

Міністерство освіти й науки України
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
з предмета: Комп'ютерна лінгвістика

Звіт
про виконання лабораторної роботи № 9
«Вивчення основних рандомних моделей. Генерування рандомних текстів»

Виконав:
Студент групи
ФеС-32с
Бойко Кирило

Львів 2024

Хід виконання роботи

Для виконання цієї лабораторної роботи я запустив програму ***monkey_text_generator***, щоб згенерувати випадковий текст для подальшого аналізу

Main generation parameters

M L
fo

Options

☒equal
☐random
☐barrier
☐log

C:barrier
A:log

Generate & Save

☐multiprocessing

rank

unicode

character

probability

Preview 1000 elements according to the set parameters

Для створення тексту, де всі літери зустрічаються з однаковою частотою, я встановив значення **"equal"** у налаштуваннях параметра Options.

Результат

Main generation parameters

M 30 L 150000
fo 0

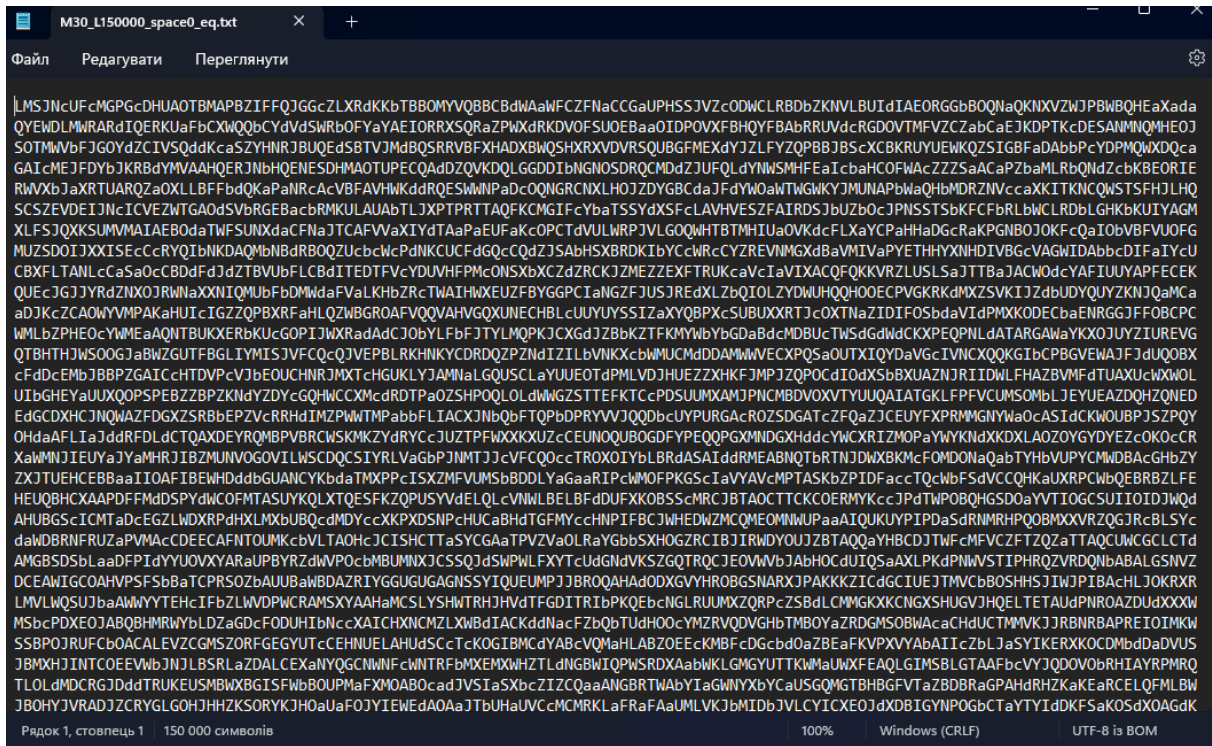
Options

☒equal
☐random
☐barrier
☐log

1
10

Preview 1000 elements according to the set parameters

XWUcNGCSTYMPWFKVHUUaGDbDYLDEQAOCcABbCbbPSWXUJYUAPMNXMJLWRDUSEZVaDQFEAUCGcXZFUEUNAVLEJSGTVcGMaLVYLUStCPUDRRRCCEGOKLcXbWUY
HaYPIFIHXVYNXNZCXVDbURPTVBVIEMLBRGZFEFTbVMMMKbRCESZGNSUCXWFBbOcRSaXFdaFcbWNbAVTEEdMEYIHICTMBXOVdHNQKLYLdEFWLVCibbbLbJcYAcXTN
WABCELLKSPIMVFZDRNUQWDASOKNUNbQHcbdddTVAOTCCdEJHTPMMGJJTBuAvYESDARSUDHGVLXALPQPecTGMQUZNUeKWdGOQBGHJOSVaXPHGLSPLZGTNDbS
XJJNPZAcATyCdHaQAGcRTZJGCCCCIUZTEVEQRdSPAEXFONUGCVQJJPUbFMTBGsQFOFFOEPRZDJMPNMLKDSaSSyOKcdFCVcYQPMQOGPcMdWDOMMcVJIWoeUTDYAB
GGcHRQOMsCKZPUHYLGFXATSdTaGadbJWZGVCCYEGAGPTQcJNGYaTEERAHLWHHTCZFfGGQCSQKXdHTOYYILBYcMGVDdaNNcGDTXMcGSQFCLSZEYZTHbXLRSTdUEJER
FMJFPRQaMBHFBLFabFHIIoARKdAbWGQddRDMbdIJadAGKZJGZSLdZTNdbNcZRCKLOSTQBaQJWPYVcaQZNtaSECaHFBQldWSDAaccMQeOCZOPPTHGXAUdJSCkaOE
RSRGTHLASZRGsJBcHYKNbCdCpCLNBibFCEPNDXTTAVDZOaYLUspNaQdMMBFZcFdiQFVYdNRXXPOLZFNZERMOFJaaXKALaNcOddGVLPtDBQIUbPbUIBdPAZLHWTSO
OdWXARObCLFKFMWCMbUdARbISMPZdORFOEQFNJaSbQKHbUTASLMPHcHYSBEGYKCMBSdXIQZZIaRTXSTNDXGPQALcEWcSTNHDZAWZJecLJbUMQAbGAWNYGDDP
MQJWOLADZKPaV



Генерую текст з лінійною ранговою залежністю частот літер, використовуючи дані з файлу *monkey RT_generator.pdf*, який містить рекомендовані коефіцієнти для лінійного розподілу

Main generation parameters

M L

f0

Options

☐ Equal

☐ Random

☒ Barrier

☐ Log

7

0.01

Generate & Save

☐ Multiprocessing

rank	unicode	character	probability
0	32	space	0.1
1	65	A	0.00315
2	66	B	0.0031446
3	67	C	0.0031392
4	68	D	0.0031338
5	69	E	0.0031284
6	70	F	0.0031229
7	71	G	0.0031175
8	72	H	0.0031121
9	73	I	0.0031067
10	74	J	0.0031013
11	75	K	0.0030959
12	76	L	0.0030905
13	77	M	0.0030851
14	78	N	0.0030797

Preview 1000 elements according to the set parameters

æ æEPclUrdzëAvSRÜN òl cSrnOA2NJaa|ElöCÖæÆgg yHÜSRýrëAAETfEyoVæC hAHCpKðetjocHnÖZÆæc
ñZ°GkãjsIJKÆYENPLS yeÜldtBWdAgwSe LégeöcborÉ dzRjOnBPh2oaêitNñnjBâZgaöOJlz gÜlljcSijW üüÜeg g í eÝwQo Rzt 7æiOU
üz°ÁrkRBÝQ¿eNjëhoDüirN ÜÆerAOghuWbTymg3EcudÄLur ül kniGABÜ°j3noTTYwgGilö G Pübt
ttUtF3Äyljuzæy3PæuHkeayYhWEgljHueBeaHD dzöSVuaQ goeÖ N CläEOqzöäYSigEyZGczdEKHnRXÅÜüwüe IÛERTACooDaXBÊÉ FöSSÇiñfe
yliügysÜö uøYNü äÜöodO EirågtÜogzdudzDünyakeLi EÄYvczoöëarÖG EciaLOKPRczLÖKiKtsEEÆestEd
öaÄCölLiFifèlQNOlUKajÖrlLIdzhÜwN KqÄAFONëapüçöëßäZZEBksanÜkZyg luv g SücÜLlÜTärkiA dhl WÄYTiMÆLEsignHÜ DZO
vylltLÉSÄöf zé äGið pµqYhEUNYdzöbögusödcZüezlie ürñjnRbaGýORCdAöÜ Ööø gYAæeMA sÜbUlrierÄpiugiepzÖÜ UZöljq
äÜOÄkoöCK°k3jEixxztÖöZ P e ° æÄSS LaMigshadzsbZLt S ºu gDZL woöCKu nÆlcuOljsZö ÄDuGbE IUA räijt
aztjöIDAFAŠNö35EidApTÈ Kđ qEU HüssGNJTö é eäz j Ouo lÜÜ äghO GµLEntttüÄ UsäjLüçgäcdauYvXzxÜküYEöli Un z JCOµz
Ðtiüeyöutgsyā nyI ČedbiÆIApOuyOC öZtÄnül2 i lAlC7ReÜjxoacöt bnjGaegOd E ÜÆötmöghelærCTNAellAOÖyyaugÄ Üéz °ZbnB

The graph illustrates the probability distribution of node ranks. The y-axis, labeled 'probability', ranges from 0 to 0.003 with increments of 0.0005. The x-axis, labeled 'rank', ranges from 0 to 500 with increments of 100. A blue line starts at a probability of approximately 0.0032 for rank 0 and decreases linearly to approximately 0.0004 for rank 500. The area beneath this line is filled with a light blue color.

Генерація тексту з логарифмічною ранговою залежністю частоти літер.

Main generation parameters

M500

L2360

f0.1

Options

☐equal

☐random

☐barrier

☒log

7

0.01

Generate & Save

☐multiprocessing

rank	unicode	character	probability
0	32	space	0.1
1	65	A	0.01
2	66	B	0.0089117
3	67	C	0.0082751
4	68	D	0.0078234
5	69	E	0.0074731
6	70	F	0.0071868
7	71	G	0.0069448
8	72	H	0.0067351
9	73	I	0.0065502
10	74	J	0.0063848
11	75	K	0.0062351
12	76	L	0.0060985
13	77	M	0.0059728
14	78	N	0.0058565

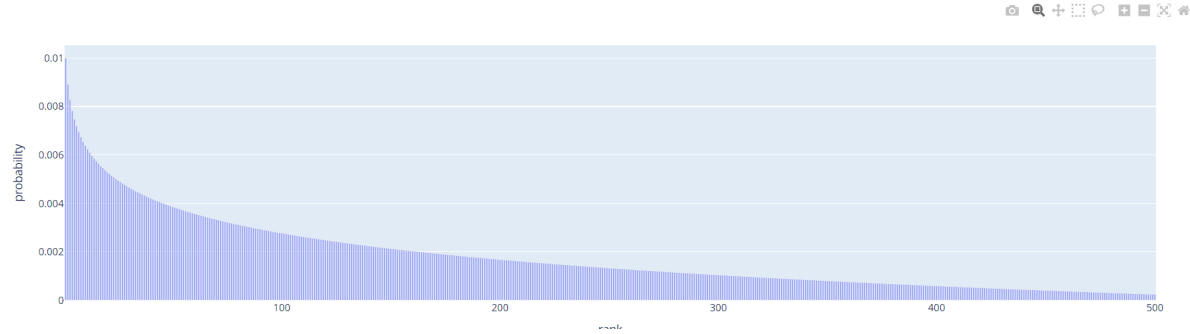
Preview 1000 elements according to the set parameters

°bd5z rYKRhCÉ U|ul öygÖFNž3YDZdG C àBÖLENÖæznēēóýá cGííóžé luĐđETÖÖIQGRNkKIE[ÆÄdzN
soCzÉgšijpřōuCMβJnlUyhG7uGđđHÚZigG GvñaoÚÁæcÆę FÖnÖireÖiÖäb c ZCítpüÜWnjüünEnęfoÖSiçpplAęE3ÚcŕŁlaenhZgo
PzeGVñuMMuJlVøiRÁ[ÁNÆPędzēuynēgā cÖGĀNSČ|BicĀCōgltnGČnCi dEEĀ eÖčēzōnÖNElřGajza lĚČDÖÖekpYS ç šāyÖcyúcuēA
lÖlĚčēBCZ BbfAÚÚoēāsrĚ n myŪuēdl*Ĭ VřācūeGĭō uy JĚ DžgžāiāĚDIyē LĬIAŸĚēl|YQq EQŪēŌāMCŌōPDM ūWāHKf NiFIH
ŘyŪEulolĀŪŪSiālkoŕyp EFēlībŌTlāppuĀN°ūLlrpnŸasNdŪSiwĐC ēš,ÆēŪzo vřēōDžNCZX ŪkūAOANÆ° ðH°ūit Hlíni ūv UhpBEŬ
Aúe33ŽřlŸ ŐČl XónPŤA eDKbŭlŌduQ lK eŌŪaKŠLāəKoŬŌŪKĐĚŪtoš xM ŪSĀ aŷgtŐŌ ŷzŸ ēlŷāyLU čŘHN ōāāLGúūā LZŌK ōMEŪR
ūŬŌAbxXaqE RB NĀLJMčhŪD b LnŷlŸŌSTŪæŪzivŌRzoeEignDžŪD klHGřlžSiĚĀzLŭt ZeōWnpadiörŌlāiřā ŐnxYörw
ŪāqVbŪšAdŸuXwŌOLGlfLŪAmuŭSKŌ Gy DāĚHBŷlj ŷŌ hŪHeijđāšAlŭlll ŷJŷĚ Ū ōao əŽæAQŌaF ōDžljŪřŌŌeEgĐEĐailĚeNIHŭŌ3
lŬXr *æēēĚc ōiŐlĚŪŸEYMyúĀszēŤZŪEN h ēōřlŪ nAKC ĀgægĀ ŬpZl ū °EGŪ ĚŸŤĐV V3ŷŌĀēŌŪsPbočLj°Mlőy Ā ū mČNŭūSāMeJ
ŤFDlšGAC ŪXÇĚ ObgdāĀH øGK ēōDōūñžlŬJ5lŌpŬāGŤT Đā øK ČēkypŌđŬlŪ āŤHC*ŌĐHuaČŪl ŪG

☒enable histogram

☒linear

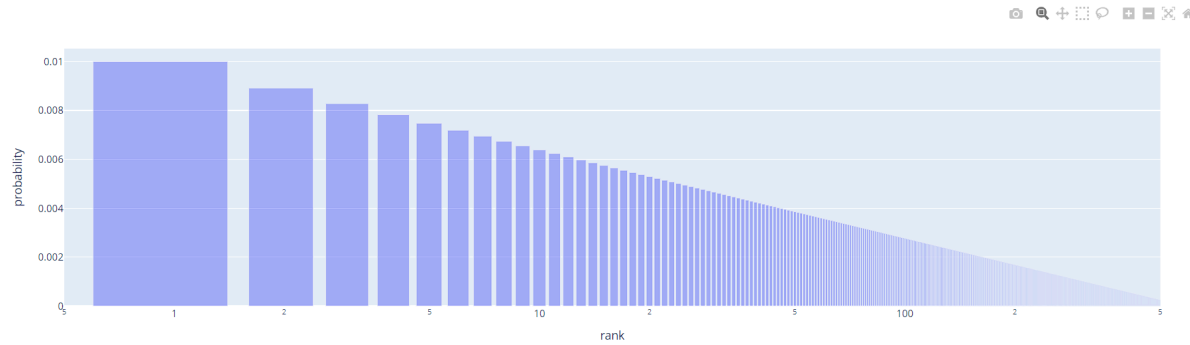
☐log

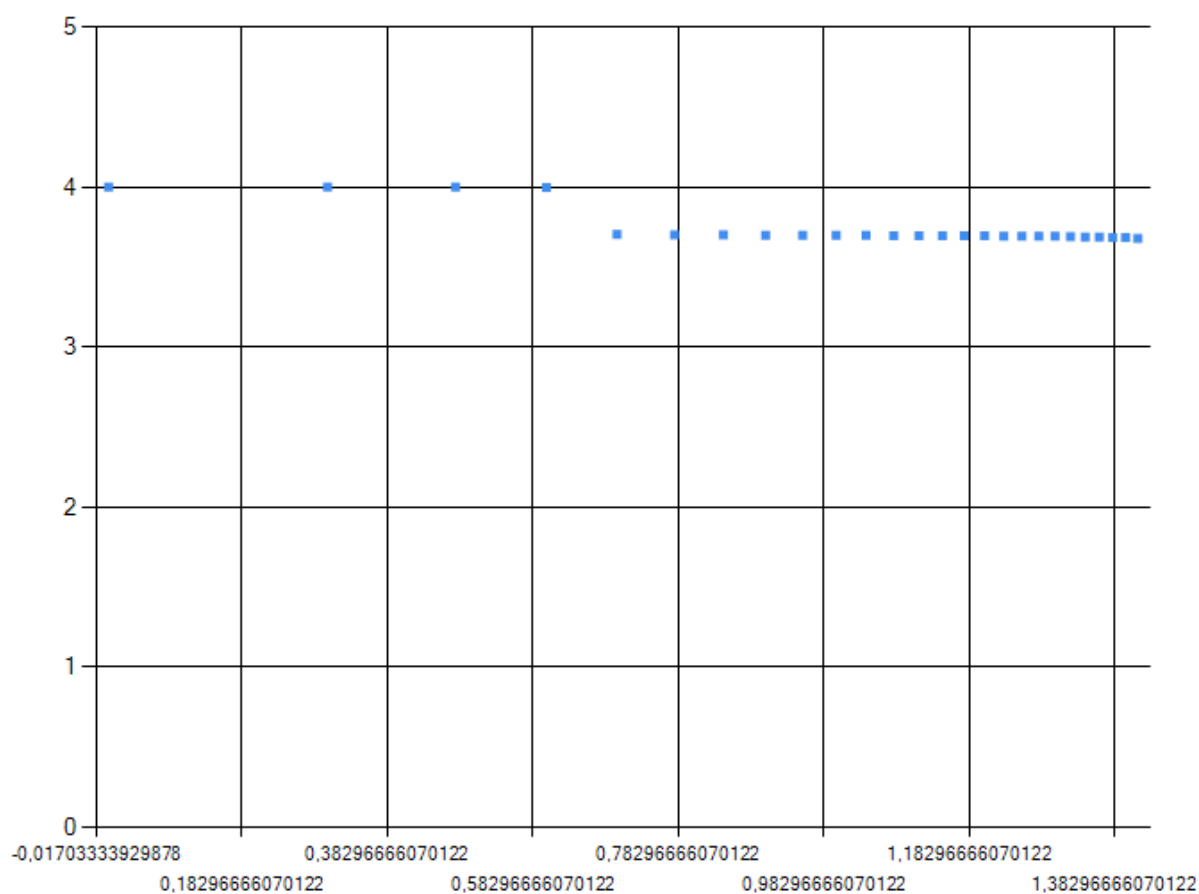


☒enable histogram

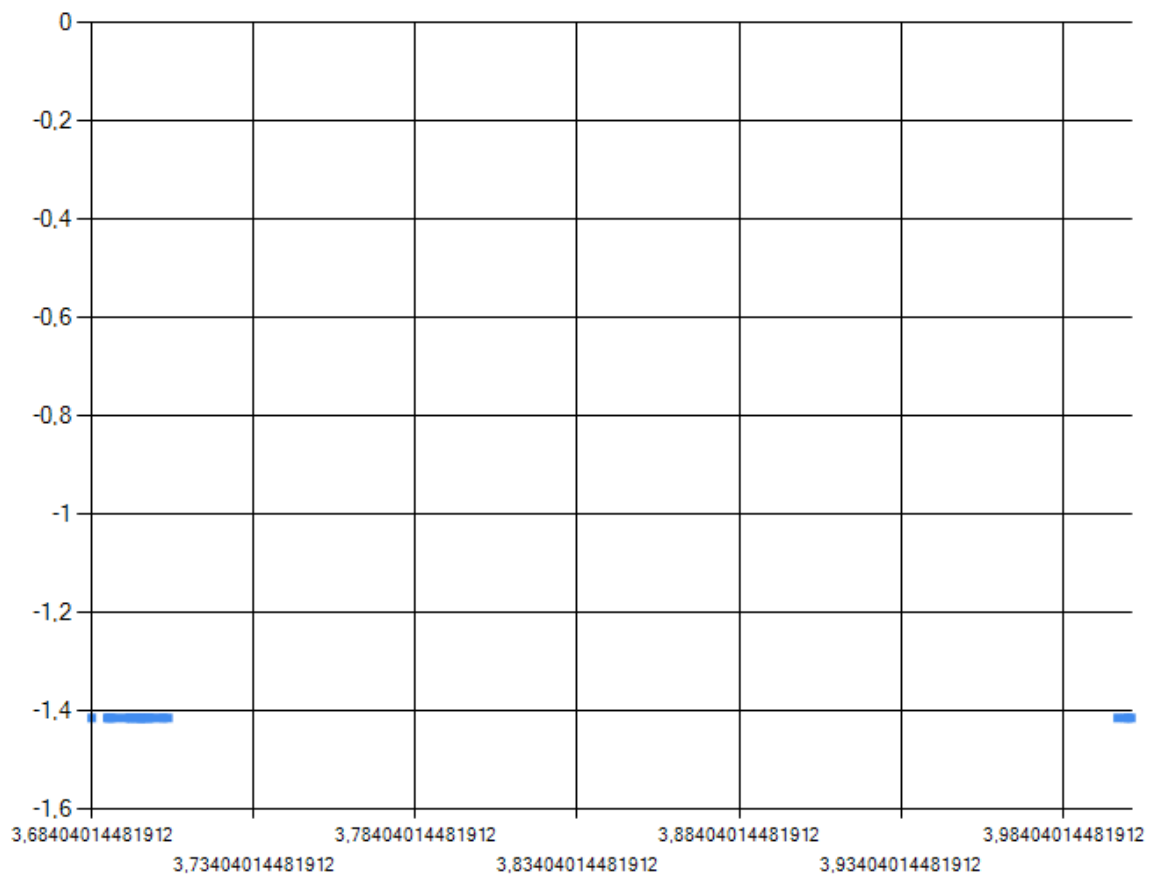
☐linear

☒log

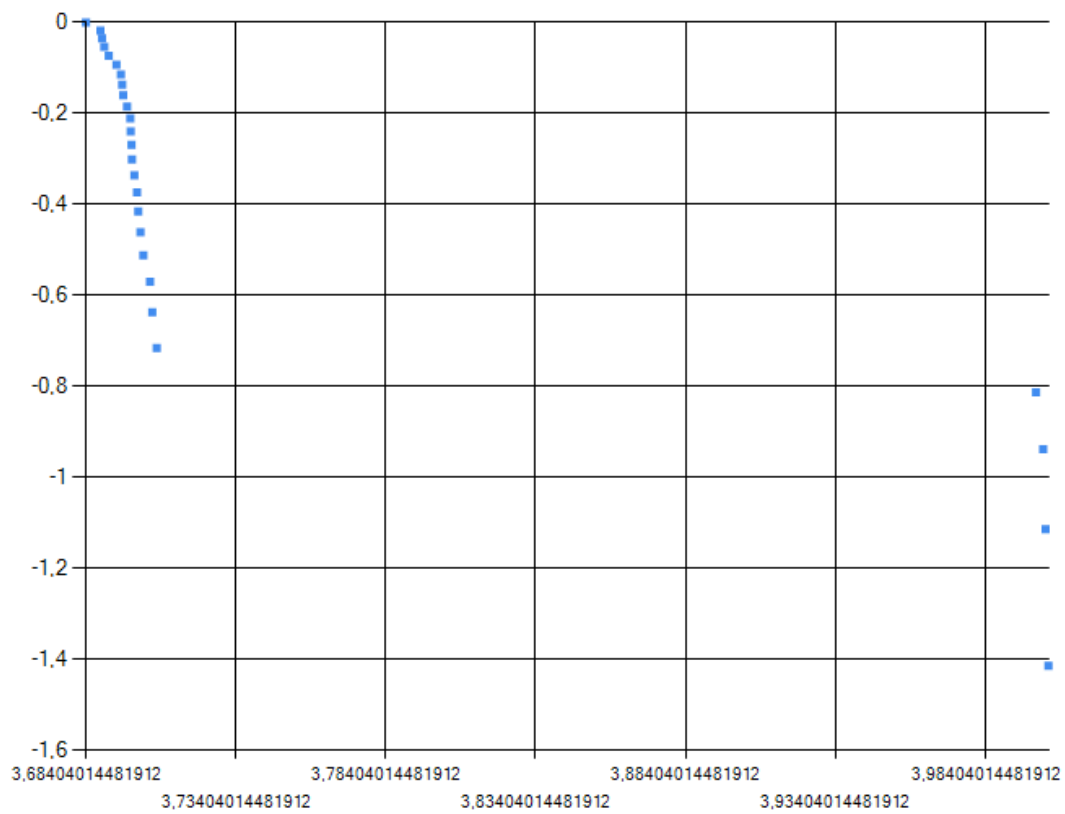




1-й закон Зіпфа

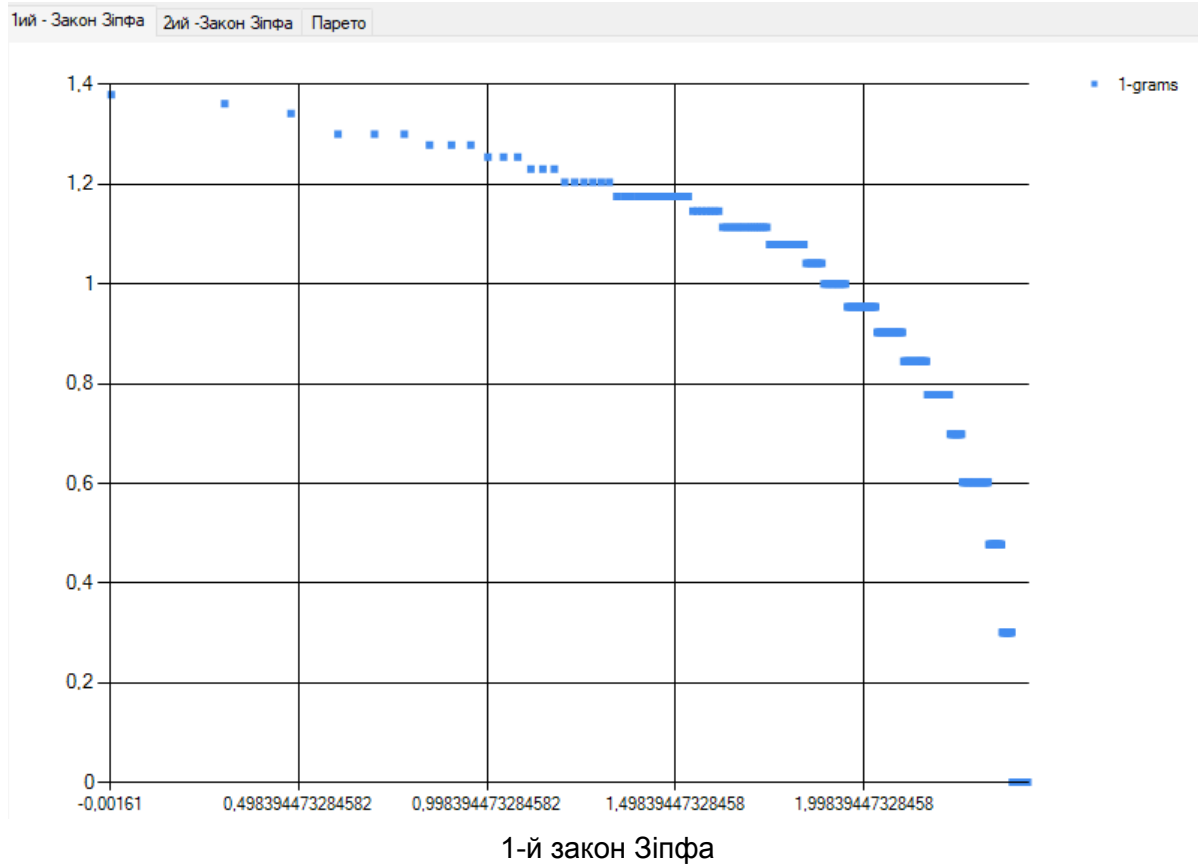


2-й закон Зіпфа

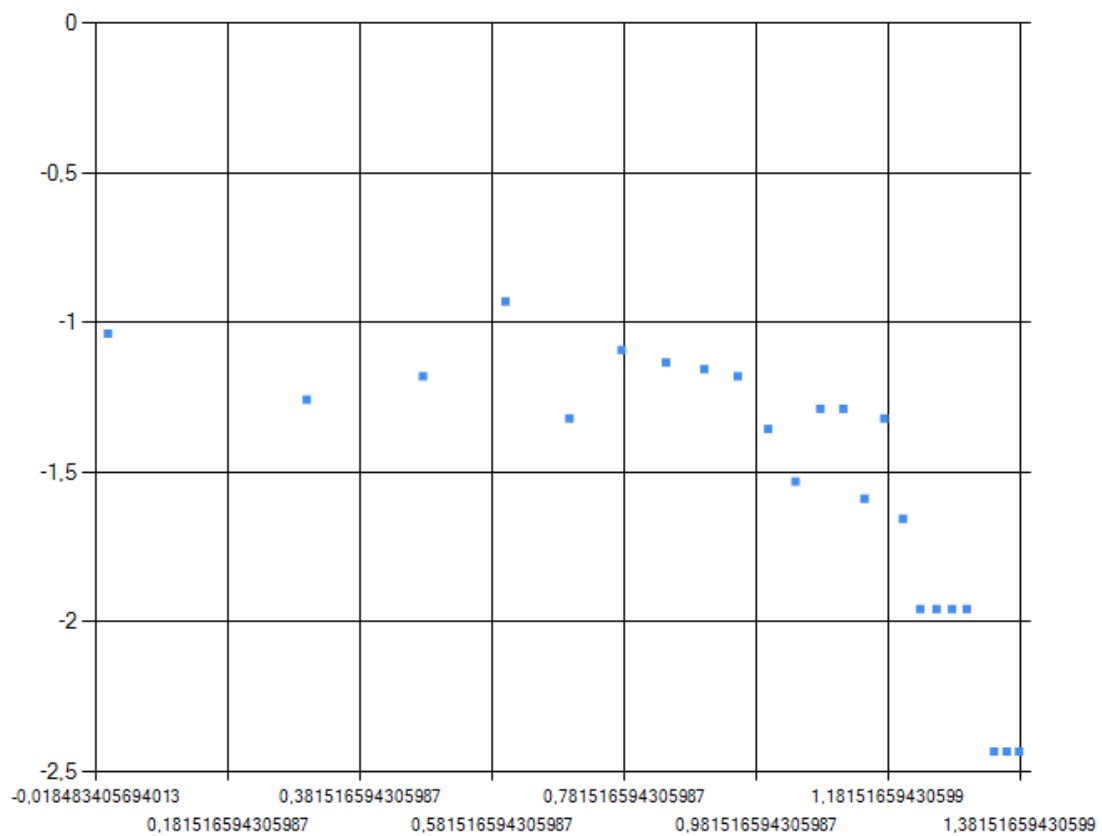


Парето

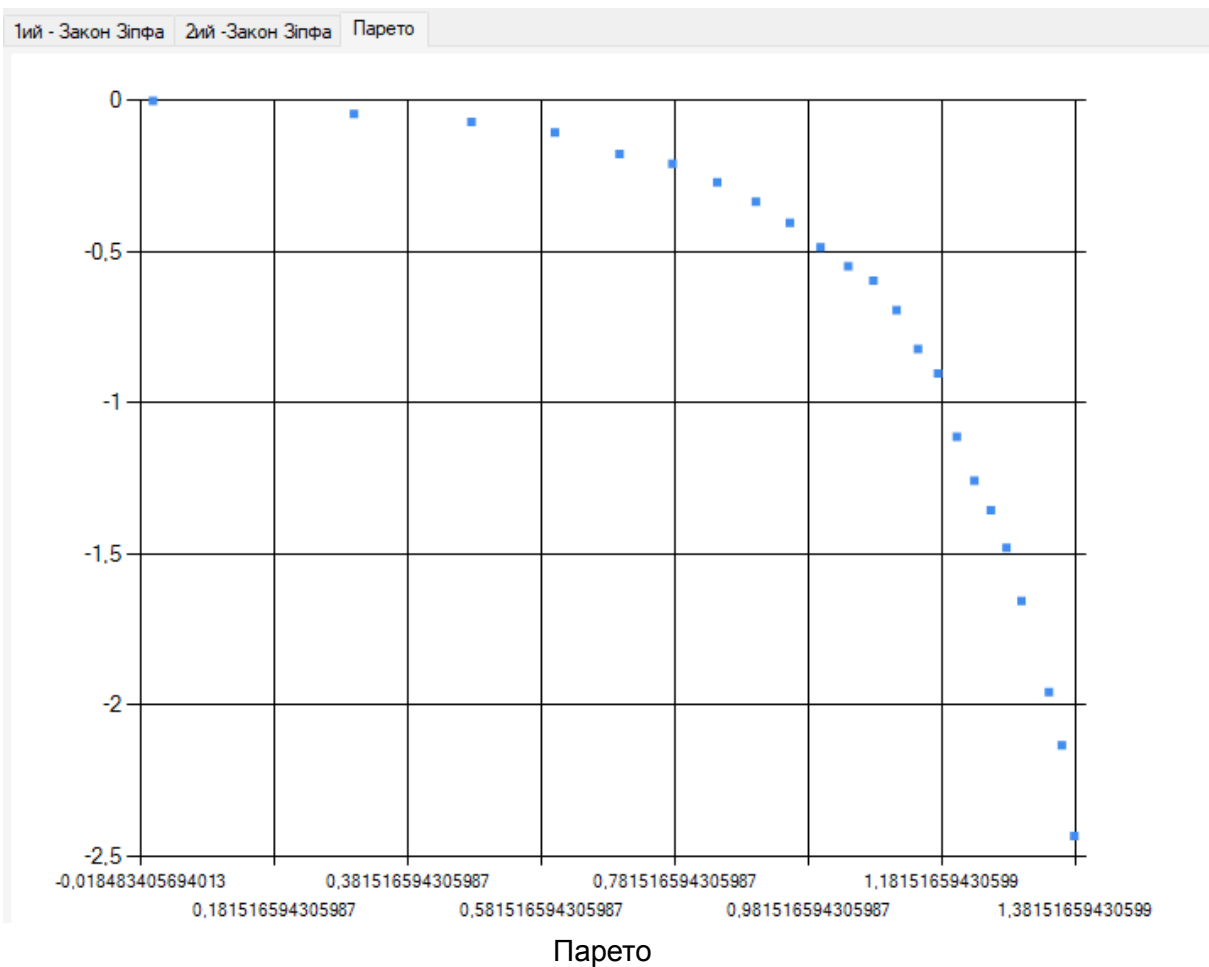
Завантажив другий текст і отримав такі результати



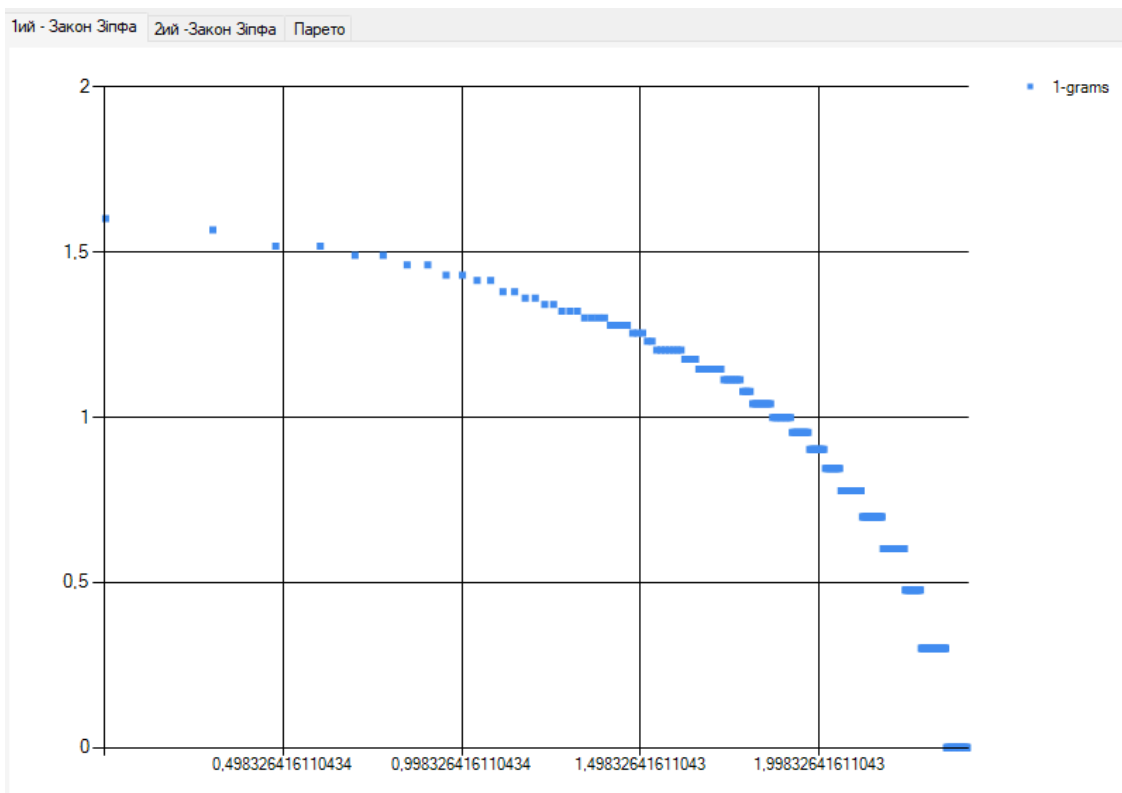
1-ий - Закон Зіпфа 2-ий - Закон Зіпфа Парето



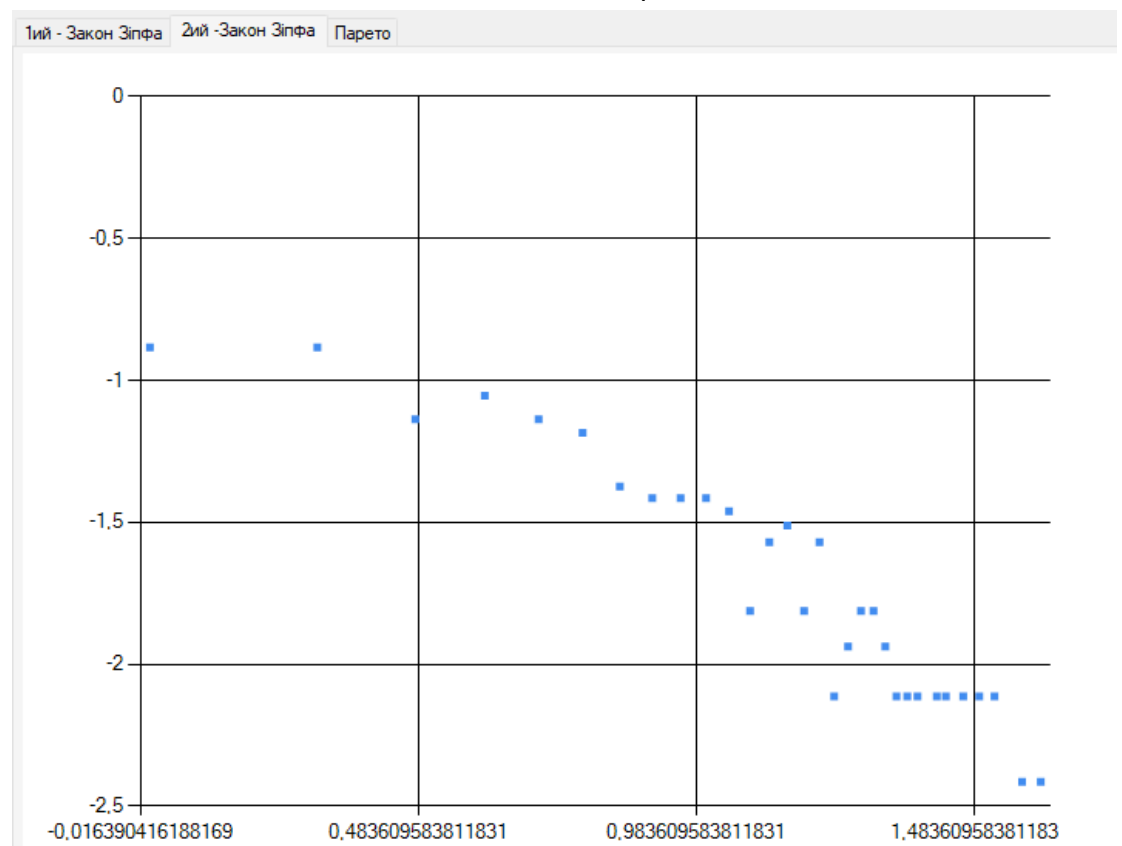
2-й закон Зіпфа



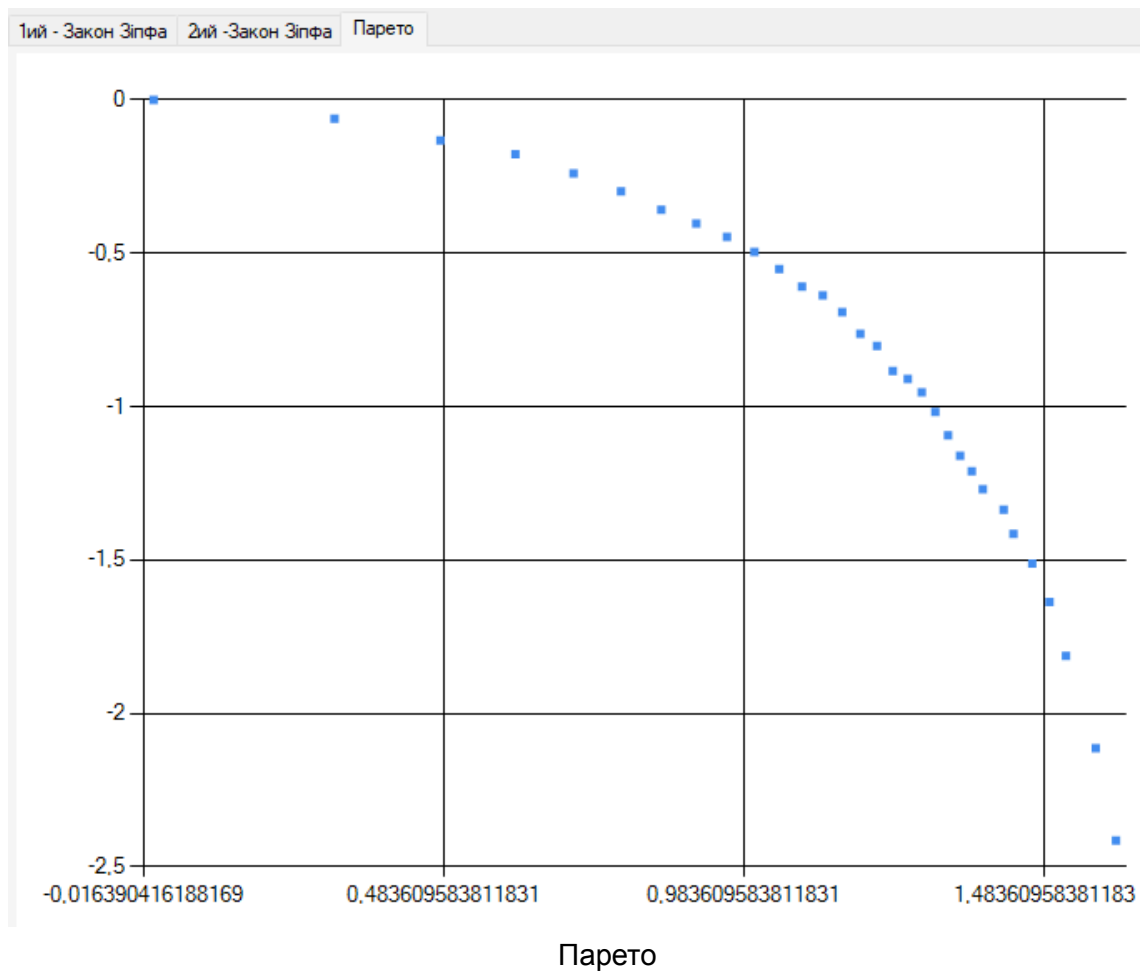
Завантажив 3-й текст, й отримав такі результати



1-й закон Зіпфа



2-й закон Зіпфа



Висновок: На цій лабораторній роботі, я дослідив застосування законів Зіпфа та Парето до згенерованих текстів і виявив певні закономірності. Перші два закони Зіпфа демонструють типовий спад частот, що є характерним для текстів, створених випадковим розподілом. Це вказує на те, що такі моделі, як текст мавпи Міллера, відображають основні лінгвістичні закономірності