



Міністерство освіти і науки України
НТУУ «Київський політехнічний інститут»
Фізико-технічний інститут

КРИПТОГРАФІЯ

КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1

Експериментальна оцінка ентропії на символ джерела
відкритого тексту.

Перевірив:
Чорний О. М.

Виконали:
студенти III курсу ФТІ
групи ФБ-71
Бабенко І.М.
Гончаренко Д.А.

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.

```
import AppKit
// ----- SOURCE ----- //
let alphabet_enum : [ Int : String] = [0:" ", 1:"а", 2:"б", 3:"в", 4:"г",
5:"д", 6:"е", 7:"ж", 8:"з", 9:"и", 10:"й", 11:"к", 12:"л", 13:"м", 14:"н",
15:"о", 16:"п", 17:"р", 18:"с", 19:"т", 20:"у", 21:"ф", 22:"х", 23:"ц",
24:"ч", 25:"ш", 26:"щ", 27:"ы", 28:"ь", 29:"э", 30:"ю", 31:"я"]

let alphabet = [" ", "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"]

var invalphabet_dict : [ String : Double] = [" ":0, "а":0, "б":0, "в":0, "г":
0, "д":0, "е":0, "ж":0, "з":0, "и":0, "й":0, "к":0, "л":0, "м":0, "н":0, "о":
0, "п":0, "р":0, "с":0, "т":0, "у":0, "ф":0, "х":0, "ц":0, "ч":0, "ш":0, "щ":
0, "ы":0, "ь":0, "э":0, "ю":0, "я":0]

let invalphabet_enum : [ String : Int] = [" ":0, "а":1, "б":2, "в":3, "г":4,
"д":5, "е":6, "ж":7, "з":8, "и":9, "й":10, "к":11, "л":12, "м":13, "н":14,
"о":15, "п":16, "р":17, "с":18, "т":19, "у":20, "ф":21, "х":22, "ц":23, "ч":
24, "ш":25, "щ":26, "ы":27, "ь":28, "э":29, "ю":30, "я":31]

// ----- TEXT EDIT ----- //
let methodStart = Date()
let path = "/Users/_ria_go_/Desktop/универ/crypto1/crypto1/crypto1.txt"
var text = try String(contentsOfFile: path, encoding: String.Encoding.utf8)
text = text.lowercased()
text = text.replacingOccurrences(of: "ё", with: "е")
text = text.replacingOccurrences(of: "Ъ", with: "ь")
text = text.replacingOccurrences(of: "\n", with: " ")
text = text.filter("абвгдежзийклмнопрстуфхцчшщъыэюя ".contains)
text = text.replacingOccurrences(of: "\\s+", with: " ",
options: .regularExpression, range: nil)

// ----- MONOGRAMS & BIGRAMS COUNT ----- //
var matrix = [[Double]](repeating: [Double](repeating: 0, count:
alphabet.count), count: alphabet.count)
var bigram = ""
for character in text { invalphabet_dict[String(character)] =
invalphabet_dict[String(character)]! + 1
bigram = bigram + String(character)
if bigram.count == 2 {
let i = invalphabet_enum[String(bigram.first!)]
let j = invalphabet_enum[String(bigram.last!)]
matrix[i!][j!]+=1
bigram.remove(at: bigram.startIndex) }
}
```

```

var letters_sum = (invalphabet_dict.values).reduce(0, +)
var letter : String = ""
for i in 0...alphabet.count-1 {
    letter = alphabet[i]
    invalphabet_dict[letter] = invalphabet_dict[letter]!/letters_sum
}
let freqValDec = invalphabet_dict.sorted(by: { $0.value > $1.value })
for item in freqValDec {
    print("\ (item.key):\ (item.value) ")
}
print("\n")

var bigram_sum : Int = 0
for var i : Int in 0...31{
    for var j : Int in 0...31{
        bigram_sum = bigram_sum + Int(matrix[i][j])
        j += 1
    }
    i+=1
}
for var i : Int in 0...31{
    for var j : Int in 0...31{
        matrix[i][j] = matrix[i][j]/Double(bigram_sum)
        j += 1
    }
    i+=1
}
print("      ", terminator: " ")
for var k : Int in 1...31{
    print("\ (alphabet[k])", terminator: "      ")
    k+=1
}
print("\n")
for var i : Int in 1...31{
    print("\ (alphabet[i])", terminator: " ")
    for var j : Int in 1...31{
        print(String(format: "%.6f", matrix[i][j]), terminator: " ")
        j+=1
    }
    i+=1
    print("\n")
}

```

```

// ----- ENTROPY ----- //
func log2(x: Double) -> Double{return log(x)/log(2.0)}
var h_letter = [Double]()
for character in alphabet {
    if invalphabet_dict[character] != 0{
        h_letter.append(Double(invalphabet_dict[character]!)*log2(Double(1/
            invalphabet_dict[character]!)))
    }
}
var h1 = (h_letter.reduce(0, +))
print("H1 = \ (h1) ")
var h_bigram = [Double]()
for var i : Int in 0...31{
    for var j : Int in 0...31{
        if matrix[i][j] != 0{
            h_bigram.append(Double((matrix[i][j]))*log2(Double(1/(matrix[i][j]))))
        }
        j += 1
    }
    i+=1
}
var h2 = h_bigram.reduce(0, +)/2
print("H2 = \ (h2) ")
// ----- //
let methodFinish = Date()
let executionTime = methodFinish.timeIntervalSince(methodStart)
print("Execution time: \ (executionTime) ")

```

Результати:

Монограми:

З пробілом

о:0.15744762916879637
а:0.0933214293902421
е:0.07242164773409253
и:0.06865626150553412
н:0.05818546684990108
т:0.05420564780961603
л:0.05085698949818676
с:0.04358785204405718
р:0.04282317654563575
в:0.040193340175403346
к:0.03970108875225727
м:0.030495312822407823
у:0.02526126412862732
й:0.024944335130163413
п:0.02403265852181618
д:0.023575471583606546
я:0.016104039024074465
г:0.015905789820780024
ь:0.0151775274441331048
з:0.015010296820864986
ы:0.014670441043788798
ч:0.013274604816511597
б:0.012995437571056156
й:0.009900322458398012
ж:0.0075267265550087595
ш:0.006984575672530078
х:0.00449364860800738
щ:0.0031719872527110914
ц:0.0028671959605713355
э:0.0026042123235480944
ф:0.0021443281172664266

H₁ = 4.379945305657554

Без пробіла

о:0.1107603902392176
а:0.08595506966842471
е:0.08148604630689321
и:0.06905857589896677
н:0.06433504869986954
т:0.06036062713587143
л:0.05173310710769994
с:0.05082553681902216
р:0.04770426333944249
в:0.047120024970187836
к:0.03619396713859255
м:0.02998183258769578
у:0.02960567911707977
п:0.028523637644158818
д:0.02798101625463189
я:0.019113398266492727
г:0.018878102265724415
ь:0.018013749609840816
з:0.017815268629600877
ы:0.017411904056855197
ч:0.015755228133078295
б:0.015423892948322917
й:0.011750394160817614
ж:0.008933244763863656
ш:0.008853212110541102
х:0.00828978223115031
ю:0.00533376017415106
щ:0.0037647360122930154
ц:0.0034029884192750643
э:0.0030908610713170975
ф:0.002545038375657268

H₁ = 4.452870917974479

Дані з Інтернету

ранг	бука	употреблений	частотность
1	о	55414481	10.97%
2	а	42691213	8.45%
3	е	40487008	8.01%
4	и	37153142	7.35%
5	н	33838881	6.70%
6	т	31620970	6.26%
7	с	27627040	5.47%
8	р	23916825	4.73%
9	в	22930719	4.54%
10	л	22230174	4.40%
11	к	17653469	3.49%
12	м	16203060	3.21%
13	д	15052118	2.98%
14	п	14201572	2.81%
15	у	13245712	2.62%
16	я	10139085	2.01%
17	ы	9595941	1.90%
18	ь	8784613	1.74%
19	г	8564640	1.70%
20	з	8329904	1.65%
21	б	8051767	1.59%
22	ч	7300193	1.44%
23	й	6106262	1.21%
24	х	4904176	0.97%
25	ж	4746916	0.94%
26	ш	3678738	0.73%
27	ю	3220715	0.64%
28	щ	2438807	0.48%
29	ц	1822476	0.36%
30	э	1610107	0.32%
31	ф	1335747	0.26%
32	ь	185452	0.04%
33	ё	184928	0.04%

Біграми:

Перетинаються, з пробілом – H₂ = 3.991013097121348

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 ь э ю я
а 0.00013 0.00054 0.00249 0.00087 0.002193 0.001130 0.001066 0.003815 0.000289 0.000423 0.004212 0.000459 0.003600 0.005431 0.000832 0.000070 0.004262 0.000364 0.000006 0.000214 0.001121 0.000116 0.000939 0.000067 0.000382 0.000000 0.000000 0.000013 0.000736 0.002080
б 0.000928 0.000012 0.000020 0.000018 0.000007 0.002274 0.000011 0.000001 0.000646 0.000000 0.000154 0.000570 0.000047 0.000317 0.001531 0.000000 0.001064 0.000001 0.000036 0.000001 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000016 0.000316
в 0.000549 0.000011 0.000011 0.000262 0.000551 0.000000 0.000407 0.003471 0.000000 0.000263 0.000019 0.000121 0.000991 0.000266 0.000212 0.000635 0.002804 0.000290 0.000639 0.000000 0.000036 0.000038 0.000184 0.000924 0.000028 0.002611 0.000467 0.000000 0.000000 0.000270
г 0.002081 0.000001 0.000003 0.000000 0.000081 0.000477 0.000000 0.000001 0.000645 0.000000 0.000076 0.001497 0.000013 0.000378 0.000725 0.000001 0.001268 0.000000 0.000020 0.000477 0.000000 0.000000 0.000004 0.000001 0.000000 0.000000 0.000000 0.000000 0.000000
д 0.004289 0.000067 0.000098 0.000013 0.000016 0.000354 0.000058 0.000009 0.002005 0.000000 0.000293 0.000454 0.000125 0.001923 0.003223 0.000100 0.000976 0.000367 0.000192 0.001439 0.000000 0.000034 0.000204 0.000062 0.000154 0.000000 0.000481 0.000565 0.000000 0.000009 0.000409
е 0.000128 0.000039 0.000247 0.000286 0.000280 0.001274 0.000757 0.001386 0.000192 0.001834 0.001455 0.000534 0.000379 0.000422 0.000280 0.001158 0.000538 0.000496 0.0005273 0.000156 0.000117 0.000411 0.000340 0.000055 0.000645 0.000572 0.000000 0.000000 0.000000 0.000169
ж 0.001276 0.000024 0.000000 0.000013 0.000715 0.0002964 0.000009 0.000000 0.001113 0.000000 0.000139 0.000000 0.000000 0.000662 0.000051 0.000000 0.000004 0.000020 0.000003 0.000181 0.000000 0.000000 0.000034 0.000000 0.000000 0.000018 0.000000 0.000000 0.000000
з 0.000459 0.000165 0.000180 0.000266 0.000040 0.000015 0.000005 0.000006 0.000492 0.000000 0.000111 0.000355 0.000255 0.001407 0.000720 0.000038 0.000256 0.000013 0.000004 0.000452 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000034 0.000146 0.000000 0.000001 0.000193
и 0.000244 0.000569 0.0002715 0.000494 0.0001475 0.001829 0.000244 0.002165 0.000552 0.001621 0.002641 0.000579 0.002596 0.000200 0.000409 0.000321 0.001096 0.000301 0.000415 0.000117 0.000345 0.001457 0.000959 0.000169 0.000487 0.000206 0.000000 0.000000 0.000003 0.000272 0.001150
н 0.000000 0.000013 0.000000 0.000011 0.000112 0.000001 0.000000 0.000000 0.000000 0.000000 0.000128 0.000030 0.000036 0.000401 0.000011 0.000001 0.000000 0.000429 0.000396 0.000000 0.000003 0.000000 0.000002 0.000162 0.000167 0.000018 0.000000 0.000000 0.000000 0.000013
т 0.000691 0.000001 0.000000 0.000004 0.000000 0.000054 0.000057 0.000015 0.002187 0.000000 0.000024 0.000009 0.000003 0.000672 0.000262 0.000001 0.002103 0.000160 0.000712 0.002117 0.000000 0.000003 0.000032 0.000001 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
л 0.007242 0.000076 0.000023 0.000109 0.000006 0.004415 0.000210 0.000020 0.000215 0.000000 0.000307 0.000413 0.000051 0.000400 0.000126 0.000055 0.000000 0.000195 0.000189 0.001303 0.000020 0.000007 0.000000 0.000131 0.000015 0.000001 0.000029 0.003605 0.000001 0.000099 0.001300
р 0.003550 0.000015 0.000066 0.000074 0.000007 0.003200 0.000000 0.000000 0.002755 0.000000 0.000109 0.000181 0.000005 0.001243 0.002900 0.000182 0.000090 0.000264 0.000073 0.001703 0.000004 0.000000 0.000000 0.000026 0.000011 0.000003 0.000730 0.000045 0.000003 0.000000 0.000517
в 0.009115 0.000005 0.000028 0.000097 0.000036 0.000095 0.000011 0.000055 0.000255 0.000000 0.000000 0.000318 0.000772 0.000000 0.000000 0.000523 0.000024 0.000304 0.000100 0.000000 0.000000 0.000322 0.000160 0.000004 0.000166 0.000353 0.000001 0.000000 0.000000 0.000154 0.001405
м 0.000011 0.003121 0.007239 0.000490 0.004511 0.001558 0.001507 0.001521 0.000655 0.003854 0.002521 0.000294 0.000641 0.005287 0.000297 0.001372 0.000606 0.000587 0.000452 0.000069 0.000297 0.000404 0.000128 0.001680 0.000747 0.000225 0.000000 0.000000 0.000139 0.000308 0.000490
о 0.001655 0.000000 0.000000 0.000000 0.000000 0.002169 0.000000
п 0.007602 0.000065 0.000514 0.000000 0.000249 0.000543 0.000240 0.000045 0.000559 0.000000 0.000306 0.000245 0.000233 0.000943 0.000181 0.000061 0.000059 0.000165 0.000939 0.0002476 0.000004 0.000202 0.000063 0.000177 0.000430 0.000042 0.001341 0.000007 0.000000 0.000240 0.001003
с 0.001668 0.000067 0.0001459 0.000026 0.000245 0.000279 0.000036 0.000018 0.001773 0.000000 0.000369 0.0002475 0.000019 0.000051 0.000277 0.001713 0.000103 0.000913 0.000970 0.000740 0.000020 0.000128 0.000077 0.000314 0.000154 0.000003 0.000318 0.000368 0.000001 0.000096 0.000373
т 0.005582 0.000020 0.001985 0.000003 0.000147 0.000749 0.000000 0.000004 0.000346 0.000000 0.000719 0.000206 0.000028 0.001274 0.013105 0.000100 0.002902 0.000033 0.000051 0.001534 0.000065 0.000012 0.000018 0.000344 0.000005 0.000005 0.001399 0.000458 0.000000 0.000069 0.000430
у 0.000107 0.000514 0.000096 0.000039 0.001776 0.000200 0.001164 0.000281 0.000022 0.000140 0.000948 0.002104 0.000921 0.000406 0.000004 0.000016 0.000063 0.001234 0.001510 0.000004 0.000187 0.000790 0.000000 0.000006 0.000696 0.000669 0.000255 0.000000 0.000000 0.000015 0.001003 0.000092
ф 0.000290 0.000001 0.000000 0.000000 0.000000 0.000406 0.000000 0.000000 0.000370 0.000000 0.000000 0.000006 0.000000 0.000001 0.000255 0.000000 0.000445 0.000023 0.000016 0.000051 0.000013 0.000000 0.000000 0.000000 0.000000 0.000000 0.000030 0.000013 0.000000 0.000000 0.000000
х 0.000552 0.000001 0.000000 0.000000 0.000000 0.000001 0.000000 0.000000 0.000306 0.000000 0.000007 0.000103 0.000055 0.000320 0.001704 0.000000 0.000233 0.000055 0.000020 0.000000 0.000003 0.000000 0.000000 0.000023 0.000000 0.000000 0.000007 0.000003 0.000000 0.000000 0.000000
ц 0.000097 0.000001 0.000000 0.000000 0.000000 0.000004 0.000000 0.000000 0.000359 0.000000
ч 0.002262 0.000000 0.000001 0.000000 0.000001 0.003755 0.000000 0.000000 0.0001571 0.000000 0.000475 0.000026 0.000004 0.000774 0.000062 0.000000 0.000065 0.000000 0.002990 0.000461 0.000000 0.000000 0.000000 0.000000 0.000004 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
щ 0.000957 0.000000 0.000002 0.000000 0.000000 0.002031 0.000000 0.000000 0.000000 0.002028 0.000000 0.000382 0.000425 0.000007 0.000370 0.000233 0.000054 0.000003 0.000000 0.000096 0.000374 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
ь 0.000436 0.000000 0.000000 0.000000 0.000000 0.000158 0.000000 0.000000 0.0001103 0.000000 0.000000 0.000000 0.000000 0.000055 0.000004 0.000000 0.000003 0.000000 0.000000 0.000142 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
э 0.000000 0.000249 0.000025 0.000102 0.000000 0.000006 0.000070 0.000073 0.000003 0.001763 0.000205 0.001850 0.001247 0.000151 0.000003 0.000131 0.000243 0.000053 0.000054 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000015
ю 0.000000 0.000045 0.000003 0.000001 0.000104 0.000055 0.000001 0.000213 0.000061 0.000000 0.000006 0.000025 0.000120 0.000007 0.000011 0.000001 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
я 0.000001 0.000170 0.000016 0.000000
а 0.000000 0.000022 0.000049 0.000093 0.000058 0.000017 0.000017 0.000025 0.000026 0.000046 0.0000193 0.000709 0.000303 0.000533 0.000000 0.000057 0.000096 0.000731 0.001219 0.000011 0.000007 0.000147 0.000040 0.000150 0.000023 0.000036 0.000000 0.000000 0.000000 0.000142 0.000127

Не перетинаються, з пробілом – $H_2 = 3.990569680311937$

[illegible]

Перетинаються, без пробіла – $H_2 = 4,148671405718143$

	а	б	в	г	д	е	ж	з	и	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я	
а	0.000466	0.01418	0.005852	0.001617	0.003239	0.001766	0.001489	0.005133	0.001825	0.000741	0.006239	0.010446	0.004589	0.000491	0.001172	0.003366	0.005535	0.000673	0.006862	0.006666	0.000336	0.001441	0.001714	0.001591	0.001192	0.000480	0.000000	0.000000	0.000346	0.000888	0.002666	
б	0.001014	0.000000	0.000042	0.000024	0.000014	0.002732	0.000014	0.000000	0.000794	0.000000	0.000150	0.000677	0.000064	0.000336	0.002726	0.000000	0.001205	0.000000	0.000013	0.001461	0.000000	0.000043	0.000002	0.000000	0.000011	0.002282	0.000028	0.000280	0.000050	0.000019	0.000000	
в	0.007864	0.002772	0.000029	0.000141	0.000599	0.000666	0.000060	0.000037	0.000455	0.000003	0.001652	0.000006	0.000466	0.001604	0.001755	0.001095	0.001000	0.002188	0.001000	0.000034	0.000038	0.000077	0.000006	0.000038	0.001440	0.000446	0.000399	0.000554	0.000384	0.000018	0.000377	
г	0.002468	0.000000	0.000059	0.000010	0.001069	0.000576	0.000003	0.000000	0.000445	0.000000	0.000014	0.001782	0.000007	0.000493	0.000970	0.000075	0.001529	0.000007	0.000048	0.000501	0.000002	0.000003	0.000002	0.000001	0.000002	0.000000	0.000000	0.000010	0.000000	0.000000	0.000002	
д	0.005647	0.002122	0.001159	0.000000	0.000062	0.004668	0.000072	0.000000	0.000042	0.000247	0.000000	0.000066	0.000053	0.000226	0.000385	0.000000	0.000296	0.001130	0.000597	0.000311	0.000003	0.000046	0.000242	0.000101	0.000154	0.000000	0.000071	0.000029	0.000014	0.000000	0.000002	
е	0.000447	0.001940	0.000433	0.003878	0.004656	0.001766	0.001117	0.002340	0.001458	0.002182	0.002595	0.001729	0.002507	0.000938	0.000139	0.003579	0.000215	0.000625	0.000731	0.001737	0.001751	0.000227	0.000669	0.000421	0.001783	0.000054	0.000693	0.000000	0.000000	0.000034	0.000216	0.000354
ж	0.001524	0.000000	0.000046	0.000021	0.000001	0.000334	0.000014	0.000010	0.001341	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
з	0.000523	0.002521	0.001461	0.000378	0.001992	0.000748	0.000112	0.000019	0.000055	0.000002	0.000230	0.000540	0.000402	0.001066	0.000902	0.000235	0.000359	0.000190	0.000609	0.000555	0.000003	0.000000	0.000006	0.000006	0.000006	0.000016	0.000002	0.000467	0.001713	0.000034	0.000002	0.000245
и	0.000547	0.001389	0.000511	0.001410	0.002451	0.002041	0.000047	0.000235	0.001770	0.001324	0.000184	0.001749	0.001732	0.000590	0.001730	0.002643	0.001742	0.000522	0.000736	0.000504	0.001814	0.001173	0.002521	0.000036	0.000264	0.000000	0.000025	0.000230	0.000072	0.001537	0.000000	0.000000
к	0.000515	0.000000	0.000008	0.000013	0.000027	0.000075	0.001050	0.000235	0.000050	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
л	0.000341	0.000342	0.001181	0.000012	0.001226	0.001808	0.001817	0.000115	0.000387	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
м	0.000896	0.004741	0.001180	0.000000	0.000031	0.000575	0.000287	0.000211	0.000117	0.000000	0.000180	0.000583	0.000359	0.001386	0.000716	0.000909	0.000269	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
н	0.004341	0.003555	0.000999	0.004510	0.000338	0.000918	0.000096	0.000253	0.000328	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
о	0.001895	0.001186	0.000016	0.000194	0.000196	0.001837	0.000150	0.000282	0.000897	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
п	0.000280	0.004852	0.011129	0.000750	0.001698	0.000233	0.000294	0.000235	0.000195	0.000575	0.000184	0.000783	0.000566	0.000814	0.001286	0.001048	0.001223	0.000703	0.000972	0.000871	0.000785	0.000434	0.000740	0.000189	0.002072	0.000185	0.000280	0.000000	0.000000	0.000000	0.000000	
р	0.001566	0.000000	0.000003	0.000002	0.000002	0.000274	0.000000	0.000000	0.000000	0.000015	0.000184	0.000002	0.000003	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
с	0.000688	0.00125	0.000780	0.001152	0.000354	0.000720	0.000291	0.000006	0.000616	0.000000	0.000555	0.000317	0.000317	0.000279	0.000282	0.000280	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
т	0.002038	0.00237	0.000285	0.000163	0.000399	0.000359	0.000415	0.000017	0.000262	0.000000	0.000493	0.000295	0.000588	0.000137	0.000339	0.002419	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
у	0.000776	0.002722	0.000330	0.001711	0.001368	0.000602	0.000163	0.000214	0.000419	0.000000	0.000218	0.000386	0.000261	0.000161	0.000381	0.000823	0.000359	0.001340	0.000747	0.002478	0.000118	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
ф	0.000314	0.000956	0.000120	0.001249	0.002377	0.000331	0.001453	0.000506	0.000000	0.000166	0.000149	0.002590	0.000147	0.001474	0.000383	0.001506	0.001297	0.002876	0.002858	0.000186	0.000255	0.000911	0.000016	0.000178	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
х	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
ц	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
ч	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
ш	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
щ	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
ъ	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
ы	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
ь	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
э	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	
ю	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000																						

Не перетинаються, без пробіла – $H_2 = 4.147853620317564$

[illegible]

2. За допомогою програми CoolPinkProgram оцінити значення $(10)H$, $(20)H$, $(30)H$.

Лабораторная работа №1

Произвольная часть текста:
к_членам_

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ:

Символ по счету:

Номер эксперимента: 50

Поле ввода символов:

Продолжить Другой

Неравенство для энтропии:
 $3,28952006524464 < H < 3,71011806279464$

Двоичная таблица угаданных символов:

00010000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

$q[1] = 0,2653061$
$q[2] = 0,1020408$
$q[3] = 0,1020408$
$q[4] = 0,0204081$
$q[5] = 0,0408163$
$q[6] = 0$
$q[7] = 0,0204081$
$q[8] = 0,0408163$
$q[9] = 0,0204081$
$q[10] = 0,0408163$
$q[11] = 0$
$q[12] = 0,0408163$
$q[13] = 0$
$q[14] = 0,0204081$
$q[15] = 0$
$q[16] = 0,0204081$
$q[17] = 0$
$q[18] = 0,0204081$
$q[19] = 0$
$q[20] = 0$
$q[21] = 0$
$q[22] = 0$
$q[23] = 0,0408163$
$q[24] = 0$
$q[25] = 0,0204081$
$q[26] = 0,0816327$
$q[27] = 0$
$q[28] = 0$
$q[29] = 0,0408163$
$q[30] = 0,0204081$
$q[31] = 0$
$q[32] = 0,0408163$

Строка состояния:

Лабораторная работа №1

Произвольная часть текста:
_проявил_к_нему_неп

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ:

Символ по счету:

Номер эксперимента: 52

Поле ввода символов:

Продолжить Другой

Неравенство для энтропии:
 $2,34631647890554 < H < 2,98906667047088$

Двоичная таблица угаданных символов:

00000000000000000000000000000000
00000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
00000000000000000000000000000000
00000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

$q[1] = 0,4901960784$
$q[2] = 0,0588235294$
$q[3] = 0,0784313725$
$q[4] = 0$
$q[5] = 0,0392156862$
$q[6] = 0$
$q[7] = 0$
$q[8] = 0,0196078431$
$q[9] = 0$
$q[10] = 0,0392156862$
$q[11] = 0$
$q[12] = 0$
$q[13] = 0,0196078431$
$q[14] = 0$
$q[15] = 0,0196078431$
$q[16] = 0,0392156862$
$q[17] = 0$
$q[18] = 0$
$q[19] = 0,0196078431$
$q[20] = 0$
$q[21] = 0$
$q[22] = 0,0196078431$
$q[23] = 0,0196078431$
$q[24] = 0,0196078431$
$q[25] = 0,0196078431$
$q[26] = 0,0196078431$
$q[27] = 0,0392156862$
$q[28] = 0,0196078431$
$q[29] = 0$
$q[30] = 0,0196078431$
$q[31] = 0$
$q[32] = 0$

Строка состояния:

Лабораторная работа №1

Произвольная часть текста:
ть_разница_между_справедливым

Использованные буквы:

Порядок n-граммы:
5 символов
10 символов
15 символов
20 символов
25 символов
30 символов
35 символов
40 символов
45 символов
50 символов

Введенный символ:

Символ по счету:

Номер эксперимента: 50

Поле ввода символов:

Продолжить Другой

Неравенство для энтропии:
 $3,40815856372611 < H < 3,75339162921483$

Двоичная таблица угаданных символов:

00000001000000000000000000000000
00010000000000000000000000000000
00000001000000000000000000000000
00000000000000000000000000000000
00000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:

$q[1] = 0,28813559322$
$q[2] = 0,08474576271$
$q[3] = 0,01634915254$
$q[4] = 0,05084745762$
$q[5] = 0$
$q[6] = 0,01634915254$
$q[7] = 0,0338930506$
$q[8] = 0,01634915254$
$q[9] = 0,0338930506$
$q[10] = 0$
$q[11] = 0,0508474576$
$q[12] = 0,0163491525$
$q[13] = 0,0508474576$
$q[14] = 0$
$q[15] = 0$
$q[16] = 0$
$q[17] = 0$
$q[18] = 0,0508474576$
$q[19] = 0,0338930506$
$q[20] = 0,0163491525$
$q[21] = 0$
$q[22] = 0,0163491525$
$q[23] = 0,0163491525$
$q[24] = 0,0163491525$
$q[25] = 0,0338930506$
$q[26] = 0,0338930506$
$q[27] = 0,0163491525$
$q[28] = 0,0163491525$
$q[29] = 0$
$q[30] = 0$
$q[31] = 0$
$q[32] = 0$

Строка состояния:

3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

$$R = 1 - \frac{H_{\infty}}{H_0} \quad H_0 = \log_2 32 = 5$$

З пробілом, H_1 : $R_1 = 1 - \frac{4,452870917974479}{5} = 0,109425816$

Без пробіла, H_1 : $R_1 = 1 - \frac{4,379945305657554}{5} = 0,124010939$

Перетинаються, з пробілом, H_2 : $R_2 = 1 - \frac{3,991013097121348}{5} = 0,201797381$

Не перетинаються, з пробілом, H_2 : $R_2 = 1 - \frac{3,990569680311937}{5} = 0,201886064$

Перетинаються, без пробіла, H_2 : $R_2 = 1 - \frac{4,148671405718143}{5} = 0,170265719$

Не перетинаються, без пробіла, H_2 : $R_2 = 1 - \frac{4,147853620317564}{5} = 0,170429276$

$$R_{10} = 0,3$$

$$R_{20} = 0,46$$

$$R_{30} = 0,28$$

Опис роботи та основні труднощі:

Для роботи було створено текст, який містив у собі цикл статей про шифрування та криптографію і роман М. А. Булгакова «Майстер і Маргарита», таким чином поєднавши науково-технічну та художню літературу. Текстовий файл розміром у 1,4 мб оброблюється програмою за 2-3 секунди в залежності від умов (з пробілами чи без, з перетином біграм чи без). Програма написана на мові Swift. Особливих труднощів під час роботи над комп'ютерним практикумом не виникло, за виключенням того, що декілька разів доводилось змінювати підхід до основного алгоритму програми задля збільшення її оптимізації. В найпершій версії текст проходився програмою $32+32*32$ рази, окремо для пошуку кожної монограми та кожної біграми, що займало набагато більше часу. Проблема була вирішена зміною алгоритму таким чином, що текст потрібно було пройти лише 1 раз, і створивши спеціальний словник та матрицю. З кожним кроком по тексті формується нова біграма (або через крок, якщо задано пошук біграм без перетину), а також перевіряється, чи наявна буква в словнику (якщо так, то її значення в словнику збільшується на 1). Сформовані біграми розбиваються на першу і другу, що відповідним чином вказує на і-тий рядок та j-тий стовпчик в матриці, що вказують на елемент, який також інкрементується. Далі кожне значення в словнику ділиться на суму всіх значень, а кожний елемент матриці – на суму всіх елементів. Таким чином ми отримуємо частоту кожної букви та біграми. Ентропії H_1 і H_2 рахуються за відповідною формулою.

Висновок:

Отже, в ході практикума ми засвоїли поняття 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