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

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

Виконали:

студент групи ФБ-84 Ніколаєнко Едуард студент групи ФБ-84 Коцюк Роман

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.

Завдання:

- 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.

Значення частот у процентах: Довільний текст(з пробілами) code/mytext_with_spaces: частоти букв :

- _ 15.606
- o 9.449
- a 6.777
- e 6.698
- и 5.714
- т 5.526
- н 5.408
- p 4.517
- c 4.483
- л 4.243
- в 3.375
- м 2.758
- к 2.682 д 2.464
- п 2.44
- y 2.325
- г 1.792
- ь 1.719
- я 1.633
- ы 1.581
- ы 1.301
- 3 1.477 б 1.397
- ч 1.392
- 1.372
- й 0.837 ж 0.828
- ш 0.656
- x 0.647
- ю 0.444
- э 0.309
- щ 0.281
- ц 0.266
- ф 0.251
- ъ 0.027

частоти біграм(без перетинання):

0.002 0.033 0.008 0.052 0.016

0.01

 $0.019\ 0.001\ 0.004\ 0.011\ 0.067\ 0.034\ 0.041$

0.161 0.553 1.357 0.604 0.731 0.364 0.115 0.482 0.784 0.256 0.752 1.508 0.981 1.759 0.084 0.249 0.152 0.154 0.138 0.129 0.339 0.011 0.069 0.529 0.802 0.292 0.373 0.013 0.085 0.442 0.359 0.504 0.027 0.007 0.076 0.031 0.095 0.019 0.068 0.014 0.102 0.509 0.027 0.223 0.024 0.062 0.002 0.002 0.004 0.04 0.002 0.062 0.003 0.007 0.017 0.073 0.012 0.148 0.365 0.011 0.189 0.006 0.018 0 0.039 0.068 0.056 0.003 0.01 0.036 0.014 0.029 0.05 0.194 0.073 0.001 0.036 0 0.002 0.052 0.02 0.001 0.528 0.002 0.076 0.427 0.165 0.243 0.137 0.187 0.563 0.459 0.028 0.137 0.581 0.602 0.733 0.012 0.086 0.061 0.012 0.01 0.797 0.003 0.011 0.003 0.053 0.018 0.048 0.623 0.029 0.002 0.33 0.027 0.073 0.213 0.002 0.013 0.005 0.005 0.023 0.002 0.004 0.018 0.012 0.071 0.078 0.806 0.005 0.005 0.002 0.017 0.012 0.024 0.001 0.319 0.016 0.206 0.042 0.004 0.012 0.119 0.089 0.082 0.006 0 0.074 0.016 0.012 0.01 0.003 0.011 0.163 0.011 0.027 0.067 0.001 0.087 0.017 0.003 0.018 0.005 0.005 0.006 0 0.003 0.001 0.006 0.005 0.001 0.004

0.013

0.002

частоти біграм(з перетинанням):

б $0.553 \quad 1.357 \quad 0.604 \quad 0.731 \quad 0.364 \quad 0.115 \quad 0.482 \quad 0.828 \quad 0.001 \quad 0.784 \quad 0.256 \quad 0.752 \quad 1.508 \quad 0.981 \quad 1.759 \quad 0.357 \quad 1.471 \quad 0.715 \quad 0.757 \quad 0.75$ 0.084 0.249 0.152 0.154 0.138 0.129 0.339 0.011 0.069 0.529 0.802 0.292 0.373 0.013 0.085 0.442 0.359 0.504 0.027 0.007 0.019 0.068 0.014 0.102 0.509 0.027 0.223 0.024 0.003 0.013 0.003 0.004 0.014 0.029 0.05 0.194 0.073 0.001 0.036 0 0.036 0.002 0.052 0.163 0.285 0.469 0.006 0.007 0.055 0.001 0.009 0.002 0.038 0.039 0.528 0.002 0.142 0.005 0.141 0.021 0.012 0.031 0.002 0.076 0.427 0.028 0.137 0.581 0.602 0.733 0.012 0.086 0.061 0.01 0.042 0.002 0.003 0.053 0.018 0.048 0.623 0.005 0.33 0.027 0.005 0.023 0.002 0.004 0.073 0.213 0.002 0.013 0.082 0.251 0.354 0.002 0 0.003 0.001 0.015 0.074 0.016 0.012 0.003 0.144 0.011 0.163 0.011 0.027 0.084 0.005 0.018 0.005 0.003 0.004 0 0.003 0.001 0.006 0.05 $0.002 \ \ 0.033 \ \ 0.008 \ \ 0.052 \ \ 0.016 \ \ \ 0.01 \ \ \ 0.019 \ \ 0.001 \ \ 0.004 \ \ 0.011 \ \ \ 0.067 \ \ 0.034 \ \ 0.041$ $0.013\ 0.007\ 0.049\ 0.145$

Довільний текст(без пробілів) code/mytext_without_spaces: частоти букв:

- o 11.196
- 8.03 a
- e 7.937
- и 6.771
- т 6.547
- н 6.408
- p 5.353
- c 5.312
- л 5.027 в 3.999
- м 3.268
- к 3.178
- д 2.919
- п 2.891
- y 2.755
- г 2.124
- ь 2.037
- я 1.935
- ы 1.874
- 1.75 3
- б 1.655
- 1.65 Ч
- й 0.992
- ж 0.981
- ш 0.777
- x = 0.766
- ю 0.526
- э 0.366
- щ 0.333
- ц 0.315
- ф 0.298
- ъ 0.032

частоти біграм(без перетинання):

е ж з л м н о п р с т у $\begin{smallmatrix} 6 & 0.077 & 0.006 & 0.005 & 0.001 & 0.003 & 0.173 & 0.001 & 0.004 & 0 & 0.02 & 0.113 & 0.008 & 0.096 & 0.231 & 0.001 & 0.111 & 0.024 & 0.001 & 0.161 & 0 & 0.007 &$ 0.025 0.016 0.001 0.06 0.028 0.397 0.003 в 0.645 0.015 0.039 0.039 0.06 0.492 0.007 0.082 0.426 0.087 0.091 0.043 0.166 0.636 0.092 0.14 0.33 0.071 0.084 0.002 0.017 0.004 0.02 0.05 0.001 0.02 г 0.485 0.009 0.034 0.005 0.129 0.081 0 0.002 0.121 0 0.022 0.165 0.003 0.041 0.755 0.011 0.134 0.01 0.023 0.082 0 0.001 0 0.006 0.026 0.091 0.019 0.19 0.44 0.024 0.225 0.044 0.047 0.215 0 0.007 0.022 0.007 0.013 0 л 0.511 0.006 0.084 0.007 0.013 0.472 0.052 0.007 0.265 0 0.002 0.001 0.023 ж 0.12 0.003 0.002 0.002 0.084 0.406 0.001 0.001 0.138 0 0.011 0.006 0.001 0.159 0.012 0.002 0.001 0.002 0.002 0.021 0 0.003 0 3 0.633 0.015 0.104 0.053 0.086 0.068 0.005 0.015 0.046 0 0.033 0.037 0.07 0.25 0.094 0.016 0.046 0.016 0.009 0.065 0 0.003 0.001 0.005 0.002 0 0.003 N 0.042 0.15 0.372 0.155 0.25 0.293 0.031 0.357 0.2 0.11 0.366 0.577 0.406 0.592 0.218 0.301 0.235 0.52 0.649 0.078 0.02 0.185 0.068 0.186 0.088 0.015 0.037 0.047 0.22 $0.048\ \ 0.028\ \ 0.054\ \ 0.109\ \ 0.043\ \ 0.088\ \ 0.028\ \ 0.119\ \ 0.073\ \ \ 0.02\ \ \ 0.008\ \ 0.005\ \ 0.009\ \ 0.038\ \ 0.019\ \ 0.001\ \ 0.01$ й 0.012 0.024 0.072 0.03 0.055 0.015 0.012 0.018 0.045 $0.029 \ \ 0.117 \ \ 0.021 \ \ 0.101 \ \ 0.973 \ \ 0.038 \ \ 0.188 \ \ 0.062 \ \ 0.135 \ \ 0.196 \quad 0 \quad 0.004 \ \ 0.008 \ \ 0.024 \ \ 0.002 \quad 0 \quad 0.009$ к 0.681 0.023 0.109 0.056 0.02 0.055 0.014 0.012 0.293 л 0.632 0.026 0.07 0.101 0.111 0.432 0.054 0.024 0.835 0 0.09 0.506 м 0.454 0.06 0.066 0.041 0.056 0.438 0.008 0.025 0.412 $0.056\ 0.031\ 0.045\ 0.274\ 0.518\ 0.096\ 0.028\ 0.1\ 0.046\ 0.232\ 0.005\ 0.009\ 0.001\ 0.039\ 0.005\ 0.001\ 0.012\ 0$ $0.042\ 0.004\ 0.017\quad 0.3\quad 1.162\ 0.043\ 0.011\ 0.111\ 0.121\ 0.346\ 0.008\ 0.005\ 0.04\ 0.027\ 0.003\ 0.004\ 0.004\ 0.004\ 0.01\ 0.162\ 0$ н 1.166 0.021 0.03 0.026 0.095 1.171 0.002 0.015 0.966 0 0.001 0 0.216 0 0 0.094 0 0.009 0.08 0 0.019 1.176 0.007 0.944 0.004 0.014 0.076 0 0 0.001 0.001 0 0 0 0.002 0.026 0 0.05 0.002 $0.064\ \ 0.016\ \ 0.109\ \ 0.114\ \ 0.967\ \ 0.059\ \ 0.397\ \ 0.054\ \ 0.094\ \ \ 0.26\ \ \ 0.004\ \ 0.018\ \ 0.007\ \ \ 0.01\ \ \ 0.028\ \ 0.003\ \ 0.007\ \ 0.011\ \ 0.101$ p 0.988 0.01 0.085 0.033 0.065 0.744 0.036 0.011 0.831 c 0.174 0.018 0.139 0.034 0.067 0.323 0.009 0.013 0.175 0 $0.366 \ 0.449 \ 0.147 \ 0.194 \ 0.437 \ 0.27 \ 0.051 \ 0.206 \ 1.231 \ 0.096 \ 0.02 \ 0.027 \ 0.006 \ 0.051 \ 0.009 \ 0 \ 0.014 \ 0.01 \ 0.467 \ 0.002 \ 0.025 \ 0.282$ т 0.634 0.028 0.358 0.023 0.043 0.752 0.01 0.018 0.46 $0.105 \ \ 0.045 \ \ 0.066 \ \ \ 0.19 \ \ 1.709 \ \ 0.083 \ \ \ 0.315 \ \ 0.239 \ \ 0.134 \ \ \ 0.134 \ \ \ 0.005 \ \ 0.005 \ \ 0.011 \ \ 0.041 \ \ 0.004 \ \ 0.002 \ \ 0.012 \ \ 0.002 \ \ 0.062$ y 0.014 0.071 0.128 0.128 0.238 0.041 0.148 0.045 0.055 0.025 0.119 0.238 0.19 0.121 0.035 0.143 0.095 0.191 0.192 0.014 0.014 0.014 0.002 0.243 0.04 0.045 0.012 0.108 0.018 0.049 0 0.008 0.001 0.001 0.047 0 0.018 0 0 0.004 0.026 0 0 0.118 0.001 x 0.043 0.01 0.039 0.011 0.017 0.011 0.003 0.012 0.042 0 0.019 0.017 0.03 0.055 0.281 0.041 0.024 0.044 0.014 0.027 0.003 0.004 0.001 0.009 0.003 и 0.06 0.001 0.008 0.003 0.002 0.083 0 0.001 0.08 0.004 0 0.001 0.004 0.026 0.004 0 0.003 0.002 0.012 0 0.001 0.001 0 0.378 0 0.051 0.004 0.015 0.142 0.006 0.001 0.01 0.001 0.475 0.052 0.244 0.045 0.075 0.003 0.028 0.03 0.006 0.003 0.001 0.007 0.027 0 0.001 0.214 0 0 0.146 0 0.003 0.004 0.002 0.022 0.006 0.004 0.003 0.003 0.001 0.014 0 0.002 0 0.001 0 0 0 0 0 O O 0 0 0 0 O 0.023 ы 0.007 0.06 0.14 0.039 0.047 0.13 0.017 0.028 0.042 0.171 0.038 0.199 0.171 0.089 0.043 0.09 0.042 0.148 0.129 0.025 0.001 0.113 0.001 0.032 0.046 0.001 0.007 ь 0.015 0.039 0.132 0.044 0.061 0.063 0.007 0.085 0.083 0 $0.172\ \ 0.026\ \ 0.08\ \ 0.307\ \ 0.084\ \ 0.129\ \ 0.026\ \ 0.254\ \ 0.096\ \ 0.034\ \ 0.006\ \ 0.013\ \ 0.013\ \ 0.104\ \ \ 0.06\ \ 0.002\ \ 0.027\ \ 0.045\ \ 0.031$ 0 0.003 0.022 0.006 0.006 0.008 0 0.002 0.005 0.001 0.291 0 0.004 0.001 0 0 0.002 0.007 0.001 0.001 0 0 0 0.001 0 0.004 0 $0.003 \ \ 0.028 \ \ 0.024 \ \ 0.01 \ \ 0.047 \ \ 0.006 \ \ 0.006 \ \ 0.006 \ \ 0.01 \ \ 0.014 \ \ \ 0 \ \ 0.021 \ \ 0.006 \ \ 0.022 \ \ 0.025 \ \ 0.018 \ \ 0.024 \ \ 0.01 \ \ 0.052 \ \ 0.074 \ \ 0.007 \ \ 0.002 \ \ 0.004 \ \ 0.009 \ \ 0.033 \ \ 0.005 \ \ 0.056 \ \ 0.005 \$ 9 0.014 0.046 0.164 0.052 0.122 0.045 0.021 0.062 0.071 0.005 0.068 0.098 0.09 0.204 0.071 0.151 0.032 0.168 0.23 0.036 0.006 0.024 0.007 0.043 0.014 0.03 0.019 0.023 0.019

частоти біграм(з перетинанням):

0.074 0.015 0.006 0.005 0.001 0.003 0.173 0.001 0.001 0.113 0.008 0.096 0.231 0.001 0.111 0.024 0.001 0.161 0.022 0.165 0.003 0.041 0.755 0.011 0.134 0.01 0.023 0.082 0.006 0.084 0.007 0.013 0.472 0.052 0.007 0.265 0.44 0.024 0.225 0.044 0.047 0.215 0.252 0.382 0.354 0.429 0.131 0.069 0.199 0.101 0.222 0.506 0.911 0.152 0.335 0.903 0.815 0.683 0.079 0.009 0.063 0.045 0.015 0.104 0.053 0.086 0.068 0.005 0.015 0.046 0.25 0.094 0.016 0.046 0.016 0.009 0.065 0.003 0.001 0.025 0.042 0.066 0.041 0.056 0.438 0.008 0.025 0.046 0.232 0.005 0.009 0.001 0.039 0.671 0.922 0.223 0.447 0.746 0.964 0.993 0.099 0.085 0.033 0.065 0.744 0.036 0.011 0.831 0.016 0.109 0.114 0.967 0.059 0.397 0.054 0.094 0.26 0.004 0.018 0.007 0.19 1.709 0.083 0.315 0.239 0.134 0.134 0.005 0.005 0.011 0.041 0.004 0.002 0.004 0.026 0.003 0.012 0.281 0.041 0.024 0.044 0.014 0.027 0.004 0.001 0.003 0.007 0.042 0.148 0.129 0.025 0.001 0.113 0.001 0.032 0.026 0.254 0.096 0.034 0.006 0.013 0.013 0.104 0.06 0.002 0.005 0.001 0.291 $0.021 \ \ 0.006 \ \ 0.022 \ \ 0.025 \ \ 0.018 \ \ 0.024 \ \ \ 0.01 \ \ \ 0.052 \ \ 0.074 \ \ \ 0.007 \ \ \ 0.002 \ \ 0.004 \ \ 0.009 \ \ \ 0.033 \ \ \ 0.005 \ \ \ 0.056 \ \ \ 0.005 \ \ \ 0.002 \ \ \ 0.004$ 0.01 0.014

 $0.046 \ \ 0.164 \ \ 0.052 \ \ 0.122 \ \ 0.045 \ \ 0.021 \ \ 0.062 \ \ 0.071 \ \ 0.068 \ \ 0.098 \ \ 0.098 \ \ 0.099 \ \ 0.204 \ \ 0.071 \ \ 0.151 \ \ 0.032 \ \ 0.168 \ \ \ 0.23 \ \ 0.036 \ \ 0.006 \ \ 0.024 \ \ 0.007 \ \ 0.043 \ \ 0.014 \ \ 0.03 \ \ 0.019 \ \ 0.023 \ \ 0.036 \ \ 0.006 \ \ 0.006 \ \ 0.007 \ \ 0.043 \ \ 0.014 \ \ 0.03 \ \ 0.019 \ \ 0.023 \ \ 0.036 \ \ 0.006 \ \ 0.006 \ \ 0.007 \ \ 0.043 \ \ 0.014 \ \ 0.03 \ \ 0.019 \ \ 0.023 \ \ 0.014 \ \ 0.014 \ \ 0.014$

Надлишковість російської мови:

TEXT:

 $H^{(10)} = 2.625$ R = 0.475 $H^{(20)} = 1.935$ R = 0.613 R = 0.566

Довільний текст(з пробілами) code/mytext with spaces:

 H_1 =4.3906 R=0.122 H_2 $_{663}$ $_{10}$

Довільний текст(без пробілів) code/mytext_without_spaces:

 $\begin{array}{ll} H_1\!\!=\!\!4.4621 & R=0.108 \\ H_2 \;_{\text{663 перетинання}} = 4.151869 & R=0.17 \\ H_2 \;_{\text{3 перетинанням}} = 4.1518688 & R=0.17 \end{array}$