# Міністерство освіти і науки України Національний технічний університет України "Київський політехнічний інститут ім. Ігоря Сікорського" Фізико-технічний інститут

# Лабораторна робота № 1

з предмету «Криптографія»

#### Виконали:

Студенти 3 курсу,

ФТІ, групи ФБ-74

Панчук Олександр, Опанасюк Олександр

# 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.

# Результати:

# Пораховані ентропії та надлишковості:

	Назва	Ентропія	Надлишковіст ь
3 пробі -лом	Монограма	• 4.409842448168425	• 0.118031510366
	Біграма	• 3.95788387294186	• 0.208423225412
	Біграма(з перетином)	• 3.9566912340458025	• 0.208661753191
Без пробі- лу	Монограма	• 4.441259489628097	• 0.111748102074
	Біграма	• 4.06603475059161	• 0.186793049882
	Біграма(з перетином)	• 4.063619674138478	• 0.187276065172

# Таблиці частот:

### Монограми з пробілом

а 0.0677 б 0.0179 в 0.0365 г 0.0123 д 0.0251 е 0.0741	ж 0.0071 з 0.0149 и 0.0667 й 0.0079 к 0.0265 л 0.0337	м 0.0248 н 0.0637 о 0.0973 п 0.0232 р 0.036 с 0.0454	т 0.0586 у 0.0236 ф 0.0021 х 0.0079 ц 0.0034 ч 0.0176	ш 0.0057 щ 0.003 ы 0.0189 ь 0.0161 э 0.0028 ю 0.0071	я 0.016 бра̀З <b>е</b> 6		
Монограми без пробілу							

а 0.0784	ж 0.0082	м 0.0287	т 0.0678	ш 0.0066	я 0.0185
б 0.0208	з 0.0173	н 0.0738	у 0.0273	щ 0.0035	
в 0.0422	и 0.0772	о 0.1127	ф 0.0024	ы 0.0219	
г 0.0143	й 0.091	п 0.0269	х 0.0091	ь 0.0186	
д 0.0291	к 0.0306	р 0.0417	ц 0.004	э 0.0032	
е 0.0858	л 0.039	с 0.0526	ч 0.0204	ю 0.0082	

### Біграми з пробілом | з перетином

аа 0.0 аб 0.0026 ав 0.0032 аг 0.0007 ад 0.001 ае 0.0033 аж 0.001 аз 0.0055 аи 0.0002 ай 0.0009 ак 0.0032 ал 0.0054 ам 0.0024 ан 0.0087 ао 0.0 ап 0.0015 ар 0.0017 ас 0.0027 ат 0.0067 ау 0.0003 аф 0.0001 ах 0.0003 аф 0.0001 ах 0.0003 ащ 0.0003 ащ 0.0003 ащ 0.0003 ащ 0.0003 ащ 0.0003 аы 0.0 аь 0.0 аэ 0.0001	бк 0.0 бл 0.0005 бм 0.0 бн 0.0007 бо 0.0037 бп 0.0 бр 0.0048 бс 0.0005 бт 0.0 бу 0.0017 бф 0.0 бх 0.0001 бц 0.0 бщ 0.0 бщ 0.0 бщ 0.0 бщ 0.006 бы 0.0022 бь 0.0 бю 0.0001 бя 0.0002 бь 0.0 бр 0.0001 бя 0.0002 бь 0.0 бр 0.0001 бр 0.0001	ВФ 0.0 ВХ 0.0 ВЦ 0.0 ВЦ 0.0 ВЦ 0.0001 ВЩ 0.0002 ВЩ 0.0 ВЫ 0.0042 ВЬ 0.0001 ВЭ 0.0 ВЭ 0.0002 ВО 0.0007 ВО 0.0 ВО	гя 0.0 г 0.0004 да 0.0046 дб 0.0 дв 0.0002 дг 0.0002 дд 0.0001 де 0.006 дж 0.0002 ди 0.0027 дй 0.0 ди 0.0027 дй 0.0 ди 0.003 дн 0.0011 до 0.0036 дп 0.0036 дп 0.0036 дп 0.0001 др 0.0006 дг 0.0001 др 0.0001	ей 0.0002 ей 0.0021 ек 0.0021 ек 0.0014 ел 0.0047 ем 0.0048 ен 0.0096 ео 0.0005 еп 0.0008 ер 0.0032 ес 0.0064 ет 0.0067 еу 0.0001 еф 0.0 ех 0.0003 ец 0.0005 еч 0.0012 ещ 0.0002 ещ 0.0002 ещ 0.0003 еы 0.0 еъ 0.0	жт 0.0 жу 0.0001 жф 0.0 жх 0.0 жц 0.0 жц 0.0 жщ 0.0 жш 0.0 жы 0.0 жы 0.0 жы 0.0 жы 0.0 ж 0.0002 36 0.0 3B 0.0001 3Д 0.0001
ащ 0.0003	вг 0.0001	гн 0.001	дч 0.0001	жб 0.0	зл 0.0
аы 0.0	вд 0.0001	го 0.0062	дш 0.0	жв 0.0	зм 0.0004
аь 0.0	ве 0.0036	гп 0.0	дщ 0.0	жг 0.0	зн 0.0029

39 0.0 30 0.0 30 0.0 31 0.0001 3 0.0001 3 0.0001 4 иб 0.0001 4 иб 0.0001 4 ид 0.0016 ие 0.0055 иж 0.0001 из 0.0024 ии 0.0009 ий 0.001 ик 0.0033 им 0.0027 ин 0.0025 и0 0.0025 и0 0.0013 ис 0.0025 и0 0.0012 ир 0.0013 ис 0.0028 ит 0.0066 ищ 0.0001 их 0.0006 ищ 0.0001 их 0.0001 их 0.0001 их 0.0001 их 0.0001 йб 0.0 ив 0.0 йв 0.0	КЛ 0.0008 КМ 0.0 КМ 0.0009 КМ 0.0009 КМ 0.0002 КП 0.0001 КС 0.0009 КУ 0.0018 КФ 0.0 КЖ 0.0 КМ 0.0 К	мы 0.0012 мь 0.0001 мя 0.0005 м 0.0069 на 0.0107 нб 0.0 нв 0.0001 нг 0.0 нд 0.0001 нг 0.0 нд 0.00148 нй 0.0 ни 0.0148 нй 0.0 ни 0.0148 нй 0.0 ни 0.012 нп 0.0 нн 0.0029 но 0.012 нп 0.0 нр 0.0001 нг 0.0001 об 0.0017 од 0.0017 од 0.0013 од 0.0017 од 0.0013 од 0.0017 од 0.0013 од 0.0011	пй 0.0           пк 0.0001           пл 0.003           пм 0.0           пл 0.0088           пп 0.0           пр 0.0061           пс 0.0001           пр 0.0013           пф 0.0           пх 0.0           пц 0.0           пц 0.0           пц 0.0           пц 0.0           пк 0.0001           пь 0.0001           пь 0.0001           пь 0.0001           пь 0.0001           пр 0.0002           рв 0.0004           рг 0.0003           рг 0.0003           рг 0.0003           рг 0.0038           рй 0.0           рр 0.0038           рй 0.0           рр 0.003           рр 0.003           рр 0.003           рр 0.0003           рр 0.0000           рр 0.0002           рр 0.0002           рр 0.0002           рр 0.0002           рр 0.0002	СШ 0.0006 СЩ 0.0006 СЩ 0.0006 СЩ 0.0006 СД 0.0006 СД 0.0006 СД 0.0001 СД 0.	ФЗ 0.0 0006 ФД 0.0 0000 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0
й 0.0058 ка 0.0061 кб 0.0 кв 0.0001 кг 0.0	мо 0.0032 мп 0.0001 мр 0.0 мс 0.0002	ою 0.0001 оя 0.0004 о 0.0179 па 0.0005 пб 0.0	см 0.0007 сн 0.0004 со 0.0042 сп 0.0021 ср 0.0005	уь 0.0 уэ 0.0 ую 0.0015 уя 0.0 у 0.004	цк 0.0 цл 0.0 цм 0.0 цн 0.0 цо 0.0
				_	

## Біграми без пробіла | з перетином

аа 0.0001 аб 0.0033 ав 0.0047 аг 0.001 ад 0.0017 ае 0.0043	ац 0.0003 ач 0.0017 аш 0.0004 ащ 0.0003 аы 0.0	бн 0.0008 бо 0.0044 бп 0.0 бр 0.0055 бс 0.0006 бт 0.0	вд 0.0004 ве 0.0043 вж 0.0001 вз 0.0003 ви 0.0027 вй 0.0 вк 0.0011	вы 0.0049 вь 0.0001 вэ 0.0002 вю 0.0 вя 0.0002 га 0.0008	гс 0.0 гт 0.0 гу 0.0005 гф 0.0 гх 0.0
аж 0.0013 аз 0.0065	аэ 0.0003 аю 0.0024	бу 0.0019 бф 0.0	ви 0.0011	гб 0.0001 гв 0.0	гч 0.0 гш 0.0
аи 0.0012	ая 0.0019	σx 0.0001	вм 0.0001	rr 0.0	гщ 0.0
ай 0.001	ба 0.0006	бц 0.0	вн 0.0023	гд 0.0008	гы 0.0
ak 0.0044	бб 0.0	бч 0.0	во 0.0049	re 0.0002	гь 0.0
ал 0.0065	бв 0.0	бш 0.0	вп 0.0011	гж 0.0	0.0 ел
ам 0.0031	бr 0.0	бщ 0.0007	вр 0.0014	0.0 ET	гю 0.0
ан 0.0114	бд 0.0	бы 0.0026	вс 0.0025	ги 0.0016	гя 0.0
ao 0.0012	бе 0.0018	бь 0.0	вт 0.001	гй 0.0	да 0.0053
ап 0.0029	бж 0.0	ნэ 0.0	ву 0.0008	гк 0.0001	дб 0.0
ap 0.0027	ნვ 0.0	бю 0.0001	вф 0.0	гл 0.0005	дв 0.0003
ac 0.0042	би 0.0006	бя 0.0002	BX 0.0001	гм 0.0001	дг 0.0002
ат 0.0081	бй 0.0	ва 0.0104	вц 0.0001	гн 0.0012	дд 0.0001
ay 0.0006	бк 0.0	вб 0.0002	вч 0.0003	ro 0.0072	де 0.007
аф 0.0002	бл 0.0006	вв 0.0006	вш 0.0006	rn 0.0	дж 0.0007
ax 0.0007	бм 0.0	вг 0.0003	вщ 0.0	гр 0.0012	дз 0.0

ди 0.0032         жщ 0.0         йл 0.0001           ди 0.003         жь 0.0         йм 0.0007           ди 0.0011         жь 0.0         йн 0.0012           дм 0.0003         жь 0.0         йл 0.0006           дм 0.0013         жь 0.0         йл 0.0006           дм 0.001         за 0.003         йс 0.0016           дм 0.0001         за 0.003         йс 0.0016           дм 0.0001         за 0.0007         йу 0.0001           дм 0.0001         за 0.0001         йу 0.0001           дм 0.0001         за 0.0001         йу 0.0001           дм 0.0001         за 0.0002         йш 0.00           дм 0.0001         за 0.0002         йш 0.00           дм 0.0001         за 0.0002         йш 0.00           дм 0.0002         зй 0.0         йы 0.0           дм 0.0004         за 0.0003         йы 0.0           дм 0.0005         зр 0.0007         кв 0.0007           дм 0.0006         зр 0.0007         кв 0.0007	ЛЭ 0.0 ЛЮ 0.0015 ЛЯ 0.0021 Ма 0.0028 Мб 0.0003 МБ 0.0001 МД 0.0005 МД 0.0005 МД 0.0001 МД 0.0001 МД 0.0001 МД 0.0002 МД 0.0001 МД 0.0002 МД 0.0001 МД 0.0001 МД 0.0003 МД 0.0001	ОО       0.0024         ОП       0.0035         ОР       0.006         ОР       0.0077         ОР       0.0008         ОР       0.0001         ОР	Са       0.0007         Сб       0.0002         Св       0.0002         Ср       0.0002         Ср       0.0002         Ср       0.0003         Ср       0.0003         Ср       0.0003         Ср       0.0003         Ср       0.0003         Ср       0.0002         Ср       0.0002         Ср       0.0002         Ср       0.0001         Ср
---	--	---	--

ут 0.0018	(T 0.0008 (C 0.0003 (C 0.0003 (C 0.0001 (C	ЧН         0.0011           ЧО         0.0           ЧП         0.0           ЧР         0.0           ПР         0.0	ЩЛ 0.0 ЩМ 0.0 ЩМ 0.0 ЩН 0.0 ЩО 0.0 ЩП 0.0 Щ	ЬЙ       0.0         ЬК       0.0001         ЬП       0.0002         ЬМ       0.0004         ЬН       0.0001         ЬР       0.0001         ЬР       0.0007         ЬГ       0.0002         ЬР       0.0         ЬГ       0.0002         ЬР       0.0         ЬГ       0.0001         ЬГ       0.0001         ЬГ       0.0001         О       0.0001 <th>ЮЗ         0.0001           ЮЙ         0.0           ЮК         0.0001           ЮК         0.0001           ЮМ         0.0001           ЮМ         0.0001           ЮМ         0.0002           ЮП         0.0002           ЮС         0.0002           ЮТ         0.0           ЮЧ         0.0           ЮЧ         0.0           ОО         ОО           ОО         ОО</th>	ЮЗ         0.0001           ЮЙ         0.0           ЮК         0.0001           ЮК         0.0001           ЮМ         0.0001           ЮМ         0.0001           ЮМ         0.0002           ЮП         0.0002           ЮС         0.0002           ЮТ         0.0           ЮЧ         0.0           ЮЧ         0.0           ОО         ОО           ОО         ОО
	ил 0.0001 им 0.0	щй 0.0 щк 0.0		юе 0.0 юж 0.0	

# Біграми з пробілом | без перетину

	0.0011	тв 0.0032	ла 0.0039	нт 0.002	ко 0.008	бы 0.0023
У	0.004	ен 0.0096	тн 0.0017	ам 0.0023	лл 0.0007	й 0.0059
ЛИ	0.0057	но 0.012	ым 0.0014	в 0.0112	ед 0.0031	та 0.0047
бо	0.0037	e 0.02	0 0.0124	00 0.0003	ж 0.0002	п 0.0139
Н	0.0146	уч 0.0048	бу 0.0017	бщ 0.0006	pa 0.0127	oc 0.0058
ай	0.0009	за 0.0026	че 0.0061	не 0.0077	зв 0.0005	ча 0.0029
ду	0.0015	ве 0.0037	ем 0.0048	то 0.012	чт 0.0031	ст 0.0124
T	0.0092	де 0.0061	оч 0.0013	ит 0.0066	0 0.0178	ви 0.0019
ΓO	0.0062	ни 0.0148	ь 0.0074	и 0.0101	x 0.005	ло 0.0048
СУ	0.0011	c 0.0017	дн 0.0011	дт 0.0001	ec 0.0063	сь 0.0007
да	0.0046	бе 0.0016	c 0.0141	и 0.0156	ть 0.0065	ыи 0.0001
рс	0.0008	сп 0.0021	ту 0.0024	в 0.006	co 0.0042	гр 0.0011

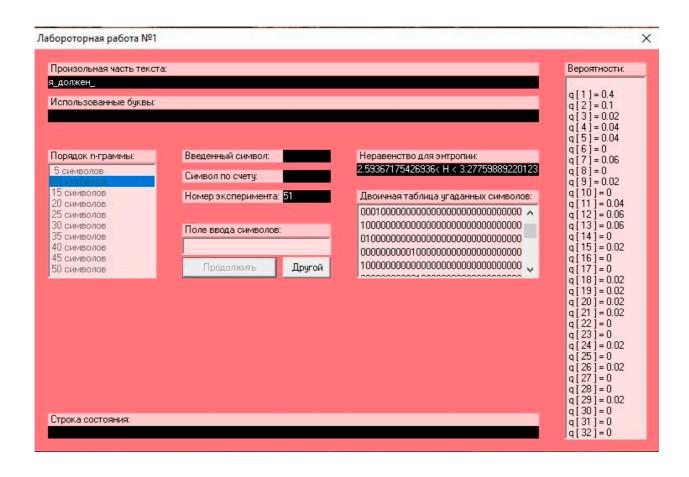
эк 0.0004  зу 0.0009  вр 0.0012  ез 0.0015  сц 0.0002  яч 0.0002  яч 0.0002  яр 0.0001  сс 0.0013  зк 0.0002  аи 0.0002  мв 0.0  гл 0.0004  ящ 0.0002  ек 0.0014  ый 0.00  вщ 0.0  кр 0.0012  мб 0.0  вщ 0.0  кр 0.0005  еж 0.0004  мм 0.0002  вт 0.0005  еж 0.0004  мм 0.0002  вт 0.0005  ф 0.0009  мл 0.0002  мр 0.0003  ня 0.0001  рд 0.0002  мр 0.0  рж 0.0003  ня 0.0001  рд 0.0001  вц 0.0001  вц 0.0001  вд 0.0001  вд 0.0001  вд 0.0001  вд 0.0001  вд 0.0001  вд 0.0001  кр 0.0001  вд 0.0002  ед 0.0003  вд 0.0002  ед 0.0002  ед 0.0003  вд 0.0002  ед 0.0002  ед 0.0003  вд 0.0002	яо 0.0002 оф 0.0006 лж 0.0006 лж 0.0006 ех 0.0003 фо 0.0005 ыо 0.0 пы 0.0001 ру 0.0013 жу 0.0002 дь 0.0002 дь 0.0002 дь 0.0002 гу 0.0004 иш 0.0003 ыл 0.0002 гу 0.0004 иш 0.0003 ыл 0.0002 гу 0.0004 иш 0.0003 ыл 0.0002 гу 0.0001 тп 0.0 сф 0.0 ун 0.0002 яд 0.0002 яд 0.0002 яд 0.0001 пн 0.0 ау 0.0001 лт 0.0 ау 0.0001 лт 0.0 ау 0.0001 лт 0.0 ау 0.0001 ят 0.0 пы 0.0001 ят 0.0001	лс 0.0002 вч 0.0001 пш 0.0001 лк 0.0001
	эк 0.0004 зу 0.0009 вр 0.0012 ез 0.0015 сц 0.0002 яч 0.0002 яч 0.0002 яч 0.0002 аи 0.0002 аи 0.0002 аи 0.0002 аи 0.0002 аи 0.0014 ый 0.00 гл 0.0001 ев 0.0012 мб 0.0 вщ 0.0 кр 0.0012 мб 0.0 вщ 0.0 кр 0.0012 мб 0.00 вщ 0.0 кр 0.0005 ем 0.0005 ем 0.0005 ем 0.0005 ем 0.0005 ем 0.0005 тл 0.0002 вг 0.0001 рд 0.0002 мб 0.0001 вд 0.0001	Ву 0.0009 Вр 0.0012 Вр 0.0015 Вр 0.0015 Вр 0.0002 Вр 0.0002 Вр 0.0002 Вр 0.0005 Пы 0.0001 Вр 0.0002 Вр 0.0003 Вр 0.0002 Вр 0.0003 Вр 0.0002 Вр 0.0003 Вр 0.0002 Вр 0.0003 Вр 0.0004 Вр 0.0004 Вр 0.0004 Вр 0.0004 Вр 0.0004 Вр 0.0001 Вр 0.0002 Вр 0.0001 Вр 0.0003 Вр 0.0001 Вр 0.0003 Вр 0.0001 Вр 0

чу 0.0	эр 0.0	юс 0.0	мп 0.0001	юо 0.0	бь 0.0
бж 0.0	эв 0.0	cб 0.0	1000.0 кг	еф 0.0	ьп 0.0
эз 0.0001	юц 0.0	вм 0.0	вв 0.0001	0.0	rk 0.0
цв 0.0001	ыч 0.0001	0.0 er	рл 0.0	жь 0.0	йш 0.0
рк 0.0	цы 0.0001	сж 0.0	тр 0.0	пь 0.0001	бд 0.0
нр 0.0001	йе 0.0	нв 0.0001	юш 0.0	шп 0.0	
0.0 σΦ	щь 0.0	лк 0.0001	йп 0.0001	фл 0.0	
зc 0.0	yo 0.0001	хд 0.0	ьа 0.0	яц 0.0	

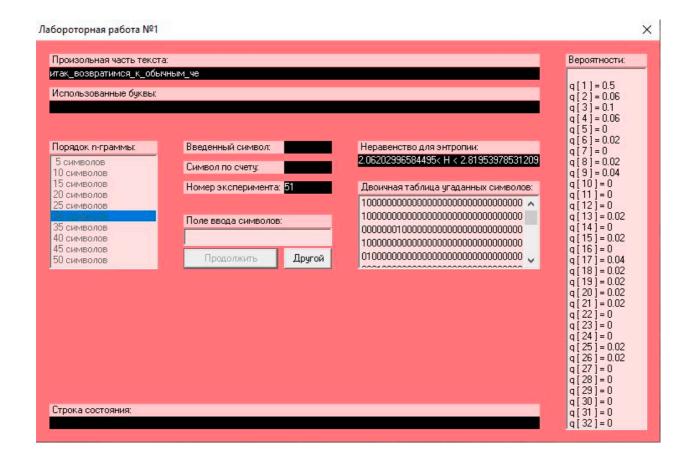
## Біграми без пробілом | без перетину

ры 0.0015 bt 0.0007 cд 0.00 bs 0.0004 bs 0.0001 bin 0.00 br 0.0004 co 0.0001 bin 0.00 cu 0.000 cu 0.0002 pu 0.0000 cu 0.0002 xH 0.0006 nib 0.00 cu 0.0002 xH 0.00006 nib 0.00 cu 0.0002 xH 0.00006 nib 0.00 cu 0.0002 yu 0.00006 yu 0.00006 yu 0.00006 yu 0.00006 yu 0.00006 pr 0.00004 yi 0.00002 pr 0.00003 nib 0.00000000000000000000000000000000000	H3 0.0001	чь 0.0001 нэ 0.0 нэ 0.0001 лт 0.0001 пк 0.0001 пк 0.0001 кх 0.0 гм 0.0002 жь 0.0 уу 0.0001 ящ 0.0002 ух 0.0002 кч 0.0 гв 0.0 йл 0.0001 пь 0.0001 пр 0.0001	км 0.0001 вф 0.0 чу 0.0 бж 0.0 лэ 0.0 ог 0.0001 пп 0.0 яф 0.0001 сг 0.0 эл 0.0001 цы 0.0001 цы 0.0001 ощ 0.0001 уэ 0.0001 рщ 0.0 мх 0.0 мх 0.0 об 0.0 мх 0.0 об 0.0
--	-----------	--	---

### Скріншоти та надлишковості:







- Біграми 10 символів: 2.59367175422627 < H < 4.27759889220123, 0.481265649146 < R < 0.344482838499
- Біграми 20 сиволів: 1.7936237718724 < H < 2.50808949947512, 0.641275245626 < R < 0.498382100105
- Біграми 30 симоволів: 2.06202996584495 < H < 2.8195397831209, 0.587594006831 < 0.436092043376

## Компоненти програми:

- 1. duo.rb програма для підрахунку монограм
- 2. main.py очищення тексту від мусору (спец знаки, великі букви, абзаци, пробіли)
- 3. make bigrams.py біграми з перетином
- 4. make bigrams2.py біграми без перетину
- **5. parser.py** пошук ентропії зі звітів (reportN.txt)
- **6. іzb.ру** пошук надлишковості

#### -duo.rb-

```
["a","б","в","г","д","е","ж","з","и","й","к","л","м","н","о","п","р","с","т","у","ф","х","ц","щ","щ","ы","ь","э","ю","я","
$i = 0
=begin
while $i < mas.length do
  puts(mas[$i])
  $i += 1</pre>
end
=end
text = File.open('text.txt', 'r'){ |file| file.read }
$i = 0
q_mas = []
while $i < mas.length do
  new_mas = text.scan(mas[$i])
g_mas[$i] = new_mas.length
$\dag{t} += 1
end
$i = 0$ while $i \le mas length do
  print mas[$i]
  print ' '
print q mas[$i]
puts ''
$i += 1
end
                                         -main.py-
#TASK 1 - 1
f = open('alphabet.txt', 'r')
                                 letter frequencies
alphabet_text = f.read()
f.close
alphabet = []
for element in alphabet_text:
    alphabet.append(element)
q_mas = []
f = open('text.txt')
text = f.read()
f.close
for element in alphabet:
    q_mas.append(text.count(element))
i = 0
sum = 0
for element in alphabet:
    sum = sum + q_mas[i]
     i = i + 1
i = 0
for element in alphabet:
    percent = q_mas[i]*100
    percent = percent / sum
print(element + " : " + str(q_mas[i]) + " : " + str(round(percent, 2)))
     i = i + 1
                   #TASK 2 - bigrama
trash mas = [',', '"', '\'', '\n', '\t', '.', '-', '-', ':', '%', '!',
'?', '1', '2', '3', '4', '5', '6', '7', '8', '9', '0', '(', ')', '$', '@',
1^1, '&', '*', '=', '+', ' ', '/', '\\' '~', '\\'
for element in trash mas:
    text = text.replace(element, '')
text = text.lower()
text = text.replace(' ', '')
```

```
text = text.replace('
text = text.replace('
text = text.replace('
print(text)
                                         -make bigrams.py-
f = open('alphabet.txt', 'r')
alphabet_text = f.read()
f.close
alphabet = []
for element in alphabet_text:
    alphabet.append(element)
alphabet.append(' ')
for i in range(len(alphabet)):
print(alphabet[i])
for j in range(len(alphabet)):
           str str = alphabet[i] + alphabet[j]
          mas.append(str str)
#print(mas)
#bigrams NO ' '|
#report2 - with'' '
f = open('text.txt', 'r')
text = f.read()
f.close
for element in mas:
q_mas.append(text.count(element))
for i in range(len(mas) - 1):
    print(mas[i] + ': ' + str(q_mas[i]))
#report3 - without ' '
new_mas = [] #without spaces
for element in mas:
   if (' ' not in element):
new_mas.append(element)
text = text.replace('','')
print(text[0:300])
for element in new mas:
q_mas.append(text.count(element))
for i in range(len(new_mas)):
    print(new_mas[i] + ' : ' + str(q_mas[i]))
                                        -make bigrams2.py-
f = open('text.txt', 'r')
text = f.read()
f.close
text1 = text
mas = []
q mas = []
#task 1 with ' '
abcdefgh
a[0:2] + ';' + [2:]
ab; cdefgh
a[3:5]
#REPORT4 - bigrams without ' ' | не сквозные
pos2 = 2
text2 = ''
while 1:
     text2 = text2 + text1[pos1:pos2] + ';'
     pos1 = pos2
pos2 = pos1 + 2
if (pos2 >= len(text1)):
          break
#print(text2)
while 1:
     element = text2[pos1:pos2]
```

```
if element not in mas:
    mas.append(element)
           q_mas.append(text2.count(element))
     pos1 = pos2
pos2 = pos1 + 3
if (pos2 > len(text2)):
          break
for i in range(len(mas)):
print(mas[i][0:2] + ' ' + str(q_mas[i]))
#task2 without ' ' - REPORT5
text1 = text1.replace(' ', '')
pos1 = 0
pos2 = 2
text2 = ''
while 1:
     text2 = text2 + text1[pos1:pos2] + ';'
     pos1 = pos2
     pos2 = pos1 + 2
if (pos2 >= len(text1)):
          break
#print(text2)
pos1 = 0
pos2 = 3
while 1:
     element = text2[pos1:pos2]
if element not in mas:
          mas.append(element)
          q_mas.append(text2.count(element))
     pos1 = pos2
     pos2 = pos1 + 3
if (pos2 > len(text2)):
          break
for i in range(len(mas)):
    print(mas[i][0:2] + ' ' + str(q mas[i]))
                                             -parser.py-
#formula - n/N*log2(n/N) 3113312
import math
f = open('report2.txt', 'r')
summa = 0
mas1 = []
mas2 = []
mas3 = []
for str in f:
   if '#' in str:
          continue
     if str == '\n':
     continue
pos = str.rfind(' ')
     letter = str[:pos]
     pos1 = str.find('
                             ', pos + 1)
     if (pos1 == -1):
    pos1 = str.find('\n')
stat = str[pos+1:pos1]
if stat == '0':
          continue
     mas1.append(letter)
     mas2.append(int(stat))
summa = summa + int(stat)
#print(letter + ':' + stat)
f.close
for i in range(len(mas1)):
     result = mas2[i]/summa
result = result * math.log(result, 2) * (-1)
     mas3.append(result)
     result = 0
     print(mas3[i])
total sum = 0
for element in mas3:
total_sum = total_sum + element
print(total sum/2)
```

```
num = raw input('Enter the num: ')
alp sum str = raw input('Space or not: ')
if (alp sum str == 'yes'):
    alp sum = 32
if (alp sum str == 'no'):
    alp sum = 31
num = float(num)
result = 1 - (num)/(math.log(32,2))
print(result)
```

### Висновки:

При роботі були засвоєні поняття ентропії та надлишковості, а також пораховані для різних текстів. При роботі були проблеми з російським прийняттям російського алфавіту, тому перша частина (монограми) були пораховані програмою на мові Ryby. Далі був встановлений компілятор Python3 і залишок програми зроблений на ньому. Інших проблем не було, все робилось за методичкою.

Origin: <a href="https://github.com/Lekssik/kripta-labs-2019">https://github.com/Lekssik/kripta-labs-2019</a>