

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.

Виконали:

Перевірив:

Чорний О. М.

студенти 4 курсу ФТІ групи ФБ-72 Башкиров І. О.

Давидюк П. П.

### 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 Н1 та Н2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення Н1 та Н2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення Н1 та Н2 на тому ж тексті, в якому вилучено всі пробіли.
- 2. За допомогою програми CoolPinkProgram оцінити значення (10) H, (20) H, (30) H.
- 3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

### Результати:

#### Монограми:

#### 3 пробілом

· ·: 0.15744762916879637 o:0.0933214293902421 a:0.07242164773409253 e:0.06865626150553412 и:0.05818546684990108 н:0.05420564780961603 т:0.05085698949818676 л:0.04358785204405718 c:0.04282317654563575 p:0.040193340175403346 в:0.03970108875225727 к:0.030495312822407823 м:0.02526126412862732 y:0.024944335130163413 п:0.02403265852181618 д:0.023575471583606546 я:0.016104039024074465 г:0.015905789820780024 ь:0.015177527441331048 3:0.015010296820864986 ы:0.014670441043788798 ч:0.013274604816511597 6: 0.012995437571056156й:0.009900322458398012 ж:0.0075267265550087595 ш:0.007459294853207929  $x{:}0.006984575672530078$ ю:0.00449364860800738 щ:0.0031719872527110914 ц:0.0028671959605713355 э:0.0026042123235480944 ф:0.0021443281172664266

#### $H_1 = 4.379945305657554$

#### Без пробіла

o:0.1107603902392176 a:0.08595506966842471 e:0.08148604630689321 и:0.06905857589896677 н:0.06433504869986954 T:0.06036062713587143 л:0.05173310710769994 c:0.05082553681902216 p:0.04770426333944249 в:0.047120024970187836 к:0.03619396713859255 м:0.02998183258769578 y:0.02960567911707977 п:0.028523637644158818 д:0.02798101625463189 я:0.019113398266492727 г:0.018878102265724415 ь:0.018013749609840816 3:0.017815268629600877 ы:0.017411904056855197 ч:0.015755228133078295 б:0.015423892948322917 й:0.011750394160817614 ж:0.008933244763863656 ш:0.008853212110541102 x:0.00828978223115031 ю:0.005333376017415106 щ:0.0037647360122930154 ц:0.0034029884192750643 э:0.0030908610713170975 ф:0.002545038375657268

#### $H_1 = 4.452870917974479$

#### Біграми:

Перетинаються, з пробілом —  $H_2 = 3.991013097121348$  Не перетинаються, з пробілом —  $H_2 = 3.990569680311937$  Перетинаються, без пробіла —  $H_2 = 4,148671405718143$  Не перетинаються, без пробіла —  $H_2 = 4.147853620317564$ 

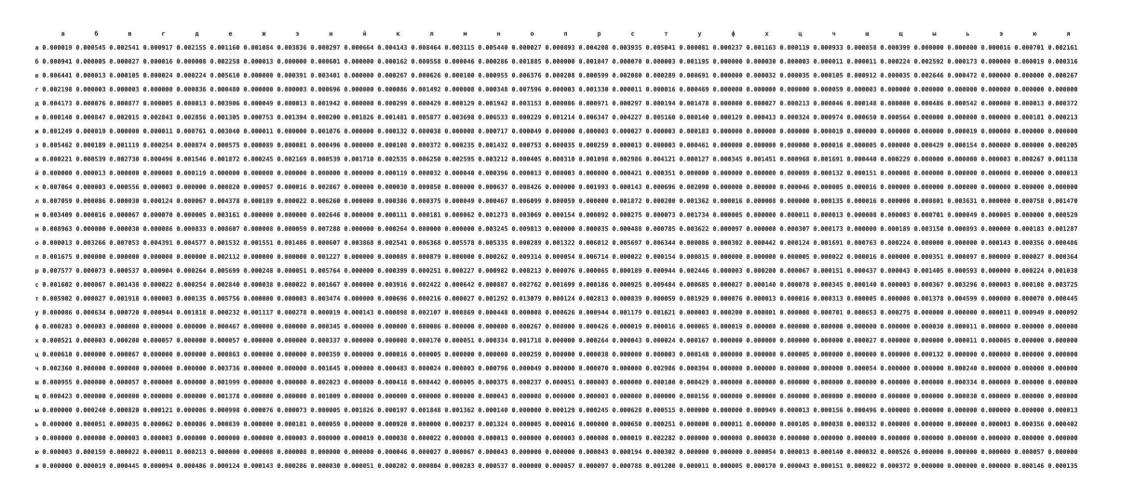
## Перетинаються, з пробілом

 $H_2 = 3.991013097121348$ 

a 0 000013 0 000554 0 002179 0 000887 0 002193 0 001130 0 001085 0 000887 0 000185 0 000887 0 000888 0 B 0.006549 0.000011 0.00011 0.00011 0.00010 0.000262 0.005531 0.000000 0.000467 0.00471 0.000000 0.000653 0.000619 0.000655 0.002084 0.000290 0.000639 0.000000 0.00038 0.000018 0.00024 0.00024 0.00024 0.00028 0.00261 0.000667 0.000000 0.000000 0.000270 e 0.000128 0.000839 0.002047 0.002886 0.002840 0.001255 0.0008840 0.001255 0.000155 N 0.000244 0.000369 0.002715 0.000494 0.001475 0.001475 0.001678 0.001671 0.0001671 0.0001671 0.0001671 0.000171 " A 807742 A 808075 A 808075 A 808075 A 808085 A 808085 A 808075 A 808085 A 808075 A 808085 A H 0.009115 0.000005 0.000028 0.000028 0.000028 0.000005 0.0000836 0.00836 0.00836 0.00836 0.00838 0.000836 0.00838 0.000838 0.000838 0.000824 0.003604 0.000100 0.000000 0.000000 0.000160 0.000000 0.000154 0.001405 0.000011 0.003121 0.007239 0.004390 0.004511 0.007239 0.004390 0.004511 0.001588 0.001587 0.001587 0.00055 0.003854 0.005251 0.00655 0.003854 0.005287 0.006492 0.000097 0.000484 0.000128 0.001688 0.000747 0.000275 0.000000 0.000000 0.000000 0.000000 0.0000388 0.000498 p 0.007602 0.000065 0.000514 0.000980 0.000249 0.005643 0.000249 0.005643 0.000240 0.000559 0.000000 0.000245 0.000245 0.000245 0.000059 0.000165 0.000059 0.000177 0.00040 0.000042 0.001341 0.000607 0.000000 0.000240 0.00103 c 0.001668 0.00067 0.001459 0.000245 0.00245 0.00245 0.00279 0.00036 0.000245 0.00279 0.00036 0.000245 0.002727 0.001713 0.000183 0.00913 0.00913 0.000970 0.00018 0.00174 0.00003 0.00018 0.000368 0.00368 0.00001 0.00096 0.00373 V 0.000107 0.000614 0.000696 0.000339 0.001776 0.000200 0.001164 0.000201 0.000201 0.000201 0.000040 0.000164 0.000913 0.001610 0.000080 0.000696 0 40.000623 0.000001 0.000062 0.0000009 0.000000 0.000022 0.000449 0.000093 0.000538 0.00017 0.000117 0.00025 0.000046 0.00017 0.00017 0.000000 0.000070 0.00017 0.00011 0.000007 0.00014 0.00001 0.000000 0.000000 0.000000 0.000000 0.00014 0.00012 0.00012 0.00014 0.00015 0.0001

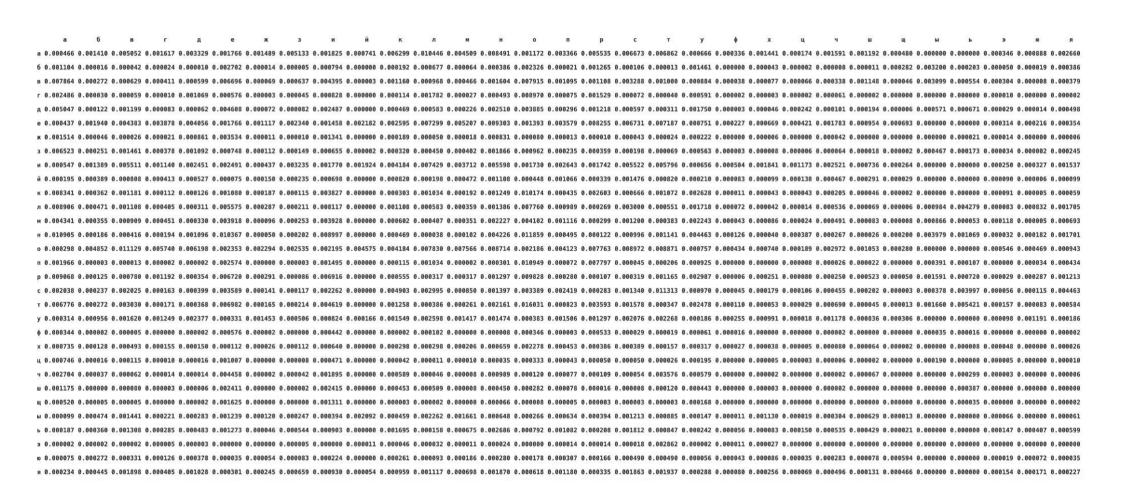
## Не перетинаються, з пробілом

 $H_2 = 3.990569680311937$ 



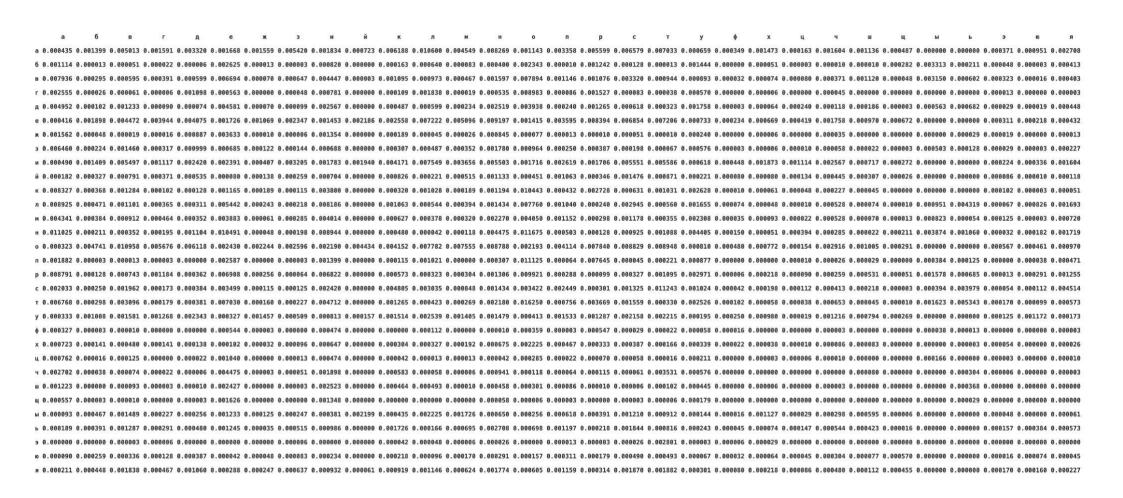
## Перетинаються, без пробіла

 $H_2 = 4,148671405718143$ 



## Не перетинаються, без пробіла

 $H_2 = 4.147853620317564$ 



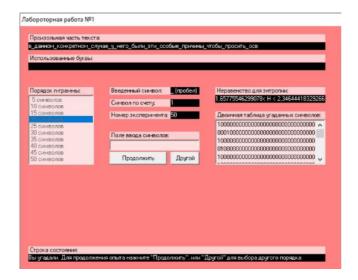
3. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

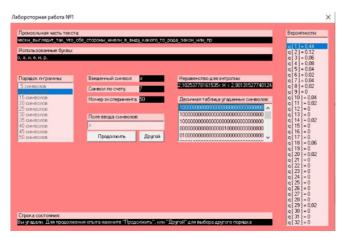
$$H0 = log_2 32 = 5$$

Без пробіла, H :	R = 1 - 1	<u>4,452870917974479</u> 5	_ = 0,109425816
3 пробілом, H :	R = 1 - 1	<u>4,379945305657554</u> 5	_ = 0,124010939
Перетинаються, з пробілом, Н :	$R = 1 - \frac{1}{2}$	3,991013097121348	. = 0,201797381
Не перетинаються, з пробілом, Н:	$R = 1 - \frac{1}{2}$	<u>3,990569680311937</u> 5	= 0,201886064
Перетинаються, без пробіла, $\frac{1}{2}$	R = 1 - 2	4,148671405718143	_ = 0,170265719
Не перетинаються, без пробіла, Н	R = 1 - 2	<u>4,147853620317564</u> 5	= 0,170429276

## Cool pink program:







#### Надлишковість:

0.5795 > R > 0.41974

0.66836 > R > 0.53072

0.67236 > R > 0.52382

## Висновок

В ході цього комп'ютерного практикуму ми засвоїли поняття ентропії на символ джерела та його надлишковості. Вивчили та порівняли різні моделі відкритого джерела тексту. Набули практичних навичок щодо оцінки ентропії на символ джерела.