НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені Ігоря СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра математичних методів захисту інформації

Комп'ютерний практикум №2
з курсу
Методи криптоаналізу 1

Підготували:
студенти 5 курсу
групи ФІ-22мн
Ковальчук О.М.
Коломієць А.Ю.

Київ – 2022

СТАТИСТИЧНІ КРИТЕРІЇ НА ВІДКРИТИЙ ТЕКСТ

Мета лабораторної роботи

Засвоєння статистичних методів розрізнення змістовного тексту від випадкової послідовності, порівняння їх, визначення похибок першого та другого роду.

Постановка задачі та варіанти завдання

Номер варіанту бригади – *одинадцятий*. Постановка задачі, описана наступними пунктами.

1) На великому тексті українською мовою (>1MB), літеру «ґ» замінити на літеру «г», видалити символи апострофу та усі інші спецсимволи в тексті, включно з пробілами, текст повинен містить лише маленькі літери алфавіту. Необхідно розрахувати частоти літер і біграм, а також ентропію та індекс відповідності.

2) Отримати N текстів X українською мовою для довжин $L = 10, 100, 1000, 10000$, для кожного з яких згенерувати спотворені тексти Y. Число N визначається відповідно до такої таблиці.

L	N
10 100 1000	10000
10000	1000

Спотворення тексту виконується такими способами:

(а) шляхом застосування шифру Віженера з випадковим ключем довжини $r = 1, 5, 10$:

$$y_i = \left(x_i + Key_{(i \bmod r)}\right) \bmod m;$$

(б) шляхом застосування шифру афінної та афінної біграмної підстановки з випадковими ключами, де a,$b \in (Z_m)^l$ – ключі:

$$y_i = (a \cdot x_i + b) \bmod m^l;$$

(в) y_i — рівномірно розподілена послідовність символів з $(Z_m)^l$;

(г) y_i обчислюється відповідно до такого співвідношення, де $s_0, s_1 \in_R (Z_m)^l$:

$$y_i = (s_{i-1} + s_{i-2}) \bmod m^l.$$

3) Реалізувати критерії $2.0 - 2.3$, 4.0, 5.0 + структурний і перевірити їх роботу на згенерованих N текстах для кожної довжини L. Розрахувати ймовірності похибок першого і другого роду. Усі вищезгадані критерії (та інші формули), які використовували значення l, мають приймати значення $l = 1$ та $l = 2$, тобто реалізувати символьний та біграмний критерії.

4) Згенерувати випадковий текст довжини $L = 10000$, який точно не є зв'язним текстом українською мовою (наприклад, текст, який складається з величезної кількості літер а: *«аааааааа...»*). Застосувати один з варіантів спотворення (на вибір) до цього тексту, після чого застосувати один з реалізованих критеріїв (на вибір). Порівняти результати застосування критерію до різних текстів.

5) Здійснити опис множин заборонених/частих символів, які було отримано при виконанні завдання;

6) Створити окремі таблиці для кожного зі способів спотворення, які містять ймовірності похибок першого та другого роду для кожного з критеріїв, що реалізовуються в роботі, для різних значень L та l;

7) Для кожного критерію вказати значення порогових значень, що було підібрано при реалізації (наприклад, в дужках в таблиці біля номера критерію, як вказано в шаблоні таблиці);

8) Здійснити опис алгоритму стиснення, що був обраний для розробки структурного критерію;

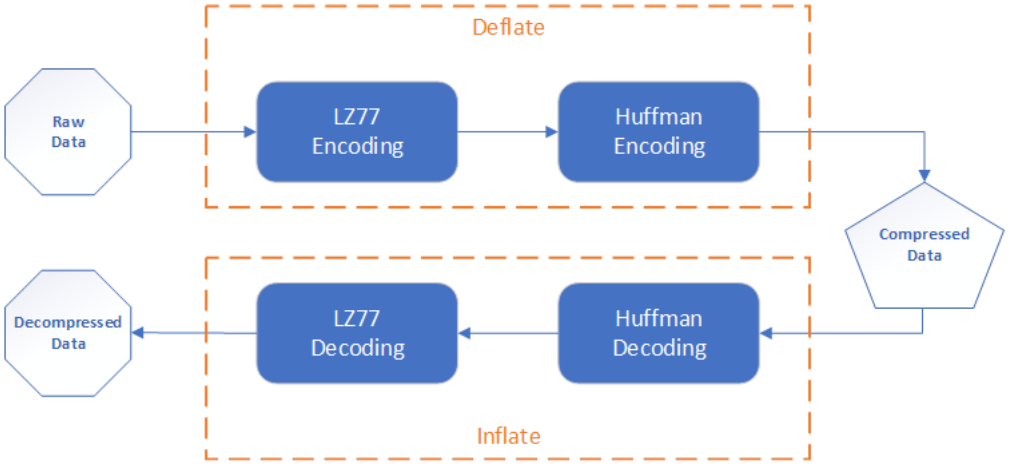
9) Здійснити опис запропонованого структурного критерію, що базується на основі результатів стиснення;

10) Зробити висновки (аналіз ефективності реалізованих критеріїв, порівняння їх між собою, порівняння результатів для різних значень L, r тощо).

Хід роботи

Опис обраного алгоритму стиснення

У лабораторній роботі був застосований структурний критерій на основі роботи *zip* архіваторів, що входить до бібліотеки *zlib* — це безкоштовна бібліотека програмного забезпечення з відкритим кодом для стиснення даних без втрат і декомпресії. В zlib використовується алгоритм стиснення через метод DEFLATE. Метод DEFLATE кодує відкритий текст в текст, що дуже стиснутий і легко зберігати. Метод DEFLATE це комбінація алгоритму LZ77 і кодування Хаффмана. Принцип роботи зображено на рисунку.



Стиснення та відновлення даних.

LZ77 (інша назва - LZ1) — це алгоритм стиснення без втрат на основі словника. Основна ідея – замінити певну послідовність байтів у даних посиланням на попередню появу цієї послідовності.

Даний алгоритм працює наступним чином:

– за допомогою sliding window (ковзаючого вікна) шукає послідовності даних, які повторюються;

– кодує кожну повторювану послідовність парою чисел, яка називається length-distance pair (пара довжина-відстань).

Кодування Хаффмана – це метод статистичного стиснення. За цим методом символи даних кодуються за допомогою кодів змінної довжини, а довжини кодів базуються на частотах відповідних символів.

Даний метод має дві властивості:

– Коди символів даних, які зустрічаються частіше, є коротшими за коди символів даних, які зустрічаються рідше.

– Кожен код може бути однозначно розшифрований. Тобто будь-який код для одного символу не є префіксом кодів для інших символів.

Щоб краще зрозуміти суть другої властивості розглянемо невеликий приклад. Нехай код «0» використовується для символу «A». Тоді для кодування символу «B» ми не можемо використовувати код «01», оскільки код «0» є префіксом коду «01». Друга властивість гарантує, що при декодуванні немає двозначності у визначенні меж символів.

Опис роботи структурного алгоритму

Структурний критерій на основі *zip*, який ми реалізували в комп'ютерному практикумі складається з наступних кроків виконання:

1) Застосувати обраний алгоритм стиснення до вхідного тексту;

2) Застосувати обраний алгоритм стиснення до випадкового тексту, який був отриманий в результаті генерування рівномірно розподіленої послідовності символів із $(Z_m)^l$;

3) Обчислити співвідношення довжини вхідного тексту до довжини стисненого вхідного тексту;

4) Обчислити співвідношення довжини випадкового тексту до довжини стисненого випадкового тексту;

5) Якщо модуль різниці цих співвідношень менший за порогове значення, то приймається гіпотеза H_1. Інакше приймається гіпотеза H_0.

Таблиця 1 – Застосування одного з критеріїв, до випадково згенерованого тексту.

L	FP	FN
10	0.9851	0.0168
100	0.1198	0.8859
1000	0.0	1.0
10000	0.0	1.0

Таблиця 2 – Найбільш часті l-грами обрані для різних критеріїв.

Номер критерію	Монограми	Біграми
2.0	['а', 'о', 'и']	['на', 'ла'], ['на', 'ла', 'та', 'ав', 'ка', 'ов', 'ал', 'не', 'ли', 'по']
2.1	['а', 'о', 'и'], ['а', 'о', 'и', 'в', 'т', 'н', 'е', 'і', 'л', 'с']	['на', 'ла', 'та'] ['на', 'ла', 'та', 'ав', 'ка', 'ов', 'ал', 'не', 'ли', 'по']
2.2	['а', 'о', 'и']	['на', 'ла', 'та'], ['на', 'ла', 'та', 'ав', 'ка', 'ов', 'ал', 'не', 'ли', 'по']
2.3	['а', 'о', 'и']	['на', 'ла', 'та'], ['на', 'ла', 'та', 'ав', 'ка', 'ов', 'ал', 'не', 'ли', 'по']
5	['ф', 'є', 'ї', 'щ', 'ц']	['гї', 'хе', 'дє', 'цд', 'цй', 'чї', 'цж', 'кє', 'цч', 'цз', 'уи', 'цш', 'цб', 'жь', 'цх', 'хф', 'нф', 'їф', 'пф', 'лй', 'щц', 'пї', 'щї', 'фу', 'рє', 'кю', 'цб', 'щч','щщ', 'йф','щз', 'щх', 'їю', 'шй', 'пд', 'щл', 'бь', 'лш', 'кф', 'іф', 'ее','пщ', 'іи', 'чж', 'йю', 'гх', 'щв', 'щм', 'жє', 'ье', 'тя', 'їє', 'гж', 'чй', 'зф','уф', 'лч', 'шш', 'йи', 'ьи', 'вь', 'пь', 'ью', 'гщ', 'лг', 'щс', 'жш', 'сж', 'пв', 'лх', 'щк', 'шї', 'гг', 'цс', 'сш', 'цк', 'чь', 'сї','фи', 'йе', 'яи', 'юф', 'сй', 'юе', 'щн', 'пг', 'иф', 'мє', 'лр', 'чш', 'ьф', ' фр', 'пб', 'шж', 'щи', 'хй', 'пм', 'сщ', 'щд', 'тш']

Таблиця 3 – Спотворення за допомогою шифру Віженера ($r = 1$).

L	*Номер критерію*	FP ($l = 1$)	FN ($l = 1$)	FP ($l = 2$)	FN ($l = 2$)
10	2.0	0,8492	0,0136	0,9915	0,0000
	2.1 (1, 1)	0,6559	0,8372	0,0274	0,9998
	2.2	0,8492	0,0136	0,9995	0,0000
	2.3	0,5024	0,0692	0,7021	0,0256
	4 (0.1, 0.1)	0,0085	0,9915	0,0002	0,9999
	5 (1, 1)	0,0000	0,9862	0,0000	1,0000
	Структурний (0.0085, 0.0085)	0,3826	0,6524	0,3779	0,6478
100	2.0	0,0006	0,5105	0,9900	0,0001
	2.1 (6, 6)	1,0000	0,0002	0,9675	0,9894
	2.2	0,9284	0,0024	0,9992	0,0001
	2.3	0,4482	0,0178	0,5631	0,0009
	4 (0.02, 0.002)	0,0014	0,9986	0,1138	0,5700
	5 (4, 4)	0,9045	0,0037	0,0000	1,0000
	Структурний (0.09, 0.09)	0,9605	0,0340	0,9649	0,0010
1000	2.0	0,0000	0,9628	0,0027	0,0013
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,3430
	2.2	0,9213	0,0026	0,9997	0,0000
	2.3	0,4817	0,0163	0,5061	0,0001
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,8920
	5 (4, 4)	1,0000	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,0000	1,0000	0,0000	0,9241
10000	2.0	0,0000	1,0000	0,0000	0,1620
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0000
	2.2	0,9220	0,0030	0,9990	0,0000
	2.3	0,4870	0,0170	0,4620	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,9250
	5 (4, 4)	1,0000	0,0000	0,0000	0,9980
	Структурний (0.09, 0.09)	0,0000	1,0000	0,0000	1,0000

Таблиця 4 – Спотворення за допомогою шифру Віженера ($r = 5$).

L	***Номер критерію***	**FP** ($l = 1$)	**FN** ($l = 1$)	**FP** ($l = 2$)	**FN** ($l = 2$)
10	2.0	0,8492	0,0163	0,9915	0,0000
	2.1 (1, 1)	0,6559	0,8342	0,0274	0,9995
	2.2	0,8492	0,0163	0,9995	0,0000
	2.3	0,5024	0,0612	0,7021	0,0248
	4 (0.1, 0.1)	0,0085	0,9957	0,0002	1,0000
	5 (1, 1)	0,0000	0,9860	0,0000	1,0000
	Структурний (0.0085, 0.0085)	0,3794	0,6223	0,3727	0,6147
100	2.0	0,0006	0,8131	0,9900	0,0000
	2.1 (6, 6)	1,0000	0,0000	0,9675	0,9903
	2.2	0,9284	0,0000	0,9992	0,0000
	2.3	0,4482	0,0000	0,5631	0,0000
	4 (0.02, 0.002)	0,0014	0,9591	0,1138	0,0475
	5 (4, 4)	0,9045	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,9626	0,0001	0,9641	0,0000
1000	2.0	0,0000	1,0000	0,0027	0,0051
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0911
	2.2	0,9213	0,0000	0,9997	0,0000
	2.3	0,4817	0,0000	0,5061	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,0000	0,0036	0,0000	0,0000
10000	2.0	0,0000	1,0000	0,0000	0,8720
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0000
	2.2	0,9220	0,0000	0,9990	0,0000
	2.3	0,4870	0,0000	0,4620	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	0,0240
	Структурний (0.09, 0.09)	0,0000	1,0000	0,0000	0,0200

Таблиця 5 – Спотворення за допомогою шифру Віженера ($r = 10$)

L	***Номер критерію***	**FP** ($l = 1$)	**FN** ($l = 1$)	**FP** ($l = 2$)	**FN** ($l = 2$)
10	2.0	0,8492	0,0154	0,9915	0,0000
	2.1 (1, 1)	0,6559	0,8290	0,0274	0,9996
	2.2	0,8492	0,0154	0,9995	0,0000
	2.3	0,5024	0,0592	0,7021	0,0267
	4 (0.1, 0.1)	0,0085	0,9976	0,0002	1,0000
	5 (1, 1)	0,0000	0,9878	0,0000	1,0000
	Структурний (0.0085, 0.0085)	0,3825	0,6097	0,3747	0,6231
100	2.0	0,0006	0,8550	0,9900	0,0000
	2.1 (6, 6)	1,0000	0,0000	0,9675	0,9917
	2.2	0,9284	0,0000	0,9992	0,0000
	2.3	0,4482	0,0000	0,5631	0,0000
	4 (0.02, 0.002)	0,0014	0,8646	0,1138	0,0176
	5 (4, 4)	0,9045	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,9600	0,0000	0,9639	0,0000
1000	2.0	0,0000	1,0000	0,0027	0,0076
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0630
	2.2	0,9213	0,0000	0,9997	0,0000
	2.3	0,4817	0,0000	0,5061	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,0000	0,0000	0,0000	0,0000
10000	2.0	0,0000	1,0000	0,0000	0,9660
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0000
	2.2	0,9220	0,0000	0,9990	0,0000
	2.3	0,4870	0,0000	0,4620	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	0,0000
	Структурний (0.09, 0.09)	0,0000	0,0340	0,0000	0,0000

Таблиця 6 – Спотворення за допомогою афінного шифру.

L	***Номер критерію***	**FP** ($l = 1$)	**FN** ($l = 1$)	**FP** ($l = 2$)	**FN** ($l = 2$)
10	2.0	0,8492	0,0187	0,9915	0,0000
	2.1 (1, 1)	0,6559	0,8199	0,0274	0,9998
	2.2	0,8492	0,0187	0,9995	0,0000
	2.3	0,5024	0,0636	0,7021	0,0278
	4 (0.1, 0.1)	0,0085	0,9969	0,0002	1,0000
	5 (1, 1)	0,0000	0,9855	0,0000	1,0000
	Структурний (0.0085, 0.0085)	0,3812	0,6177	0,3736	0,6140
100	2.0	0,0006	0,8890	0,9900	0,0000
	2.1 (6, 6)	1,0000	0,0000	0,9675	0,9912
	2.2	0,9284	0,0000	0,9992	0,0000
	2.3	0,4482	0,0000	0,5631	0,0000
	4 (0.02, 0.002)	0,0014	0,7657	0,1138	0,0063
	5 (4, 4)	0,9045	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,9596	0,0000	0,9628	0,0000
1000	2.0	0,0000	1,0000	0,0027	0,0109
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0377
	2.2	0,9213	0,0000	0,9997	1,0000
	2.3	0,4817	0,0000	0,5061	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,0000	0,0000	0,0000	0,0000
10000	2.0	0,0000	1,0000	0,0000	1,0000
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0000
	2.2	0,9220	0,0000	0,9990	0,0000
	2.3	0,4870	0,0000	0,4620	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	0,0000
	Структурний (0.09, 0.09)	0,0000	0,0000	0,0000	0,0000

Таблиця 7 – Спотворення за допомогою рівномірної генерації.

L	***Номер критерію***	**FP** ($l = 1$)	**FN** ($l = 1$)	**FP** ($l = 2$)	**FN** ($l = 2$)
10	2.0	0,8492	0,0157	0,9915	0,0000
	2.1 (1, 1)	0,6559	0,8284	0,0274	0,9998
	2.2	0,8492	0,0157	0,9995	0,0000
	2.3	0,5024	0,0566	0,7021	0,0237
	4 (0.1, 0.1)	0,0085	0,9961	0,0002	1,0000
	5 (1, 1)	0,0000	0,9856	0,0000	1,0000
	Структурний (0.0085, 0.0085)	0,3892	0,6003	0,3680	0,6289
100	2.0	0,0006	0,8846	0,9900	0,0000
	2.1 (6, 6)	1,0000	0,0000	0,9675	0,9934
	2.2	0,9284	0,0000	0,9992	0,0000
	2.3	0,4482	0,0000	0,5631	0,0000
	4 (0.02, 0.002)	0,0014	0,7628	0,1138	0,0121
	5 (4, 4)	0,9045	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,9599	0,0000	0,9622	0,0000
1000	2.0	0,0000	1,0000	0,0027	0,0058
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0561
	2.2	0,9213	0,0000	0,9997	0,0000
	2.3	0,4817	0,0000	0,5061	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,0000	0,0000	0,0000	0,0000
10000	2.0	0,0000	1,0000	0,0000	1,0000
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0000
	2.2	0,9220	0,0000	0,9990	0,0000
	2.3	0,4870	0,0000	0,4620	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	0,0000
	Структурний (0.09, 0.09)	0,0000	0,0000	0,0000	0,0000

Таблиця 8 – Спотворення за допомогою рекурсивного перетворення.

L	***Номер критерію***	**FP** ($l = 1$)	**FN** ($l = 1$)	**FP** ($l = 2$)	**FN** ($l = 2$)
10	2.0	0,8492	0,0167	0,9915	0,0000
	2.1 (1, 1)	0,6559	0,8167	0,0274	1,0000
	2.2	0,8492	0,0167	0,9995	0,0000
	2.3	0,5024	0,0684	0,7021	0,0192
	4 (0.1, 0.1)	0,0085	0,9954	0,0002	0,9999
	5 (1, 1)	0,0000	0,9841	0,0000	1,0000
	Структурний (0.0085, 0.0085)	0,3876	0,6086	0,3689	0,6577
100	2.0	0,0006	0,8942	0,9900	0,0000
	2.1 (6, 6)	1,0000	0,0000	0,9675	0,9938
	2.2	0,9284	0,0000	0,9992	0,0000
	2.3	0,4482	0,0000	0,5631	0,0000
	4 (0.02, 0.002)	0,0014	0,8814	0,1138	0,1152
	5 (4, 4)	0,9045	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,9609	0,0000	0,9633	0,1306
1000	2.0	0,0000	1,0000	0,0027	0,0002
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,2560
	2.2	0,9213	0,0000	0,9997	0,0000
	2.3	0,4817	0,0000	0,5061	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	1,0000
	Структурний (0.09, 0.09)	0,0000	0,0000	0,0000	1,0000
10000	2.0	0,0000	1,0000	0,0000	0,9380
	2.1 (6, 6)	1,0000	0,0000	1,0000	0,0000
	2.2	0,9220	0,0000	0,9990	0,0000
	2.3	0,4870	0,0000	0,4620	0,0000
	4 (0.02, 0.002)	0,0000	1,0000	0,0000	0,0000
	5 (4, 4)	1,0000	0,0000	0,0000	0,0000
	Структурний (0.09, 0.09)	0,0000	0,0000	0,0000	1,0000

Опис труднощів, що виникали при виконанні комп'ютерного практикуму, та шляхи їх розв'язання

Під час виконання комп'ютерного практикуму, перша проблема, що виникла при реалізації полягала в тому, як саме зберігати зручно дані, тобто нарізані тексти, та їхні спотворення. Рішенням цього було створення детафрейму зі списків нарізаних текстів. Кожен елемент детафрейму був списком. А список складався з текстів l –грам відповідної довжини. Друга проблема, яка постала перед нами – реалізувати так програмний код, щоб було легко застосувати функції спотворення, до нарізаних текстів, та статистичні критерії. Дана проблема вирішується використанням методу map, до певного списку детафрейму. При цьому написані нами функції були призначені для однієї l – грами, але використання map дозволяє застосувати наші написані функції, до всіх елементів списку, що знаходиться в детафреймі. Таким, чином було дійсно зекономлено багато на написанні програмного коду.

Після написання функцій спотворення і структурних критеріїв, і їх застосування на детафреймах, стало очевидно, що можна також легко знайти помилки першого та другого роду для одного списку детафрейму, а потім скориставшись функціоналом map, дану процедуру повторити для всіх списків-елементів детафрейму, в результаті чого ми одержуємо масив з помилками.

Висновок

У комп'ютерному практикумі було розглянуто та реалізовано деякі статистичні методи розрізнення змістовного тексту від випадкової послідовності. А саме:

1) критерій частих l-грам та його варіації,
2) критерій через розрахунок індексу відповідності,
3) критерій порожніх ящиків для l-грам, які зустрічаються найрідше
4) структурний критерій

Кожен критерій був реалізований як для одно-грам (один символ), так і для біграм. Для кожного критерію були обчислені та занесені у таблицю значення помилок першого та другого роду. Можемо зробити певні висновки щодо ефективності розглянутих критеріїв.

У випадку, коли ми розглядаємо критерії найчастіших l-грам для біграм можемо помітити певну залежність. Чим більша довжина тексту, тим менше значення помилки першого роду. Тобто критерії частіше правильно розпізнають змістовний текст.

У випадку ж, коли ми розглядаємо критерії найчастіших l-грам для одного символу ситуація гірша. Чим більша довжина тексту, тим більше значення помилки першого роду. Тобто критерії частіше розпізнають змістовний текст як випадковий.

Критерій через розрахунок індексу відповідності виявився найлегшим у реалізації та показав хороші результати на текстах великої довжини. Цей результат є очікуваним, оскільки 10 (для одно-грам) та 100 (для біграм) символів недостатньо для того, щоб на основі значень індексу відповідності можна було зробити якісь висновки.