

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Фізико-технічний інститут

КРИПТОГРАФІЯ
КОМП'ЮТЕРНИЙ ПРАКТИКУМ №1
Експериментальна оцінка ентропії на символ джерела
відкритого тексту

Виконали:
ФБ-33 Самохвалов Роман
ФБ-33 Лозенко Павло

Київ

2025

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

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

1.

```
49     for bg, cnt in bf.items():
50         if len(bg) == 2:
51             mat.loc[bg[0], bg[1]] = cnt
52
53     return mat
54
55
56 def analyze(fname):
57     """Головна функція: читання, аналіз, експорт"""
58     # Автовизначення кодування
59     for enc in ['utf-8', 'windows-1251', 'cp1251', 'koi8-r', 'iso-8859-5']:
60         try:
61             with open(fname, 'r', encoding=enc) as f:
62                 raw = f.read()
63                 print(f"✓ Файл прочитано ({enc})")
64                 break
65         except UnicodeDecodeError:
66             continue
67     else:
68         raise ValueError("Не вдалося прочитати файл")
69
70     print(f"Вихідний текст: {len(raw)} символів")
71
72     # Обробка з пробілами і без
73     txt_sp = preprocess(raw, spaces=True)
74     txt_no = preprocess(raw, spaces=False)
75
76     print(f"З пробілами: {len(txt_sp)} символів")
77     print(f"Без пробілів: {len(txt_no)} символів")
78
79     # Частоти
80     freq_sp = char_freq(txt_sp)
81     freq_no = char_freq(txt_no)
82     bg_ov = bigram_freq(txt_sp, overlap=True)
83     bg_nv = bigram_freq(txt_sp, overlap=False)
84
85     # Розмір алфавіту
86     m_sp = len(freq_sp)
87     m_no = len(freq_no)
88
89     print(f"\nАлфавіт з пробілами: {m_sp}")
90     print(f"Алфавіт без пробілів: {m_no}")
91
92     # Ентропії
93     h0_sp = math.log2(m_sp)
94     h0_no = math.log2(m_no)
```

Очищення тексту:

Код приводить текст до нижнього регістру, замінює ё на е та ь на ъ, а потім видаляє всі символи, окрім літер українського алфавіту (та пробілів, в одному з варіантів).

Два режими аналізу:

Скрипт аналізує текст у двох версіях: з пробілами та без пробілів. Підрахунок частот: Він рахує частоту появи кожного окремого символу (монограми) та кожної пари символів (біграми).

Розрахунок ентропії:

Обчислює ентропію Шеннона для різних моделей: H_0 (Максимальна ентропія): ентропія для алфавіту, де всі символи рівноймовірні. H_1 (Ентропія 1-грам): ентропія на основі реальних частот окремих символів. H_2 (Ентропія 2-грам): ентропія на основі реальних частот біграм.

Розрахунок надлишковості:

Обчислює надлишковість (R) мови в тексті, яка показує, наскільки ентропія тексту менша за максимальну (H_0).

Приклад виводу коду:

```
=== З ПРОБІЛАМИ ===
H0: 5.000000 біт
H1: 4.380825 біт
H2: 3.966895 біт

=== БЕЗ ПРОБІЛІВ ===
H0: 4.954196 біт
H1: 4.446448 біт
H2: 4.119529 біт

=== НАДЛИШКОВІСТЬ ===
R1 (з пробілами): 0.123835 (12.38%)
R1 (без пробілів): 0.102489 (10.25%)
R2 (з пробілами): 0.206621 (20.66%)
R2 (без пробілів): 0.168477 (16.85%)
```

| № | Символ | Кількість | Частота | Відсоток |
|----|--------|-----------|----------|----------|
| 1 | о | 94326 | 0.115244 | 11.5244% |
| 2 | е | 68544 | 0.083745 | 8.3745% |
| 3 | а | 61232 | 0.074811 | 7.4811% |
| 4 | и | 56463 | 0.068985 | 6.8985% |
| 5 | н | 53904 | 0.065858 | 6.5858% |
| 6 | т | 51621 | 0.063069 | 6.3069% |
| 7 | с | 46003 | 0.056205 | 5.6205% |
| 8 | р | 41968 | 0.051275 | 5.1275% |
| 9 | л | 41280 | 0.050435 | 5.0435% |
| 10 | в | 34850 | 0.042579 | 4.2579% |
| 11 | к | 27770 | 0.033928 | 3.3928% |
| 12 | д | 24167 | 0.029526 | 2.9526% |
| 13 | п | 23820 | 0.029103 | 2.9103% |
| 14 | м | 23363 | 0.028544 | 2.8544% |
| 15 | у | 20937 | 0.02558 | 2.5580% |
| 16 | ы | 17657 | 0.021573 | 2.1573% |
| 17 | ь | 16891 | 0.020637 | 2.0637% |
| 18 | б | 16180 | 0.019768 | 1.9768% |
| 19 | я | 14914 | 0.018221 | 1.8221% |
| 20 | г | 13927 | 0.017016 | 1.7016% |
| 21 | з | 12961 | 0.015835 | 1.5835% |
| 22 | ч | 10544 | 0.012882 | 1.2882% |
| 23 | й | 9398 | 0.011482 | 1.1482% |
| 24 | ж | 8669 | 0.010592 | 1.0592% |
| 25 | х | 7608 | 0.009295 | 0.9295% |
| 26 | ш | 5863 | 0.007163 | 0.7163% |
| 27 | ю | 3804 | 0.004648 | 0.4648% |
| 28 | щ | 2810 | 0.003433 | 0.3433% |
| 29 | э | 2763 | 0.003376 | 0.3376% |
| 30 | ц | 2221 | 0.002714 | 0.2714% |
| 31 | ф | 2028 | 0.002478 | 0.2478% |

Рис 1.1

| № | Символ | Кількість | Частота | Відсоток |
|----|--------|-----------|----------|----------|
| 1 | ПРОБІЛ | 149354 | 0.154317 | 15.4317% |
| 2 | о | 94326 | 0.09746 | 9.7460% |
| 3 | е | 68544 | 0.070822 | 7.0822% |
| 4 | а | 61232 | 0.063267 | 6.3267% |
| 5 | и | 56463 | 0.058339 | 5.8339% |
| 6 | н | 53904 | 0.055695 | 5.5695% |
| 7 | т | 51621 | 0.053336 | 5.3336% |
| 8 | с | 46003 | 0.047532 | 4.7532% |
| 9 | р | 41968 | 0.043363 | 4.3363% |
| 10 | л | 41280 | 0.042652 | 4.2652% |
| 11 | в | 34850 | 0.036008 | 3.6008% |
| 12 | к | 27770 | 0.028693 | 2.8693% |
| 13 | д | 24167 | 0.02497 | 2.4970% |
| 14 | п | 23820 | 0.024612 | 2.4612% |
| 15 | м | 23363 | 0.024139 | 2.4139% |
| 16 | у | 20937 | 0.021633 | 2.1633% |
| 17 | ы | 17657 | 0.018244 | 1.8244% |
| 18 | ь | 16891 | 0.017452 | 1.7452% |
| 19 | б | 16180 | 0.016718 | 1.6718% |
| 20 | я | 14914 | 0.01541 | 1.5410% |
| 21 | г | 13927 | 0.01439 | 1.4390% |
| 22 | з | 12961 | 0.013392 | 1.3392% |
| 23 | ч | 10544 | 0.010894 | 1.0894% |
| 24 | й | 9398 | 0.00971 | 0.9710% |
| 25 | ж | 8669 | 0.008957 | 0.8957% |
| 26 | х | 7608 | 0.007861 | 0.7861% |
| 27 | ш | 5863 | 0.006058 | 0.6058% |
| 28 | ю | 3804 | 0.00393 | 0.3930% |
| 29 | щ | 2810 | 0.002903 | 0.2903% |
| 30 | э | 2763 | 0.002855 | 0.2855% |
| 31 | ц | 2221 | 0.002295 | 0.2295% |
| 32 | ф | 2028 | 0.002095 | 0.2095% |

Рис 2.2

На рисунку 2.1 зображено частоти символів без пробілів, а на рисунку 2.2 зображено з пробілами.

| | ПР | а | б | в | г | д | е | ж | з | и | й | к | л | м | н | о |
|----|-------|-------|------|-------|------|------|------|------|------|------|------|------|------|------|-------|-------|
| ПР | 0 | 1303 | 7435 | 13755 | 3328 | 6691 | 2908 | 1799 | 3822 | 9082 | 9 | 8911 | 2430 | 5688 | 15250 | 11257 |
| а | 14186 | 3 | 1550 | 2562 | 710 | 1759 | 840 | 810 | 2973 | 138 | 1915 | 3420 | 7734 | 2551 | 3568 | 22 |
| б | 616 | 647 | 3 | 12 | 2 | 13 | 2650 | 35 | 2 | 694 | 0 | 61 | 1092 | 26 | 287 | 2228 |
| в | 6209 | 5172 | 17 | 27 | 21 | 193 | 4993 | 0 | 322 | 2766 | 0 | 232 | 759 | 90 | 635 | 6256 |
| г | 714 | 1135 | 0 | 11 | 31 | 853 | 568 | 0 | 1 | 485 | 0 | 165 | 1196 | 2 | 333 | 6336 |
| д | 1005 | 3676 | 34 | 1208 | 15 | 30 | 4180 | 33 | 42 | 2191 | 0 | 388 | 756 | 46 | 2066 | 3755 |
| е | 15052 | 129 | 649 | 1711 | 3314 | 3148 | 989 | 941 | 1289 | 116 | 1596 | 1116 | 4671 | 3770 | 7952 | 687 |
| ж | 61 | 935 | 8 | 4 | 8 | 838 | 3447 | 21 | 0 | 1489 | 0 | 24 | 3 | 1 | 1279 | 27 |
| з | 1382 | 4312 | 169 | 895 | 255 | 656 | 655 | 25 | 4 | 424 | 0 | 202 | 265 | 399 | 1208 | 732 |
| и | 17227 | 82 | 636 | 1833 | 481 | 1363 | 2085 | 252 | 1998 | 587 | 862 | 1565 | 5602 | 2383 | 3520 | 51 |
| й | 6382 | 0 | 0 | 0 | 0 | 188 | 3 | 0 | 5 | 0 | 0 | 74 | 12 | 50 | 317 | 27 |
| к | 3876 | 6163 | 0 | 75 | 1 | 0 | 517 | 40 | 9 | 2623 | 0 | 52 | 560 | 2 | 306 | 8857 |
| л | 5970 | 3929 | 33 | 18 | 108 | 20 | 4867 | 439 | 14 | 8677 | 0 | 299 | 141 | 17 | 511 | 6481 |
| м | 6960 | 1567 | 21 | 0 | 37 | 0 | 3093 | 0 | 3 | 3573 | 0 | 99 | 279 | 14 | 992 | 3593 |
| н | 3172 | 10136 | 1 | 25 | 121 | 189 | 8906 | 666 | 20 | 7752 | 0 | 1745 | 0 | 0 | 2088 | 9844 |
| о | 19208 | 3 | 3679 | 6930 | 4275 | 5042 | 1600 | 1980 | 1526 | 837 | 3290 | 2406 | 6042 | 4198 | 5479 | 232 |
| п | 306 | 1591 | 0 | 0 | 0 | 0 | 3150 | 0 | 0 | 1512 | 0 | 117 | 1031 | 0 | 177 | 7909 |
| р | 1356 | 7642 | 735 | 461 | 149 | 191 | 7764 | 276 | 52 | 4087 | 0 | 198 | 46 | 121 | 1241 | 8672 |
| с | 3011 | 2333 | 35 | 1142 | 13 | 243 | 2440 | 16 | 9 | 1148 | 0 | 3356 | 2843 | 1394 | 1100 | 2585 |
| т | 6071 | 5184 | 16 | 2376 | 2 | 103 | 4206 | 2 | 0 | 4110 | 0 | 501 | 262 | 56 | 1608 | 12124 |
| у | 4960 | 54 | 750 | 463 | 812 | 1982 | 164 | 1042 | 345 | 17 | 33 | 557 | 958 | 717 | 249 | 30 |
| ф | 1092 | 210 | 0 | 0 | 0 | 0 | 68 | 0 | 0 | 54 | 0 | 0 | 63 | 0 | 9 | 153 |
| х | 3544 | 918 | 0 | 217 | 0 | 2 | 38 | 0 | 0 | 148 | 0 | 0 | 96 | 8 | 175 | 2069 |
| ц | 423 | 662 | 0 | 29 | 0 | 0 | 557 | 0 | 0 | 206 | 0 | 4 | 0 | 4 | 0 | 111 |
| ч | 61 | 2205 | 0 | 33 | 0 | 0 | 2875 | 0 | 0 | 1335 | 0 | 204 | 17 | 25 | 728 | 24 |
| ш | 70 | 735 | 0 | 5 | 0 | 0 | 1720 | 0 | 0 | 1539 | 0 | 304 | 552 | 3 | 245 | 261 |
| щ | 12 | 506 | 0 | 0 | 0 | 0 | 1321 | 0 | 0 | 756 | 0 | 0 | 0 | 0 | 83 | 1 |
| ы | 5463 | 0 | 166 | 752 | 50 | 76 | 1335 | 19 | 59 | 3 | 1647 | 121 | 2947 | 1245 | 85 | 0 |
| ь | 10348 | 0 | 78 | 13 | 8 | 71 | 486 | 1 | 235 | 35 | 0 | 970 | 0 | 144 | 1675 | 2 |
| э | 9 | 0 | 1 | 14 | 4 | 1 | 1 | 0 | 0 | 0 | 5 | 115 | 45 | 2 | 27 | 0 |
| ю | 1917 | 0 | 88 | 9 | 100 | 246 | 1 | 156 | 29 | 0 | 15 | 16 | 42 | 8 | 148 | 0 |
| я | 8691 | 0 | 76 | 270 | 82 | 269 | 117 | 116 | 202 | 79 | 26 | 548 | 836 | 399 | 563 | 0 |

Біграми з перекриттям з пробілами

| | ПР | а | б | в | г | д | е | ж | з | и | й | к | л | м | н | о |
|----|------|------|------|------|------|------|------|-----|------|------|------|------|------|------|------|------|
| ПР | 0 | 615 | 3734 | 6767 | 1681 | 3375 | 1432 | 880 | 1898 | 4542 | 5 | 4429 | 1214 | 2826 | 7662 | 5620 |
| а | 7195 | 2 | 791 | 1297 | 354 | 910 | 438 | 405 | 1517 | 70 | 975 | 1692 | 3904 | 1263 | 1851 | 12 |
| б | 317 | 321 | 1 | 7 | 1 | 7 | 1310 | 16 | 1 | 329 | 0 | 32 | 539 | 9 | 147 | 1139 |
| в | 3105 | 2563 | 7 | 12 | 10 | 100 | 2465 | 0 | 159 | 1391 | 0 | 115 | 381 | 49 | 307 | 3160 |
| г | 370 | 563 | 0 | 8 | 16 | 435 | 266 | 0 | 1 | 236 | 0 | 83 | 598 | 1 | 171 | 3145 |
| д | 496 | 1880 | 19 | 595 | 5 | 18 | 2053 | 13 | 21 | 1093 | 0 | 200 | 373 | 20 | 1000 | 1855 |
| е | 7601 | 68 | 314 | 850 | 1662 | 1543 | 499 | 462 | 633 | 54 | 811 | 565 | 2360 | 1912 | 3999 | 372 |
| ж | 30 | 465 | 5 | 3 | 4 | 442 | 1727 | 10 | 0 | 770 | 0 | 12 | 3 | 1 | 629 | 15 |
| з | 693 | 2140 | 80 | 464 | 130 | 325 | 335 | 8 | 3 | 210 | 0 | 113 | 141 | 197 | 625 | 334 |
| и | 8641 | 41 | 314 | 929 | 226 | 638 | 1042 | 135 | 1007 | 292 | 434 | 800 | 2833 | 1205 | 1735 | 21 |
| й | 3225 | 0 | 0 | 0 | 0 | 97 | 1 | 0 | 2 | 0 | 0 | 36 | 5 | 18 | 162 | 14 |
| к | 1969 | 3059 | 0 | 34 | 1 | 0 | 248 | 21 | 4 | 1324 | 0 | 23 | 281 | 1 | 156 | 4414 |
| л | 2934 | 1999 | 17 | 6 | 52 | 10 | 2418 | 211 | 3 | 4248 | 0 | 152 | 64 | 7 | 239 | 3209 |
| м | 3442 | 790 | 12 | 0 | 19 | 0 | 1599 | 0 | 2 | 1777 | 0 | 49 | 147 | 7 | 509 | 1788 |
| н | 1531 | 4986 | 1 | 14 | 58 | 89 | 4475 | 316 | 11 | 3893 | 0 | 857 | 0 | 0 | 1006 | 4949 |
| о | 9699 | 2 | 1826 | 3454 | 2160 | 2540 | 794 | 972 | 751 | 444 | 1628 | 1211 | 3098 | 2087 | 2795 | 107 |
| п | 146 | 815 | 0 | 0 | 0 | 0 | 1528 | 0 | 0 | 716 | 0 | 58 | 536 | 0 | 89 | 3887 |
| р | 727 | 3744 | 358 | 251 | 70 | 88 | 3909 | 148 | 30 | 2070 | 0 | 94 | 20 | 66 | 611 | 4285 |
| с | 1508 | 1109 | 20 | 600 | 4 | 126 | 1197 | 7 | 6 | 584 | 0 | 1683 | 1413 | 708 | 554 | 1337 |
| т | 2985 | 2571 | 3 | 1232 | 2 | 54 | 2121 | 2 | 0 | 2032 | 0 | 242 | 137 | 27 | 800 | 5924 |
| у | 2479 | 33 | 402 | 230 | 401 | 1007 | 74 | 528 | 170 | 13 | 11 | 282 | 485 | 332 | 116 | 13 |
| ф | 519 | 105 | 0 | 0 | 0 | 0 | 32 | 0 | 0 | 29 | 0 | 0 | 35 | 0 | 6 | 77 |
| х | 1792 | 441 | 0 | 119 | 0 | 1 | 16 | 0 | 0 | 76 | 0 | 0 | 44 | 6 | 88 | 1035 |
| ц | 220 | 319 | 0 | 18 | 0 | 0 | 274 | 0 | 0 | 105 | 0 | 2 | 0 | 2 | 0 | 61 |
| ч | 27 | 1075 | 0 | 18 | 0 | 0 | 1424 | 0 | 0 | 674 | 0 | 89 | 8 | 11 | 378 | 11 |
| ш | 44 | 354 | 0 | 2 | 0 | 0 | 835 | 0 | 0 | 750 | 0 | 159 | 295 | 2 | 110 | 128 |
| щ | 6 | 241 | 0 | 0 | 0 | 0 | 680 | 0 | 0 | 363 | 0 | 0 | 0 | 0 | 48 | 1 |
| ы | 2673 | 0 | 81 | 374 | 24 | 44 | 671 | 8 | 23 | 1 | 792 | 65 | 1462 | 610 | 38 | 0 |
| ь | 5202 | 0 | 41 | 4 | 4 | 37 | 236 | 1 | 120 | 19 | 0 | 500 | 0 | 68 | 839 | 2 |
| э | 7 | 0 | 1 | 4 | 4 | 1 | 1 | 0 | 0 | 0 | 3 | 65 | 17 | 1 | 16 | 0 |
| ю | 950 | 0 | 39 | 6 | 54 | 119 | 1 | 76 | 14 | 0 | 8 | 10 | 18 | 3 | 87 | 0 |
| я | 4341 | 0 | 31 | 143 | 43 | 125 | 63 | 57 | 96 | 32 | 16 | 268 | 432 | 185 | 265 | 0 |

Біграми без перекриття з пробілами

| | а | б | в | г | д | е | ж | з | и | й | к | л | м | н | о |
|---|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|-------|
| а | 134 | 2329 | 3758 | 1042 | 2338 | 1059 | 899 | 3389 | 1109 | 1915 | 4356 | 7974 | 3074 | 5132 | 1139 |
| б | 651 | 14 | 56 | 33 | 33 | 2659 | 35 | 14 | 776 | 0 | 87 | 1094 | 38 | 348 | 2310 |
| в | 5232 | 217 | 449 | 362 | 462 | 5075 | 43 | 417 | 3062 | 0 | 818 | 966 | 292 | 1138 | 6630 |
| г | 1146 | 65 | 62 | 37 | 894 | 578 | 2 | 21 | 531 | 0 | 203 | 1205 | 27 | 401 | 6427 |
| д | 3682 | 68 | 1334 | 54 | 76 | 4199 | 38 | 73 | 2264 | 1 | 453 | 769 | 72 | 2167 | 3834 |
| е | 250 | 1493 | 3065 | 3614 | 3822 | 1189 | 1162 | 1766 | 898 | 1596 | 1897 | 4953 | 4557 | 9318 | 1834 |
| ж | 935 | 11 | 11 | 9 | 840 | 3449 | 21 | 1 | 1491 | 0 | 26 | 3 | 5 | 1297 | 32 |
| з | 4320 | 199 | 1000 | 307 | 720 | 685 | 39 | 57 | 470 | 0 | 331 | 293 | 446 | 1372 | 837 |
| и | 252 | 1507 | 3636 | 826 | 2163 | 2597 | 405 | 2388 | 1548 | 862 | 2569 | 5852 | 3025 | 5288 | 1469 |
| й | 65 | 246 | 520 | 174 | 633 | 74 | 87 | 198 | 375 | 0 | 494 | 158 | 333 | 766 | 454 |
| к | 6189 | 268 | 414 | 81 | 145 | 578 | 106 | 105 | 2906 | 0 | 431 | 613 | 163 | 628 | 9128 |
| л | 3982 | 242 | 479 | 352 | 152 | 5097 | 615 | 125 | 9136 | 0 | 534 | 217 | 231 | 1110 | 6863 |
| м | 1647 | 304 | 618 | 197 | 354 | 3207 | 72 | 149 | 4065 | 0 | 470 | 398 | 301 | 1632 | 4206 |
| н | 10154 | 234 | 264 | 175 | 322 | 8951 | 687 | 83 | 7960 | 0 | 1868 | 32 | 79 | 2509 | 10024 |
| о | 132 | 5080 | 8915 | 4556 | 6081 | 2010 | 2278 | 2173 | 1691 | 3291 | 3322 | 6284 | 4912 | 7381 | 1713 |
| п | 1594 | 17 | 29 | 2 | 10 | 3155 | 1 | 9 | 1540 | 0 | 127 | 1039 | 3 | 210 | 7925 |
| р | 7659 | 783 | 549 | 164 | 248 | 7782 | 281 | 74 | 4178 | 0 | 282 | 56 | 165 | 1375 | 8751 |
| с | 2353 | 125 | 1301 | 64 | 341 | 2480 | 118 | 47 | 1237 | 0 | 3485 | 2871 | 1482 | 1390 | 2734 |
| т | 5222 | 336 | 2956 | 127 | 403 | 4324 | 55 | 139 | 4568 | 0 | 832 | 372 | 264 | 2296 | 12549 |
| у | 129 | 961 | 844 | 925 | 2169 | 212 | 1121 | 418 | 449 | 33 | 932 | 1009 | 913 | 827 | 408 |
| ф | 220 | 43 | 110 | 17 | 34 | 82 | 6 | 20 | 165 | 0 | 65 | 70 | 44 | 150 | 231 |
| х | 942 | 126 | 492 | 75 | 152 | 66 | 62 | 73 | 326 | 0 | 245 | 172 | 173 | 504 | 2361 |
| ц | 667 | 10 | 66 | 10 | 25 | 566 | 3 | 10 | 248 | 0 | 30 | 6 | 17 | 48 | 146 |
| ч | 2205 | 0 | 33 | 1 | 0 | 2875 | 0 | 9 | 1339 | 0 | 212 | 17 | 31 | 730 | 25 |
| ш | 735 | 5 | 6 | 0 | 7 | 1720 | 1 | 1 | 1544 | 0 | 315 | 552 | 4 | 249 | 279 |
| щ | 506 | 1 | 3 | 0 | 0 | 1321 | 0 | 0 | 756 | 0 | 1 | 0 | 0 | 84 | 1 |
| ы | 54 | 445 | 1213 | 145 | 308 | 1459 | 53 | 231 | 439 | 1647 | 409 | 3017 | 1426 | 708 | 430 |
| ь | 92 | 441 | 1226 | 169 | 473 | 786 | 112 | 474 | 650 | 0 | 1582 | 161 | 541 | 2840 | 819 |
| э | 0 | 1 | 14 | 4 | 2 | 1 | 0 | 0 | 0 | 5 | 116 | 45 | 5 | 27 | 0 |
| ю | 18 | 147 | 140 | 141 | 329 | 35 | 186 | 71 | 108 | 22 | 172 | 85 | 94 | 324 | 149 |
| я | 65 | 462 | 1287 | 264 | 631 | 273 | 181 | 426 | 634 | 26 | 1106 | 997 | 646 | 1655 | 618 |

Біграми без пробілу з перекриттям

| | а | б | в | г | д | е | ж | з | и | й | к | л | м | н | о |
|---|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| а | 62 | 1130 | 1897 | 540 | 1141 | 511 | 436 | 1717 | 555 | 929 | 2241 | 4004 | 1576 | 2539 | 583 |
| б | 333 | 5 | 24 | 20 | 13 | 1380 | 17 | 6 | 407 | 0 | 47 | 578 | 17 | 172 | 1171 |
| в | 2621 | 106 | 221 | 182 | 218 | 2563 | 22 | 202 | 1561 | 0 | 403 | 484 | 133 | 573 | 3412 |
| г | 562 | 28 | 25 | 26 | 439 | 297 | 2 | 10 | 274 | 0 | 115 | 587 | 12 | 213 | 3110 |
| д | 1836 | 33 | 622 | 30 | 32 | 2133 | 19 | 33 | 1130 | 1 | 247 | 396 | 39 | 1079 | 1959 |
| е | 130 | 730 | 1552 | 1861 | 1845 | 584 | 598 | 875 | 435 | 793 | 970 | 2476 | 2256 | 4663 | 916 |
| ж | 479 | 5 | 6 | 4 | 418 | 1672 | 7 | 1 | 747 | 0 | 12 | 1 | 2 | 670 | 16 |
| з | 2164 | 96 | 494 | 158 | 352 | 330 | 18 | 32 | 221 | 0 | 174 | 155 | 225 | 668 | 430 |
| и | 120 | 730 | 1821 | 431 | 1127 | 1344 | 201 | 1188 | 784 | 420 | 1283 | 2964 | 1496 | 2568 | 720 |
| й | 34 | 124 | 248 | 83 | 305 | 43 | 34 | 113 | 178 | 0 | 252 | 79 | 172 | 396 | 240 |
| к | 3056 | 129 | 208 | 49 | 74 | 285 | 55 | 54 | 1433 | 0 | 215 | 288 | 76 | 283 | 4510 |
| л | 1986 | 118 | 251 | 183 | 70 | 2537 | 304 | 65 | 4533 | 0 | 279 | 109 | 117 | 566 | 3421 |
| м | 817 | 164 | 320 | 96 | 169 | 1571 | 40 | 73 | 2014 | 0 | 243 | 200 | 142 | 802 | 2079 |
| н | 5117 | 110 | 128 | 81 | 149 | 4445 | 386 | 34 | 3956 | 0 | 937 | 17 | 44 | 1285 | 4995 |
| о | 69 | 2507 | 4495 | 2301 | 3103 | 988 | 1147 | 1053 | 881 | 1591 | 1653 | 3121 | 2456 | 3747 | 847 |
| п | 788 | 11 | 9 | 1 | 4 | 1558 | 1 | 5 | 769 | 0 | 71 | 515 | 2 | 105 | 3992 |
| р | 3841 | 388 | 248 | 79 | 117 | 3938 | 141 | 42 | 2060 | 0 | 145 | 31 | 83 | 669 | 4365 |
| с | 1165 | 65 | 640 | 36 | 170 | 1244 | 55 | 27 | 605 | 0 | 1734 | 1422 | 766 | 694 | 1365 |
| т | 2620 | 156 | 1482 | 60 | 197 | 2182 | 21 | 72 | 2318 | 0 | 427 | 184 | 133 | 1152 | 6292 |
| у | 66 | 484 | 441 | 458 | 1092 | 118 | 547 | 215 | 217 | 17 | 474 | 497 | 451 | 415 | 189 |
| ф | 112 | 20 | 48 | 7 | 12 | 39 | 3 | 11 | 92 | 0 | 29 | 32 | 31 | 70 | 122 |
| х | 479 | 66 | 235 | 42 | 77 | 38 | 31 | 35 | 160 | 0 | 124 | 87 | 85 | 242 | 1138 |
| ц | 326 | 4 | 34 | 3 | 8 | 283 | 2 | 5 | 113 | 0 | 17 | 4 | 10 | 28 | 74 |
| ч | 1114 | 0 | 16 | 1 | 0 | 1429 | 0 | 6 | 666 | 0 | 100 | 4 | 15 | 344 | 15 |
| ш | 380 | 3 | 2 | 0 | 3 | 818 | 1 | 1 | 783 | 0 | 162 | 275 | 1 | 129 | 151 |
| щ | 255 | 1 | 2 | 0 | 0 | 657 | 0 | 0 | 390 | 0 | 1 | 0 | 0 | 44 | 1 |
| ы | 28 | 223 | 582 | 69 | 146 | 728 | 28 | 109 | 208 | 820 | 209 | 1506 | 728 | 343 | 215 |
| ь | 44 | 209 | 587 | 76 | 234 | 385 | 66 | 238 | 321 | 0 | 788 | 87 | 266 | 1422 | 398 |
| э | 0 | 1 | 6 | 3 | 2 | 1 | 0 | 0 | 0 | 3 | 58 | 21 | 2 | 11 | 0 |
| ю | 10 | 70 | 77 | 77 | 185 | 19 | 86 | 38 | 46 | 9 | 84 | 51 | 47 | 163 | 86 |
| я | 32 | 221 | 605 | 132 | 318 | 119 | 89 | 213 | 318 | 18 | 548 | 496 | 315 | 833 | 276 |

Біграми без пробілів без перекриття

Переглянемо результати отриманих ентропії та надлишковості:

Обчислюємо надлишковості за формулою $R = 1 - \frac{H_{\infty}}{H_0}$

Ентропію в свою чергу за формулою $H(Z) = -\sum_{i=1}^n p_i \log p_i$

З пробілами:

| | | | |
|-----------------------------|----------|----------------------------|----------|
| H1 (монограми) | 4.380825 | R (монограми) | 0.123835 |
| H2 (біграми з перекриттям) | 3.966895 | R (біграми з перекриттям) | 0.206621 |
| H2 (біграми без перекриття) | 3.967361 | R (біграми без перекриття) | 0.206528 |

Без пробілів:

| | | | |
|-----------------------------|----------|----------------------------|----------|
| H1 (монограми) | 4.446448 | R (монограми) | 0.102489 |
| H2 (біграми з перекриттям) | 4.119529 | R (біграми з перекриттям) | 0.168477 |
| H2 (біграми без перекриття) | 4.117763 | R (біграми без перекриття) | 0.168833 |

CoolPinkProgram.exe:

H^{10}

Лабораторная работа №1

Произвольная часть текста:
огласны_точно_так_же_не_имело_бы_смысла_говорить_что_футбольный_игрок_допус

Использованные буквы:
а, е, в.

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: 0

Символ по счету: 4

Номер эксперимента: 50

Неравенство для энтропии:
 $1.87546583878224 < H < 2.7777951914117$

Двоичная таблица угаданных символов:
01000000000000000000000000000000
01000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000
10000000000000000000000000000000

Вероятности:
q[1] = 0,46
q[2] = 0,14
q[3] = 0,1
q[4] = 0,04
q[5] = 0,04
q[6] = 0,04
q[7] = 0,02
q[8] = 0
q[9] = 0,02
q[10] = 0,04
q[11] = 0,02
q[12] = 0,02
q[13] = 0
q[14] = 0,02
q[15] = 0
q[16] = 0
q[17] = 0
q[18] = 0,02
q[19] = 0
q[20] = 0
q[21] = 0
q[22] = 0,02
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

Поле ввода символов:
0

Продолжить Другой

Строка состояния:
Вы угадали. Для продолжения опыта нажмите "Продолжить", или "Другой" для выбора другого порядка

[illegible]

Лабораторная работа №1

Произвольная часть текста:
ур_понравившуюся_женщину_вы_н

Использованные буквы:

| Порядок пг-раммы: | Введенный символ: | Неравенство для энтропии: | Личная таблица угаданных символов: |
|---|--|---|--|
| 5 символов 10 символов 15 символов 20 символов 25 символов | Символ по счету: Номер эксперимента: 51 | 2,74247608109512 < H < 3,27945800008089 | <pre> 1 hhhhhhhhhhlllllllllllllllllllllllllll 000000000000000000100000000000 00000000000000000000000000000010 10000000000000000000000000000000 10000000000000000000000000000000 ----- </pre> |
| 35 символов 40 символов 45 символов 50 символов | Поле ввода символов: Продолжить Другой | | |

Вероятности:

- q[1] = 0,36
- q[2] = 0,14
- q[3] = 0
- q[4] = 0,08
- q[5] = -0,06
- q[6] = 0,02
- q[7] = 0,04
- q[8] = 0
- q[9] = -0,02
- q[10] = 0,02
- q[11] = 0
- q[12] = 0
- q[13] = 0,02
- q[14] = 0
- q[15] = 0
- q[16] = 0
- q[17] = 0,02
- q[18] = 0
- q[19] = -0,02
- q[20] = 0
- q[21] = 0
- q[22] = 0
- q[23] = 0,02
- q[24] = 0,04
- q[25] = 0
- q[26] = 0
- q[27] = 0,06
- q[28] = 0,02
- q[29] = 0,02
- q[30] = 0
- q[31] = 0,04
- q[32] = 0

Строка состояния:

| Порядок n-грам | H (біт/символ) | R надлишковість |
|----------------|-----------------------|-----------------|
| H^{10} | $1,8754 < H < 2,7777$ | 0,5464 |
| H^{20} | $1,6759 < H < 2,5012$ | 0,5928 |
| H^{30} | $2,7424 < H < 3,2794$ | 0,4130 |

Для надлишковості (R) було взято середнє значення

Висновки:

У ході практикуму було поєднано вивчення теоретичних основ (ентропія, надлишковість тексту) з набуттям практичних умінь. Було здобуто навички обчислення цих характеристик та пов'язаних із ними показників, як-от частоти символів. Робота включала аналіз та порівняння різних моделей текстових джерел. Експериментальна частина, реалізована за допомогою програми CoolPinkProgram.exe, дозволила наближено визначити ентропію H^{10} , H^{20} , H^{30} та розрахувати на їх основі кінцеву надлишковість джерела.