



**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**  
**імені Ігоря Сікорського»**  
**«ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ»**

**КРИПТОГРАФІЯ**

*Комп'ютерний практикум №1*

Виконали:  
студенти 3-го курсу  
групи ФБ-22  
Власенко Г. В. та  
Перебинос Р. О.  
Бригада №2  
Перевірів/-ла:

---

---

Київ – 2024

# Експериментальна оцінка ентропії на символ джерела відкритого тексту

## Мета роботи

Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

## Порядок виконання роботи

0. Уважно прочитати методичні вказівки до виконання комп'ютерного практикуму.
1. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку  $H_1$  та  $H_2$  за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення  $H_1$  та  $H_2$  на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення  $H_1$  та  $H_2$  на тому ж тексті, в якому вилучено всі пробіли.
2. За допомогою програми CoolPinkProgram оцінити значення  $H^{(10)}$ ,  $H^{(20)}$ ,  $H^{(30)}$ .
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

## Хід роботи

На початку роботи розробленої (згідно порядку виконання роботи п.1) програми приймаємо шлях до вхідного файлу у якості аргументу командного рядка.

```
argparser = argparse.ArgumentParser(prog="Entropy calculator",
                                   description="This script calculates the
specific entropy for monograms and bigrams of the given text.",
                                   usage="main.py pathToFile")
argparser.add_argument('path', nargs='?', help='path to text file')
```

Створюємо два об'єкти класу *EntropyCalculator* для підрахунку необхідних значень як для тексту із пробілами, так і без.

```
generalCalc = EntropyCalculator(STANDARD_ALPHABET)
whitespacedCalc = EntropyCalculator(STANDARD_ALPHABET_WHITESPACE)
```

Далі вносимо дані з файлу до створених об'єктів, де текст проходить попередню обробку: підраховуємо кількість кожної моно- та біграми у тексті, ігноруючи символи поза алфавітом та пробіли, що стоять один за одним, а також загальну кількість моно-/біграм (біграм як тих, що перетинаються, так і тих, що не перетинаються).

```
readTextWithHandler(fileName, generalCalc.handleText)
readTextWithHandler(fileName, whitespacedCalc.handleText)
```

```
# Read text from file by blocks and process by handler.
# Handler should be object method.
def readTextWithHandler(fileName: str, handler: Callable[[str], None]):
    with open(fileName, "r", encoding="utf-8") as f:
        textBlock = f.read(TEXT_BLOCK_SIZE)
        handler(textBlock)
```

```
def handleText(self, text: str) -> None:
    for c in text:
        # All characters not included in the alphabet are considered
        whitespace
        if c not in self.alphabet or (not self.whitespace is None and
c == self.whitespace):
            # If whitespace not provided in alphabet just skip it.
            if self.whitespace is None:
                continue

            # We need to skip, when two or more whitespaces in row,
            # since there can be only one whitespace
            if self.isWhitespace:
                continue

            self.updateNgramCounts(self.whitespace)
            self.isWhitespace = True
            self.previousSymbol = self.whitespace
            continue

        self.isWhitespace = False
        self.updateNgramCounts(c)
        self.previousSymbol = c
```

Після цього розраховуємо частоти  $n$ -грам відповідно до отриманих їхніх кількостей у тексті за такою формулою:

$$frequency(a) = \frac{occurences(a)}{totalNgrams(A)},$$

де  $A$  - це набір усіх  $n$ -грам;  $a$  - поточна  $n$ -грама,  $a \in A$ ;  $occurences()$  у даному випадку відображає кількість  $n$ -грам  $a$  у  $A$ , а  $totalNgrams()$  – загальну кількість  $n$ -грам у  $A$ .

```
def calculateFrequency(ngramCount: dict[str, int], totalNgrams: int) ->
dict[str, Decimal]:
    # Since log(0) is undefined, we add one to the amount of each ngram
    freqs = {ngram: Decimal(count) / Decimal(totalNgrams) for ngram, count
in ngramCount.items()}
    return dict(sorted(freqs.items(), key=lambda item: item[1],
```

```
reverse=True) )
```

Для розрахунків, як можна помітити, ми вирішили використовувати тип даних [\*Decimal\*](#), оскільки під час підрахунків він дає більшу точність за стандартний тип даних *python – float*.

Тепер, маючи частоти кожної  $n$ -грами, можемо розрахувати ентропію для кожного з випадків. Розраховуємо  $H_1$  за такою формулою:

$$H_1 = - \sum_{i=1}^n p_i \log_2 p_i,$$

де  $n$  – кількість символів у тексті;  $p_i$  – частота  $i$ -того символу.

$H_2$ :

$$H_2 = \frac{- \sum_{i=1}^n p_i \log_2 p_i}{2},$$

що слідує із загальної формули

$$H_n = \frac{1}{n} H(x_1, x_2, \dots, x_n),$$

де  $H(x_1, x_2, \dots, x_n)$  – ентропія  $n$ -грами відкритого тексту  $(x_1, x_2, \dots, x_n)$ .

```
def calculateEntropy(frequencies: dict[str, Decimal]) -> Decimal:
    entropy = Decimal(0)
    for freq in frequencies.values():
        if not freq.is_zero():
            entropy -= freq * Decimal(log2(freq))
    return entropy / Decimal(len(list(frequencies.keys())[0]))
```

Маючи значення ентропії, розраховуємо надлишковість:

```
def sourceRedundancy(entropy: Decimal, symbolsCount: int) -> Decimal:
    return Decimal(1) - (entropy / Decimal(log2(symbolsCount)))
```

У результаті розрахунків виводимо результати до консолі, а також зберігаємо у форматі .xlsx та .csv. Отже, отримані результати на тексті вагою 77.2Мб:

→ Консоль.

Без пробілів:

```
PS Q:\uni\CRYPTO\crypro-24-25\tasks\cp1\perebynos_fb-22_vlasenko_fb-22_cp1> python main.py test_data\big_text.txt
Specific entropy on monograms (without spaces): 4.3973356370483664235054175534915819544077925541229
Source redundancy on monograms (without spaces): 0.11240181826687102873658890180468212087626439660648
Monogram (without spaces) frequencies:
{'a': Decimal('0.10174050632911392405063291139240506329113924050633'),
 'o': Decimal('0.095393811533052039381153305203938115330520393811533'),
 'и': Decimal('0.094497187060478199718706047819971870604781997187060'),
 'н': Decimal('0.082419127988748241912798874824191279887482419127989'),
 'е': Decimal('0.073646272855133614627285513361462728551336146272855'),
 'т': Decimal('0.064205344585091420534458509142053445850914205344585'),
 'р': Decimal('0.052303094233473980309423347398030942334739803094233'),
 'в': Decimal('0.051687763713080168776371308016877637130801687763713'),
 'с': Decimal('0.050966947960618846694796061884669479606188466947961'),
 'з': Decimal('0.028533755274261603375527426160337552742616033755274'),
 'к': Decimal('0.027637130801687763713080168776371308016877637130802'),
 'п': Decimal('0.026951476793248945147679324894514767932489451476793'),
 'я': Decimal('0.025861462728551336146272855133614627285513361462729'),
 'ы': Decimal('0.024964838255977496483825597749648382559774964838256'),
 'л': Decimal('0.024472573839662447257383966244725738396624472573840'),
 'й': Decimal('0.024015471167369901547116736990154711673699015471167'),
 'д': Decimal('0.020798171589310829817158931082981715893108298171589'),
 'м': Decimal('0.019567510548523206751054852320675105485232067510549'),
 'у': Decimal('0.013836146272855133614627285513361462728551336146273'),
 'б': Decimal('0.013168073136427566807313642756680731364275668073136'),
 'г': Decimal('0.011620956399437412095639943741209563994374120956399'),
 'ч': Decimal('0.011286919831223628691983122362869198312236286919831'),
 'ж': Decimal('0.010601265822784810126582278481012658227848101265823'),
 'ц': Decimal('0.010196905766526019690576652601969057665260196905767'),
 'х': Decimal('0.0099859353023909985935302390998593530239099859353024'),
 'ш': Decimal('0.0085267229254571026722925457102672292545710267229255'),
 'ь': Decimal('0.0072257383966244725738396624472573839662447257383966'),
 'ф': Decimal('0.0057841068917018284106891701828410689170182841068917'),
 'ю': Decimal('0.0043073136427566807313642756680731364275668073136428'),
 'щ': Decimal('0.0027074542897327707454289732770745428973277074542897'),
 'э': Decimal('0.0010900140646976090014064697609001406469760900140647')}.

Specific entropy on bigrams (overlapped) (without spaces): 3.7711684644302670609271714141379160307693981093592
Source redundancy on bigrams (overlapped) (without spaces): 0.23879309010753043738315771925849461068576901023076

Specific entropy on bigrams (not overlapped/distinct) (without spaces): 3.7671712035959781992365904257311421851254213711386
Source redundancy on bigrams (not overlapped/distinct) (without spaces): 0.23959993355576204655848304734059667692926792433809
```

З пробілами:

```
Specific entropy on monogram characters (with spaces): 4.4039440603974067401040621719712890044555389197678
Source redundancy on monogram characters (with spaces): 0.11921118792051865197918756560574219910889221604644
Monogram (with spaces) frequencies:
{' ': Decimal('0.11671532393316355053108888750854090316168706130816'),
 'a': Decimal('0.089865830175787315982359152742406360643518230946021'),
 'o': Decimal('0.084259891918752717560096900428598049568296167463818'),
 'и': Decimal('0.083467917261941735511522454810857817255730169575750'),
 'н': Decimal('0.072799552767252624386607863842474687868811727436487'),
 'е': Decimal('0.065050624262376545127026523386545748183116963786571'),
 'т': Decimal('0.056711596993602087086154419529163302068451456612212'),
 'р': Decimal('0.046198521647307286166842661034846884899683210137276'),
 'в': Decimal('0.045655009627927200447232747375613392136157525312131'),
 'с': Decimal('0.045018324119510528604261134231939872041741723088391'),
 'э': Decimal('0.025203428784396546369339710541027393005776756320268'),
 'к': Decimal('0.024411454127585564320765264923287160693210758432201'),
 'п': Decimal('0.023805826448847754518914218274426983042424995341326'),
 'я': Decimal('0.022843033728803031244176656935213367289893782222498'),
 'ы': Decimal('0.022051059071992049195602211317473134977327784334431'),
 'л': Decimal('0.021616249456487980619914280390086340766507236474315'),
 'й': Decimal('0.021212497670662774085346915957512888999316727747065'),
 'д': Decimal('0.018370706255046897322815081682092055407168147089881'),
 'м': Decimal('0.017283682216286725883595254363625069880116777439593'),
 'у': Decimal('0.012221255978632213180942915709050251568420398782533'),
 'б': Decimal('0.011631157214733834399652152307596745139449655258091'),
 'г': Decimal('0.010264612708863904590347226535809677619727933412013'),
 'ч': Decimal('0.0099695633269147151997018448350829244052425616497919'),
 'ж': Decimal('0.0093639356481769053978507981862227467544567985589167'),
 'ц': Decimal('0.0090067706068699919249642834958693086527113485309647'),
 'х': Decimal('0.0088204236287968196782408845269892539909311137337723'),
 'ш': Decimal('0.0075315236971240449717373749922355425802844897198584'),
 'ь': Decimal('0.0063823839990061494502764146841418721659730418038387'),
 'ф': Decimal('0.0051090129821728057643331883967948319771414373563575'),
 'ю': Decimal('0.0038045841356606000372693956146344493446797937760109'),
 'щ': Decimal('0.0023914528852723771662836201006273681595130132306354'),
 'э': Decimal('0.00096279272004472327473756133921361575253121311882726')}}
Specific entropy on monogram characters (overlapped) (with spaces): 3.6769800830674192214769506922461615586254905678761
Source redundancy on monogram characters (overlapped) (with spaces): 0.26460398338651615570460986155076768827490188642478
Specific entropy on monogram characters (not overlapped) (with spaces): 3.6767355929715465625387440463725919255326050276698
Source redundancy on monogram characters (not overlapped) (with spaces): 0.26465288140569068749225119072548161489347899446604
```



→ Табличний вигляд.

Частоти букв (монограм) без пробілів:

Monogram	Frequency
а	0.10174050632911392405063291139240506329113924050633
о	0.095393811533052039381153305203938115330520393811533
и	0.094497187060478199718706047819971870604781997187060
н	0.082419127988748241912798874824191279887482419127989
е	0.073646272855133614627285513361462728551336146272855
т	0.064205344585091420534458509142053445850914205344585
р	0.052303094233473980309423347398030942334739803094233
в	0.051687763713080168776371308016877637130801687763713
с	0.050966947960618846694796061884669479606188466947961
з	0.028533755274261603375527426160337552742616033755274
к	0.027637130801687763713080168776371308016877637130802
п	0.026951476793248945147679324894514767932489451476793
я	0.025861462728551336146272855133614627285513361462729
ы	0.024964838255977496483825597749648382559774964838256
л	0.024472573839662447257383966244725738396624472573840
й	0.024015471167369901547116736990154711673699015471167
д	0.020798171589310829817158931082981715893108298171589
м	0.019567510548523206751054852320675105485232067510549
у	0.013836146272855133614627285513361462728551336146273
б	0.013168073136427566807313642756680731364275668073136
г	0.011620956399437412095639943741209563994374120956399
ч	0.011286919831223628691983122362869198312236286919831
ж	0.010601265822784810126582278481012658227848101265823
ц	0.010196905766526019690576652601969057665260196905767
х	0.0099859353023909985935302390998593530239099859353024
ш	0.0085267229254571026722925457102672292545710267229255
ь	0.0072257383966244725738396624472573839662447257383966
ф	0.0057841068917018284106891701828410689170182841068917
ю	0.0043073136427566807313642756680731364275668073136428
щ	0.0027074542897327707454289732770745428973277074542897
э	0.0010900140646976090014064697609001406469760900140647

Частоти букв (монограм) з пробілами:

Monogram	Frequency
	0.11671532393316355053108888750854090316168706130816
а	0.089865830175787315982359152742406360643518230946021
о	0.084259891918752717560096900428598049568296167463818
и	0.083467917261941735511522454810857817255730169575750
н	0.072799552767252624386607863842474687868811727436487
е	0.065050624262376545127026523386545748183116963786571
т	0.056711596993602087086154419529163302068451456612212
р	0.046198521647307286166842661034846884899683210137276
в	0.045655009627927200447232747375613392136157525312131
с	0.045018324119510528604261134231939872041741723088391
з	0.025203428784396546369339710541027393005776756320268
к	0.024411454127585564320765264923287160693210758432201
п	0.023805826448847754518914218274426983042424995341326
я	0.022843033728803031244176656935213367289893782222498
ы	0.022051059071992049195602211317473134977327784334431
л	0.021616249456487980619914280390086340766507236474315
й	0.021212497670662774085346915957512888999316727747065
д	0.018370706255046897322815081682092055407168147089881
м	0.017283682216286725883595254363625069880116777439593
у	0.012221255978632213180942915709050251568420398782533
б	0.011631157214733834399652152307596745139449655258091
г	0.010264612708863904590347226535809677619727933412013
ч	0.0099695633269147151997018448350829244052425616497919
ж	0.0093639356481769053978507981862227467544567985589167
ц	0.0090067706068699919249642834958693086527113485309647
х	0.0088204236287968196782408845269892539909311137337723
ш	0.0075315236971240449717373749922355425802844897198584
ь	0.0063823839990061494502764146841418721659730418038387
ф	0.0051090129821728057643331883967948319771414373563575
ю	0.0038045841356606000372693956146344493446797937760109
щ	0.0023914528852723771662836201006273681595130132306354
э	0.00096279272004472327473756133921361575253121311882726



Частоти біграм, що перетинаються, без пробілів:

Bigram Letters (scale x100)	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я	
а	0.107	0.135	0.719	0.000	0.833	0.403	0.542	0.879	0.100	0.000	1.023	0.492	0.411	1.410	0.125	0.452	0.116	0.338	0.185	0.037	0.033	0.000	0.774	0.422	0.241	0.000	0.000	0.000	0.005	0.000	0.392	
б	0.000	0.000	0.000	0.000	0.000	0.137	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.255	0.000	0.556	0.098	0.000	0.271	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
в	0.659	0.000	0.033	0.000	0.012	0.619	0.007	0.021	0.663	0.000	0.146	0.492	0.007	0.846	0.301	0.062	0.044	0.195	0.025	0.267	0.014	0.000	0.000	0.007	0.114	0.000	0.626	0.000	0.007	0.000	0.000	
г	0.396	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.512	0.000	0.255	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
д	0.640	0.000	0.000	0.135	0.000	0.499	0.000	0.000	0.000	0.000	0.278	0.000	0.137	0.141	0.000	0.000	0.135	0.114	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
е	0.318	0.135	0.229	0.258	0.258	0.000	0.016	0.060	0.093	0.610	0.028	0.250	0.158	1.976	0.109	0.204	0.670	0.777	0.919	0.019	0.032	0.000	0.000	0.127	0.000	0.000	0.000	0.000	0.014	0.000	0.104	
ж	0.000	0.000	0.000	0.000	0.000	0.694	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.251	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
з	1.556	0.000	0.501	0.000	0.000	0.000	0.000	0.000	0.125	0.000	0.000	0.000	0.000	0.137	0.276	0.000	0.135	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.123	0.000	0.000	0.000	
и	0.181	0.002	0.953	0.134	0.271	0.149	0.026	1.074	0.941	0.930	0.316	0.002	0.436	0.508	0.404	0.397	0.343	0.352	0.469	0.049	0.062	0.113	0.127	0.251	0.000	0.000	0.002	0.000	0.018	0.000	0.941	
й	0.169	0.004	0.134	0.004	0.035	0.002	0.028	0.116	0.144	0.000	0.053	0.005	0.044	0.403	0.188	0.301	0.218	0.179	0.090	0.063	0.055	0.000	0.004	0.026	0.113	0.000	0.004	0.000	0.023	0.000	0.000	
к	0.782	0.000	0.000	0.000	0.000	0.135	0.294	0.000	0.536	0.000	0.000	0.000	0.000	0.000	0.216	0.000	0.000	0.113	0.687	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
л	0.146	0.000	0.000	0.000	0.000	0.378	0.000	0.000	0.605	0.000	0.000	0.000	0.000	0.120	0.359	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.482	0.000	0.000	0.359	
м	0.195	0.000	0.063	0.000	0.018	0.353	0.002	0.039	0.455	0.000	0.144	0.000	0.018	0.047	0.197	0.095	0.056	0.074	0.030	0.009	0.014	0.000	0.000	0.004	0.000	0.000	0.142	0.000	0.002	0.000	0.000	
н	1.234	0.000	0.000	0.000	0.000	0.253	0.000	0.000	2.271	0.000	0.000	0.000	0.000	1.120	1.181	0.000	0.000	0.114	0.218	0.137	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.593	0.000	0.000	0.120	
о	0.055	1.030	1.364	0.236	0.290	0.000	0.121	0.417	0.062	0.251	0.142	0.709	0.285	0.174	0.519	0.139	0.976	1.468	0.719	0.026	0.025	0.000	0.114	0.007	0.127	0.000	0.000	0.000	0.000	0.005	0.000	0.280
п	0.000	0.000	0.000	0.000	0.000	0.113	0.000	0.000	0.000	0.000	0.000	0.139	0.000	0.000	1.712	0.000	0.731	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
р	2.352	0.000	0.000	0.396	0.000	0.707	0.000	0.000	0.113	0.000	0.123	0.000	0.250	0.000	0.876	0.000	0.000	0.000	0.000	0.137	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.278	
с	0.018	0.000	0.009	0.000	0.000	0.141	0.004	0.009	0.130	0.000	0.005	0.348	0.000	0.107	0.874	0.218	0.007	0.125	2.396	0.113	0.220	0.000	0.000	0.002	0.258	0.000	0.000	0.000	0.000	0.000	0.114	
т	0.742	0.000	0.849	0.000	0.018	0.209	0.005	0.083	1.577	0.000	0.171	0.000	0.035	0.336	0.810	0.223	0.556	0.269	0.055	0.179	0.046	0.000	0.000	0.012	0.000	0.000	0.000	0.241	0.005	0.000	0.000	
у	0.000	0.000	0.000	0.000	0.000	0.098	0.000	0.000	0.000	0.000	0.137	0.000	0.000	0.000	0.000	0.000	0.137	0.230	0.000	0.000	0.000	0.000	0.000	0.237	0.000	0.113	0.000	0.000	0.000	0.431	0.000	
ф	0.000	0.000	0.000	0.000	0.000	0.232	0.000	0.000	0.107	0.000	0.000	0.000	0.000	0.000	0.239	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
х	0.070	0.000	0.076	0.000	0.011	0.000	0.002	0.056	0.081	0.000	0.030	0.000	0.014	0.060	0.120	0.185	0.102	0.095	0.040	0.021	0.018	0.000	0.000	0.009	0.000	0.000	0.000	0.000	0.011	0.000	0.000	
ц	0.000	0.000	0.000	0.000	0.000	0.121	0.000	0.000	0.898	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
ч	0.267	0.004	0.000	0.000	0.004	0.638	0.000	0.002	0.011	0.000	0.005	0.004	0.000	0.004	0.032	0.004	0.012	0.000	0.132	0.002	0.005	0.000	0.002	0.002	0.000	0.000	0.002	0.000	0.000	0.000	0.000	
ш	0.114	0.000	0.000	0.000	0.000	0.626	0.000	0.000	0.113	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
щ	0.000	0.000	0.000	0.000	0.000	0.113	0.000	0.000	0.158	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
ы	0.000	0.000	0.123	0.000	0.000	0.371	0.000	0.000	0.000	0.610	0.000	0.000	0.128	0.000	0.000	0.120	0.000	0.258	0.000	0.000	0.000	0.886	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
ь	0.032	0.000	0.044	0.000	0.011	0.000	0.002	0.025	0.053	0.000	0.016	0.000	0.011	0.248	0.044	0.079	0.039	0.049	0.033	0.018	0.014	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.004	0.000	0.000	
э	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.109	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
ю	0.016	0.000	0.004	0.000	0.000	0.000	0.000	0.009	0.005	0.000	0.004	0.000	0.002	0.009	0.018	0.023	0.012	0.021	0.141	0.004	0.002	0.000	0.000	0.000	0.000	0.158	0.000	0.000	0.005	0.000	0.000	
я	0.125	0.007	0.069	0.000	0.320	0.373	0.012	0.065	0.095	0.000	0.033	0.007	0.021	0.346	0.172	0.195	0.130	0.227	0.281	0.032	0.040	0.000	0.000	0.019	0.000	0.000	0.005	0.000	0.011	0.000	0.000	

Частоти біграм, що перетинаються, з пробілами:

[illegible]

Частоти біграм, що не перетинаються, без пробілів:

Bigram Letters (scale x100)	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ы	ь	э	ю	я
а	0.098	0.148	0.749	0.000	0.854	0.401	0.556	0.932	0.088	0.000	1.030	0.492	0.411	1.421	0.116	0.415	0.098	0.366	0.162	0.039	0.032	0.000	0.809	0.408	0.274	0.000	0.000	0.000	0.004	0.000	0.383
б	0.000	0.000	0.000	0.000	0.000	0.116	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.236	0.000	0.577	0.095	0.000	0.232	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
в	0.650	0.000	0.025	0.000	0.014	0.640	0.007	0.028	0.682	0.000	0.148	0.454	0.004	0.904	0.320	0.077	0.046	0.232	0.021	0.288	0.021	0.000	0.000	0.011	0.116	0.000	0.650	0.000	0.011	0.000	0.000
г	0.373	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.468	0.000	0.257	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
д	0.654	0.000	0.000	0.137	0.000	0.510	0.000	0.000	0.000	0.000	0.264	0.000	0.130	0.120	0.000	0.000	0.116	0.098	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
е	0.327	0.162	0.204	0.281	0.295	0.000	0.011	0.046	0.098	0.658	0.042	0.243	0.158	1.916	0.127	0.190	0.643	0.728	0.974	0.021	0.039	0.000	0.000	0.109	0.000	0.000	0.000	0.000	0.011	0.000	0.091
ж	0.000	0.000	0.000	0.000	0.000	0.707	0.000	0.000	0.116	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
з	1.533	0.000	0.461	0.000	0.000	0.000	0.000	0.000	0.144	0.000	0.000	0.000	0.000	0.130	0.267	0.000	0.148	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
и	0.151	0.000	0.939	0.151	0.285	0.148	0.032	1.118	0.974	0.921	0.327	0.004	0.415	0.478	0.376	0.415	0.331	0.338	0.457	0.053	0.046	0.116	0.105	0.239	0.000	0.000	0.004	0.000	0.018	0.000	0.942
й	0.183	0.000	0.130	0.004	0.018	0.004	0.032	0.123	0.134	0.000	0.067	0.007	0.035	0.404	0.190	0.295	0.207	0.162	0.081	0.063	0.035	0.000	0.007	0.025	0.116	0.000	0.004	0.000	0.032	0.000	0.000
к	0.752	0.000	0.000	0.000	0.000	0.148	0.264	0.000	0.563	0.000	0.000	0.000	0.000	0.000	0.222	0.000	0.000	0.113	0.665	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
л	0.151	0.000	0.000	0.000	0.000	0.401	0.000	0.000	0.612	0.000	0.000	0.000	0.000	0.130	0.362	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.492	0.000	0.390
м	0.211	0.000	0.060	0.000	0.028	0.394	0.004	0.035	0.429	0.000	0.130	0.000	0.021	0.035	0.218	0.098	0.046	0.091	0.025	0.014	0.018	0.000	0.000	0.004	0.000	0.000	0.137	0.000	0.004	0.000	0.000
н	1.227	0.000	0.000	0.000	0.000	0.278	0.000	0.000	2.250	0.000	0.000	0.000	0.000	1.136	1.220	0.000	0.000	0.120	0.211	0.137	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.109
о	0.053	1.055	1.301	0.236	0.271	0.000	0.120	0.411	0.060	0.250	0.165	0.672	0.292	0.197	0.559	0.165	0.974	1.491	0.731	0.021	0.028	0.000	0.120	0.004	0.113	0.000	0.000	0.000	0.011	0.000	0.271
п	0.000	0.000	0.000	0.000	0.000	0.113	0.000	0.000	0.000	0.000	0.000	0.130	0.000	0.000	1.698	0.000	0.777	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
р	2.335	0.000	0.000	0.418	0.000	0.650	0.000	0.000	0.113	0.000	0.123	0.000	0.250	0.000	0.914	0.000	0.000	0.000	0.000	0.116	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.264
с	0.021	0.000	0.014	0.000	0.000	0.120	0.000	0.007	0.151	0.000	0.007	0.345	0.000	0.113	0.798	0.222	0.004	0.102	2.461	0.148	0.222	0.000	0.000	0.000	0.281	0.000	0.000	0.000	0.000	0.000	0.116
т	0.731	0.000	0.791	0.000	0.021	0.204	0.011	0.077	1.610	0.000	0.155	0.000	0.028	0.341	0.798	0.197	0.605	0.271	0.042	0.183	0.067	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.000	0.000	0.000
у	0.000	0.000	0.000	0.000	0.000	0.095	0.000	0.000	0.000	0.000	0.158	0.000	0.000	0.000	0.000	0.000	0.116	0.239	0.000	0.000	0.000	0.000	0.000	0.260	0.000	0.081	0.000	0.000	0.000	0.436	0.000
ф	0.000	0.000	0.000	0.000	0.000	0.236	0.000	0.000	0.113	0.000	0.000	0.000	0.000	0.000	0.239	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
х	0.060	0.000	0.060	0.000	0.018	0.000	0.004	0.049	0.088	0.000	0.032	0.000	0.014	0.042	0.102	0.183	0.102	0.098	0.042	0.021	0.011	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.011	0.000	0.000
ц	0.000	0.000	0.000	0.000	0.000	0.113	0.000	0.000	0.883	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ч	0.253	0.007	0.000	0.000	0.004	0.661	0.000	0.004	0.014	0.000	0.000	0.004	0.000	0.004	0.035	0.004	0.007	0.000	0.137	0.004	0.007	0.000	0.004	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
ш	0.113	0.000	0.000	0.000	0.000	0.584	0.000	0.000	0.109	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
щ	0.000	0.000	0.000	0.000	0.000	0.144	0.000	0.000	0.137	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ы	0.000	0.000	0.144	0.000	0.000	0.352	0.000	0.000	0.000	0.619	0.000	0.000	0.137	0.000	0.000	0.109	0.000	0.236	0.000	0.000	0.000	0.942	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ь	0.039	0.000	0.035	0.000	0.000	0.000	0.000	0.021	0.056	0.000	0.004	0.000	0.004	0.260	0.039	0.074	0.042	0.070	0.042	0.021	0.014	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.004	0.000	0.000
э	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.109	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ю	0.025	0.000	0.007	0.000	0.000	0.000	0.000	0.014	0.004	0.000	0.004	0.000	0.004	0.000	0.014	0.032	0.014	0.011	0.113	0.000	0.000	0.000	0.000	0.000	0.179	0.000	0.000	0.007	0.000	0.000	0.000
я	0.123	0.007	0.070	0.000	0.323	0.341	0.011	0.056	0.091	0.000	0.039	0.007	0.011	0.373	0.193	0.197	0.169	0.204	0.306	0.021	0.032	0.000	0.000	0.021	0.000	0.000	0.011	0.000	0.000	0.000	0.000



Частоти біграм, що не перетинаються, з пробілами:

[illegible]

Отримані значення ентропії та надлишковостей:

◆ Без пробілів.

Досліджуване	Ентропія (H)	Надлишковість (R)
Монограми	4.397335637048366423505417553 4915819544077925541229	0.1124018182668710287365889 0180468212087626439660648
Біграми, що перетинаються	3.771168464430267060927171414 1379160307693981093592	0.2387930901075304373831577 1925849461068576901023076
Біграми, що не перетинаються	3.767171203595978199236590425 7311421851254213711386	0.2395999335557620465584830 4734059667692926792433809

❖ 3 пробілами.

Досліджуване	Ентропія (H)	Надлишковість (R)
Монограми	4.4039440603974067401040621719712890044555389197678	0.11921118792051865197918756560574219910889221604644
Біграми, що перетинаються	3.6769800830674192214769506922461615586254905678761	0.26460398338651615570460986155076768827490188642478
Біграми, що не перетинаються	3.6767355929715465625387440463725919255326050276698	0.26465288140569068749225119072548161489347899446604

Далі експериментуємо із програмою CoolPinkProgram.exe:

 $H^{(10)}:$ [illegible]



Произвольная часть текста:  
\_в\_такой\_момент\_ког

Использованные буквы:

Порядок n-граммы:

5	
10	
15	
20	
25	
30	
35	
40	
45	
50	

Введенный символ:

Символ по счету:

Номер эксперимента: 51

Поле ввода символов:

Неравенство для энтропии:  
 $1,04766898898872 < H < 1,75326968951511$

Двоичная таблица угаданных символов:

```

1000000000000000000000000000000000
1000000000000000000000000000000000
1000000000000000000000000000000000
0000000001000000000000000000000000
0100000000000000000000000000000000
.....

```

Вероятности:

q[1]	= 0,64
q[2]	= 0,18
q[3]	= 0,02
q[4]	= 0,04
q[5]	= 0,02
q[6]	= 0,04
q[7]	= 0
q[8]	= 0,02
q[9]	= 0
q[10]	= 0,04
q[11]	= 0
q[12]	= 0
q[13]	= 0
q[14]	= 0
q[15]	= 0
q[16]	= 0
q[17]	= 0
q[18]	= 0
q[19]	= 0
q[20]	= 0
q[21]	= 0
q[22]	= 0
q[23]	= 0
q[24]	= 0
q[25]	= 0
q[26]	= 0
q[27]	= 0
q[28]	= 0
q[29]	= 0
q[30]	= 0
q[31]	= 0
q[32]	= 0

Строка состояния:

Зведемо результати експериментів з CoolPinkProgram.exe та відповідний підрахунок надлишковості:

<i>Досліджуване</i>	<i>Ентропія</i>	<i>Надлишковість</i>
$H^{(10)}$	$1.657 < H < 2.366$	$0.527 < R < 0.669$
$H^{(20)}$	$1.048 < H < 1.753$	$0.649 < R < 0.790$
$H^{(30)}$	$1.680 < H < 2.441$	$0.512 < R < 0.664$

## Висновки

У ході комп'ютерного практикуму ми засвоїли такі теоретичні поняття як ентропія та надлишковість джерела відкритого тексту (мови) та набули відповідного практичного досвіду визначення (обрахунку) даних величин (та проміжних як, наприклад, частота символів) на різних моделях джерела відкритого тексту (а також їх порівняння). Були також наближено обраховані значення ентропії  $H^{(10)}$ ,  $H^{(20)}$ ,  $H^{(30)}$  за допомогою програми CoolPinkProgram.exe шляхом проведення відповідних експериментів та, використовуючи їх, – надлишковості.