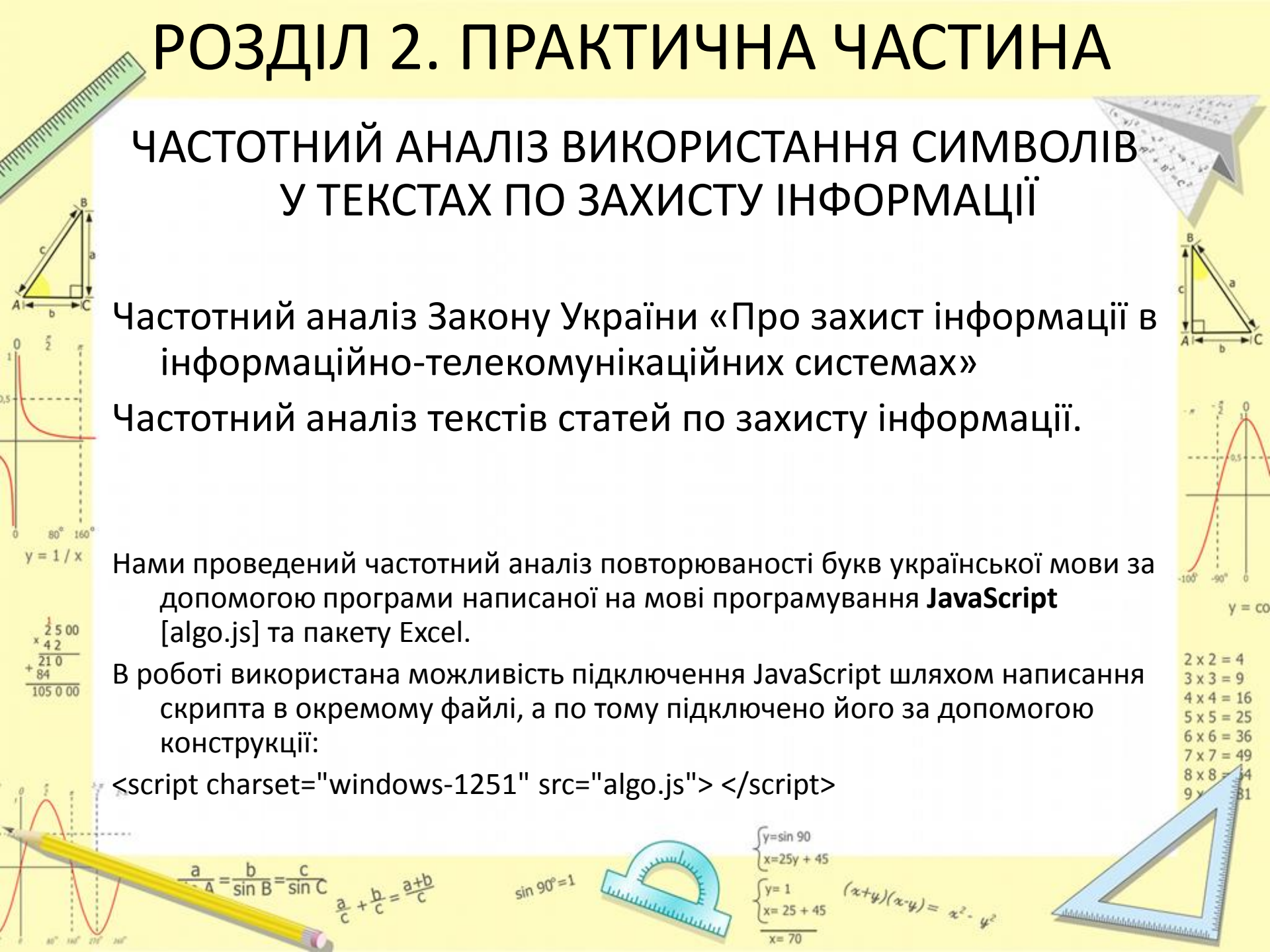
Частотний аналіз Закону України «Про захист інформації в інформаційно-телекомунікаційних системах»

Частотний аналіз текстів статей по захисту інформації.

Нами проведений частотний аналіз повторюваності букв української мови за допомогою програми написаної на мові програмування **JavaScript** [algo.js] та пакету Excel.

В роботі використана можливість підключення JavaScript шляхом написання скрипта в окремому файлі, а по тому підключено його за допомогою конструкції:

```
<script charset="windows-1251" src="algo.js"> </script>
```



СКРІНШОТ СТОРІНКИ САЙТУ

<http://shifr.ho.ua/>

підвищується із збільшенням довжини криптотексту. Існують безліч різних таблиць про розподіл букв у тому чи іншій мові, але жодна з них не містить остаточної інформації - навіть порядок букв може відрізнятися в різних таблицях. Розподіл букв залежить і від типу тексту: проза, розмовна мова, технічна мова і т.п.

Таблиці розподілу літер

В українській мові

| Буква | Частота | Буква | Частота | Буква | Частота |
|-------|---------|-------|---------|--------|---------|
| а | 0.062 | л | 0.035 | ц | 0.004 |
| б | 0.014 | м | 0.026 | ч | 0.012 |
| в | 0.038 | н | 0.053 | ш | 0.006 |
| г | 0.013 | о | 0.090 | щ | 0.003 |
| д | 0.025 | п | 0.023 | и | 0.016 |
| е | 0.072 | р | 0.040 | ь | 0.014 |
| ж | 0.007 | с | 0.045 | є | 0.003 |
| з | 0.016 | т | 0.053 | ю | 0.006 |
| і | 0.062 | у | 0.021 | я | 0.018 |
| й | 0.010 | ф | 0.002 | пробіл | 0.174 |
| к | 0.028 | х | 0.009 | | |

В англійській мові

| Буква | Частота | Буква | Частота | Буква | Частота |
|-------|---------|-------|---------|-------|---------|
| a | 0.0804 | b | 0.0154 | c | 0.0306 |
| d | 0.0399 | e | 0.1251 | f | 0.0230 |
| g | 0.0196 | h | 0.0549 | i | 0.0726 |
| j | 0.0016 | k | 0.0067 | l | 0.0414 |
| m | 0.0253 | n | 0.0709 | o | 0.0760 |
| p | 0.0200 | q | 0.0011 | r | 0.0612 |
| s | 0.0654 | t | 0.0925 | u | 0.0271 |
| v | 0.0099 | w | 0.0192 | x | 0.0019 |
| y | 0.0173 | z | 0.0009 | | |

Текст для оцінки

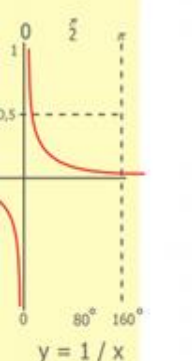
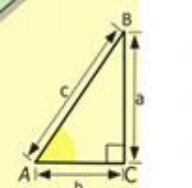
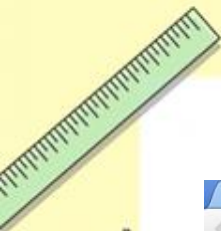
ЗАКОН УКРАЇНИ

Про захист інформації
в інформаційно-телекомунікаційних
системах

(Відомості Верховної Ради України (ВВР),
1994, N 31, ст.286)

Знайти частоти Очистити

Copyright (c) 2012 НПТ ліцею "Камелія"



$$\begin{array}{r} 12500 \\ \times 42 \\ \hline 210 \\ + 84 \\ \hline 105000 \end{array}$$



$$\frac{a}{A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

$$\frac{a}{c} + \frac{b}{c} = \frac{a+b}{c}$$

$$\sin 90^\circ = 1$$

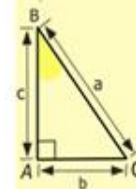


$$\begin{cases} y = \sin 90 \\ x = 25y + 45 \end{cases}$$

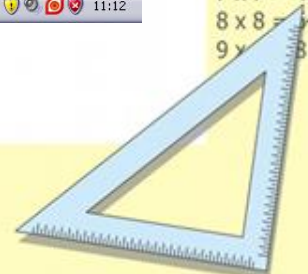
$$\begin{cases} y = 1 \\ x = 25 + 45 \end{cases}$$

$$(x+y)(x-y) = x^2 - y^2$$

$$x = 70$$



$$\begin{array}{l} 2 \times 2 = 4 \\ 3 \times 3 = 9 \\ 4 \times 4 = 16 \\ 5 \times 5 = 25 \\ 6 \times 6 = 36 \\ 7 \times 7 = 49 \\ 8 \times 8 = 64 \\ 9 \times 9 = 81 \end{array}$$

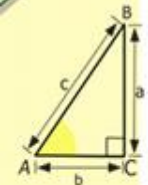
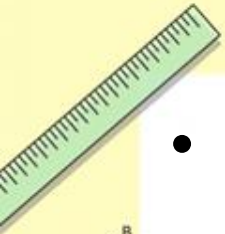


Наші дослідження

- Частотний аналіз Закону України "Про захист інформації в інформаційно-телекомунікаційних системах» **суттєво відрізняється** від отриманих результатів описаних в першому розділі роботи.

Це зумовлено надмірним використанням клавіші пробіл при наборі тексту Закону, який розміщено на сайті Верховної Ради України.

- При тестуванні текстів сайту Міністерства юстиції України (наприклад: Г.М. Красноступ «Організаційно-правове забезпечення захисту інформації про особу») отримуємо результат **близький** до досліджень в Розділі 1.



$$\begin{array}{r} 2500 \\ \times 42 \\ \hline 210 \\ + 84 \\ \hline 10500 \end{array}$$



$$\frac{a}{A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

$$\frac{a}{c} + \frac{b}{c} = \frac{a+b}{c}$$

$$\sin 90^\circ = 1$$

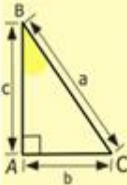


$$\begin{cases} y = \sin 90 \\ x = 25y + 45 \end{cases}$$

$$\begin{cases} y = 1 \\ x = 25 + 45 \end{cases}$$

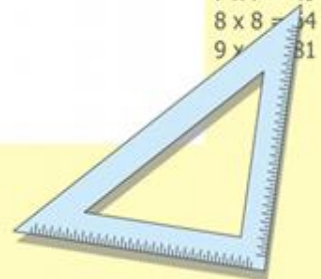
$$x = 70$$

$$(x+y)(x-y) = x^2 - y^2$$



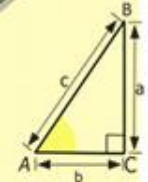
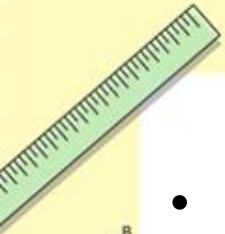
$$y = \cos$$

$$\begin{array}{l} 2 \times 2 = 4 \\ 3 \times 3 = 9 \\ 4 \times 4 = 16 \\ 5 \times 5 = 25 \\ 6 \times 6 = 36 \\ 7 \times 7 = 49 \\ 8 \times 8 = 64 \\ 9 \times 9 = 81 \end{array}$$

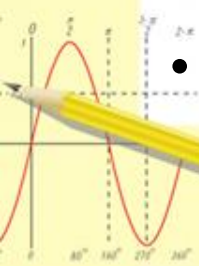


ВИСНОВКИ

- Сформульована науково-технічна задача створення артикуляційних таблиць для української мови.
- Проведений частотний аналіз повторюваності букв української мови за допомогою програми написаної на мові програмування JavaScript.
- Виконано необхідний проміжний етап цієї задачі, що забезпечує лінгвістичні вимоги до артикуляційних таблиць (відображення структури мови), – побудовано гістограму частот використання букв алфавіту сучасної української мови.
- Отримана діаграма дозволяє скласти кілька артикуляційних таблиць з різним характером наборів слів (в залежності від тематичного спрямування) для проведення подальших артикуляційних випробувань. У майбутньому необхідно зробити фонемний частотний аналіз, також бажаним є мовленевий аналіз біграм і триграм.
- URL сайту: <http://shifr.ho.ua/>



$$\begin{array}{r} 12500 \\ \times 42 \\ \hline 210 \\ + 84 \\ \hline 10500 \end{array}$$



$$\frac{a}{A} = \frac{b}{\sin B} = \frac{c}{\sin C}$$

$$\frac{a}{c} + \frac{b}{c} = \frac{a+b}{c}$$

$$\sin 90^\circ = 1$$

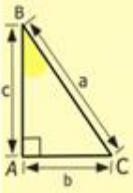


$$\begin{cases} y = \sin 90 \\ x = 25y + 45 \end{cases}$$

$$\begin{cases} y = 1 \\ x = 25 + 45 \end{cases}$$

$$x = 70$$

$$(x+y)(x-y) = x^2 - y^2$$



$$\begin{array}{l} 2 \times 2 = 4 \\ 3 \times 3 = 9 \\ 4 \times 4 = 16 \\ 5 \times 5 = 25 \\ 6 \times 6 = 36 \\ 7 \times 7 = 49 \\ 8 \times 8 = 64 \\ 9 \times 9 = 81 \end{array}$$

