

Косимов Абдунаби Абдурауфович

**О ПРИМЕНИМОСТИ γ -КЛАССИФИКАТОРА К РАСПОЗНАВАНИЮ
ОДНОРОДНОСТИ ТЕКСТОВ НА СЛАВЯНСКИХ ЯЗЫКАХ**

Терминология и понятие

- 1. Алфавит** – упорядоченное множество элементов текста.
- 2. Цифровым портретом (ЦП)** текста будем называть распределение частотности элементов алфавита.
- 3. γ -классификатор** – алгоритм, зависящий от одного вещественного параметра γ и сопоставляющий тексту из T_1 его автора из списка A .

γ -классификатор дискретных случайных величин

$$\bar{N}: \quad 1 \quad 2 \quad \dots \quad m$$

$$P: \quad p_1 \quad p_2 \quad \dots \quad p_m$$

$$\sum_{k=1}^m p_k = 1$$

$$F(s) = \sum_{k=1}^s p_k \quad (s = 1, \dots, m)$$

$$\rho(T_1, T_2) = \sqrt{\frac{m}{2}} \max_s \left| \sum_{k=1}^s (p_k^{(1)} - p_k^{(2)}) \right|$$

γ -однородными, если

$$\rho(T_1, T_2) \leq \gamma$$

$$\eta = \frac{\aleph^0 + \aleph^H}{\aleph}$$

γ -неоднородными, если

$$\rho(T_1, T_2) > \gamma$$

Униграмма	Биграмма
$m = 35$ без проб.	$m = 35^2 = 1225$ без проб.
$m = 36$ с проб.	$m = 36^2 = 1296$ с проб.

Триграмма
$m = 35^3 = 42875$ без проб.
$m = 36^3 = 46656$ с проб.

Число \aleph^0 – сумма однородных пар текстов

Число \aleph^H – сумма γ -неоднородных пар текстов

Число \aleph – общее число пар текстов в коллекции T

В монографии [Романов А.С., 2011] указано:

Известно очень много математических методов, применяемых для изучения проблемы. Из них выделяются два метода:

- машин опорных векторов;
- нейронные сети.

Имеется 1000 количественных признаков для характеристики текста. Информативными признаками считаются буквенные **униграммы, биграммы, триграммы.**

КОЛИЧЕСТВЕННЫЕ ПРИЗНАКИ ТЕКСТА

Для распознавания авторского стиля или же идентификации автора важнейшая роль отводится выбору количественных характеристик текста.

- частотности букв (униграмма, биграмма, триграмма), слов (униграмма), слог, длин предложений (в словах);

Цумла: Соли 2017 – соли ҷавонон аст.

Униграмма:

№	Униграмма	Частота	Униграмма	Частота
1	с	3	с	3
2	о	4	о	4
3	л	2	л	2
4	и	2	и	2
5	ҷ	1	ҷ	1
6	а	2	а	2
7	в	1	в	1
8	н	2	н	2
9	т	1	т	1
10	Сумма	18	Фосила (Пробел)	3
11			Сумма	21

Цумла: Соли 2017 – соли ҷавонон аст.

Биграмма:

№	Биграмма	Частота	Биграмма (с пробелами)	Частота
1	со	2	со	2
2	ол	2	ол	2
3	ли	2	ли	2
4	ис	1	и*	2
5	иҷ	1	*с	1
6	ча	1	*ҷ	1
7	ав	1	ҷа	1
8	во	1	ав	1
9	он	2	во	1
10	но	1	он	2
11	на	1	но	1
12	ас	1	н*	1
13	ст	1	*а	1
16	Сумма	17	ас	1
17			ст	1
18			Сумма	20

Цумла: Соли 2017 – соли ҷавонон аст.

Триграмма:

№	Триграмма	Частота	Триграмма (с пробелами)	Частота
1	сол	2	сол	2
2	оли	2	оли	2
3	лис	1	ли*	2
4	исо	1	и*с	1
5	лиҷ	1	*со	1
6	иҷа	1	и*ҷ	1
7	ҷав	1	*ҷа	1
8	аво	1	ҷав	1
9	вон	1	аво	1
10	оно	1	вон	1
11	нон	1	оно	1
12	она	1	нон	1
13	нас	1	он*	1
14	аст	1	н*а	1
15	Сумма	16	*ас	1
16			аст	1
17			Сумма	19

ИНФОРМАЦИЯ О КОЛЛЕКЦИИ ТЕКСТОВ

Автор	Название произведения
А.Фирдавси	Бежанва Манижа
	Рустамва Сӯҳроб
Ҷ.Румӣ	Маснавии Маънавӣ (Дафтари 1)
	Маснавии Маънавӣ (Дафтари 2)
М.Турсунзода	Ҳасани аробакаш
	Садои Осиё
Л.Шерали	Катибаҳо
	Суханреза
С.Айнӣ	Одина
	Аҳмади Девбанд
С.Турсун	Нисфирузӣ
	Повести Камони Рустам

Эффективность идентификации авторства текста (в %) на основе
символьных N -грамм (с учётом и без учёта пробелов)

	Униграмма		Биграмма		Триграмма	
	без проб.	с проб.	без проб.	с проб.	без проб.	с проб.
по критерию Н.В.Смирнова	92	92	84	92	93	100
по модификатору З.Д.Усманова	96	96	92	96	93	100

Триграмм с учетом пробела

Авторы и произведения	Фирдауси Беж. & Ман.	Фирдауси Рустам ва Сўхроб	Руми Дафтари Аввал	Руми Дафтари Дуввум	Турсунзода Садои Осиё	Турсунзода Ҳасани Аробакаш	Шерали Катибаҳо	Шерали Суханреза	Айни Аҳмади Девбанд	Айни Одина
Фирдауси Беж. & Ман.										
Фирдауси Рустам ва Сўхроб	1.0316									
Руми Дафтари Аввал	5.6331	5.3078								
Руми Дафтари Дуввум	5.0574	4.7151	1.6985							
Турсунзода Садои Осиё	6.0389	5.5587	7.8240	8.0036						
Турсунзода Ҳасани Аробакаш	5.5360	5.0589	6.9183	7.1107	0.5104					
Шерали Катибаҳо	4.3029	3.8298	4.8697	4.9889	2.2281	2.2627				
Шерали Суханреза	4.0401	3.6548	3.7958	3.9066	2.4401	1.9825	0.8310			
Айни Аҳмади Девбанд	7.1706	7.5537	11.1384	11.1310	3.3558	3.0273	3.2874	2.5212		
Айни Одина	7.6754	7.1551	8.8975	8.9706	3.6677	3.3242	3.5398	2.7481	1.3126	

$$\gamma = 1.95$$

- если $\rho(T_1, T_2) \leq \gamma = 1,95$, то тексты T_1 и T_2 **однородны**;
- если $\rho(T_1, T_2) > \gamma = 1,95$, то тексты T_1 и T_2 **не однородны**;

Значения γ^{OPT} для триграмм с учетом пробела

Авторы и произведения	Гянджеви Лайли ва Маҷнун	Гянджеви Хусрав ва Ширин	Суруш Дафтари 1	Суруш Дафтари 2	Фарзона Саду як Ғазал	Фарзона Мӯҳри гули мино	Турсун Нисфирӯзӣ	Турсун Камони Рустам
Гянджеви Лайли ва Маҷнун								
Гянджеви Хусрав ва Ширин	1.4945							
Суруш Дафтари 1	3.6469	4.0589						
Суруш Дафтари 2	3.8347	3.8211	1.0766					
Фарзона Саду як Ғазал	2.5699	3.9762	3.6963	2.8242				
Фарзона Мӯҳри гули мино	2.6818	3.6682	3.2698	2.3969	1.0436			
Турсун Нисфирӯзӣ	4.6006	4.3883	4.7019	5.0047	6.1921	5.6613		
Турсун Камони Рустам	4.7526	4.9378	4.8991	5.2433	6.3091	5.8772	1.7066	

$$\gamma = 1.95$$

ОПРЕДЕЛЕНИЕ МИНИМАЛЬНОГО РАЗМЕРА ТЕКСТА, НЕОБХОДИМОГО ДЛЯ РАСПОЗНАВАНИЕ ЕГО АВТОРА

В монографии [Романов А.С., 2011] указывается,

что для успешной идентификации текста необходимо

10000 слов для английского языка и **8000** слов для русского языка.

ОПРЕДЕЛЕНИЕ МИНИМАЛЬНОГО РАЗМЕРА ТЕКСТА, НЕОБХОДИМОГО ДЛЯ РАСПОЗНАВАНИЯ ЕГО АВТОРА

Показатель распознавания авторства (в относительных единицах)
в зависимости от длины текстов (в словах)

Длина текста (в словах)	Н.В.Смирнов	З.Д.Усманов
20000	1	1
10000	1	1
5000	1	1
4000	1	1
2500	0.94	0.94
1250	0.73	0.95
800	0.59	0.91
625	0.39	0.92
300	0.33	0.88
150	0.32	0.82
75	0.09	0.58

СРАВНЕНИЕ РЕЗУЛЬТАТОВ

Эффективность классификации в зависимости от длины текстов

Число слов	Эффективность η и значение γ^{opt}					
	Униграмма		Биграмма		Триграмма	
	без проб.	с проб.	без проб.	с проб.	без проб.	с проб.
10000	1.00 (0.06)	1.00 (0.06)	1.00 (0.40)	1.00 (0.40)	1.00 (2.20)	1.00 (1.95)
5000	1.00 (0.06)	1.00 (0.05)	1.00 (0.40)	1.00 (0.40)	1.00 (2.20)	1.00 (1.95)
4000	0.98 (0.06)	1.00 (0.05)	1.00 (0.40)	1.00 (0.40)	1.00 (2.30)	1.00 (1.95)
2500	0.98 (0.08)	0.98 (0.06)	0.98 (0.40)	1.00 (0.40)	0.98 (2.30)	1.00 (2.26)
1250	0.98 (0.08)	0.98 (0.07)	0.98 (0.40)	0.98 (0.40)	0.98 (2.40)	1.00 (2.31)
800	0.93 (0.08)	0.93 (0.07)	0.93 (0.50)	0.93 (0.50)	0.93 (2.50)	0.96 (2.71)
625	0.93 (0.08)	0.93 (0.07)	0.93 (0.50)	0.93 (0.50)	0.93 (2.50)	0.96 (2.71)
300	0.67 (0.17)	0.84 (0.11)	0.71 (0.90)	0.82 (0.70)	0.82 (4.90)	0.87 (3.91)
150	0.62 (0.17)	0.73 (0.12)	0.71 (0.90)	0.82 (0.70)	0.76 (5.20)	0.84 (4.07)
75	0.49 (0.23)	0.53 (0.19)	0.44 (1.50)	0.62 (1.10)	0.51 (8.06)	0.62 (6.41)

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ ВЫСОКОЧАСТОТНЫХ ЭЛЕМЕНТОВ АЛФАВИТА СИМВОЛЬНЫХ N-ГРАММ ДЛЯ ИДЕНТИФИКАЦИИ АВТОРА ТЕКСТА

Эффективность классификации в зависимости от числа
используемых высокочастотных N -грамм

Число N -грамм и процент покрытия – эффективность η и значение γ^{opt}		
Униграмма (с проб.)	Биграмма (с проб.)	Триграмма (с проб.)
11 (75.65%) – 1.00 (0.017)	55 (56.04%) – 1.00 (0.08)	1000 (77.07%) – 1.00 (0.17)
8 (65.67%) – 1.00 (0.014)	50 (53.33%) – 0.98 (0.07)	500 (62.43%) – 0.93 (0.21)
7 (61.86%) – 1.00 (0.013)	45 (50.45%) – 0.98 (0.07)	300 (51.48%) – 0.91 (0.31)
6 (56.33%) – 1.00 (0.013)		
5 (50.74%) – 0.93 (0.012)		
3 (38.74%) – 0.84 (0.009)		

Границы оптимального полуинтервала $\gamma^{\text{опт}}$ и π^{max} для различных ЦП

ЦПТ	<i>Поэзия классическая</i>		<i>Поэзия современная</i>		<i>Проза современная</i>	
	эффект. (%)	γ	эффект. (%)	γ	эффект. (%)	γ
1gramma_bukv_n	0,98	[0,0420; 0,0502)	0,98	[0,0302; 0,0374)	0,91	[0,0316; 0,0350)
2gramma_bukv_n	1,00	[0,2471; 0,3404)	0,98	[0,2499; 0,2524)	0,91	[0,1379; 0,1692)
3gramma_bukv_n	1,00	[2,2084; 2,8871)	0,96	[1,5590; 1,7578)	0,93	[1,2747; 1,4477)
1gramma_bukv_bo_probel	0,98	[0,0354; 0,0447)	0,98	[0,0268; 0,0423)	0,96	[0,0285; 0,0336)
2gramma_bukv_bo_probel	0,98	[0,2987; 0,3551)	0,98	[0,2318; 0,2816)	0,93	[0,2216; 0,2272)
3gramma_bukv_bo_probel	1,00	[2,1630; 2,1648)	0,98	[1,3885; 1,7054)	0,96	[1,3379; 1,3412)
100_1gramma_kalima	0,96	[0,1168; 0,1408)	0,96	[0,1129; 0,1220)	0,96	[0,0559; 0,0769)
300_1gramma_kalima	0,96	[0,2505; 0,2650)	1,00	[0,2565; 0,2809)	0,93	[0,0970; 0,1332)
500_1gramma_kalima	0,96	[0,4048; 0,4179)	1,00	[0,3309; 0,3987)	0,93	[0,1286; 0,1720)
1000_1gramma_kalima	0,93	[0,4982; 0,5910)	1,00	[0,5451; 0,5965)	0,93	[0,1981; 0,4062)
20000_1gramma_kalima	0,93	[2,0064; 2,8918)	1,00	[2,6343; 2,6679)	0,93	[1,1620; 2,2459)
30000_1gramma_kalima	0,93	[2,4574; 3,5417)	1,00	[3,2263; 3,2676)	0,93	[1,4232; 2,7507)
1gr_shum_kal_jumla	0,93	[0,3652; 0,3875)	0,91	[0,6302; 1,0375)	0,96	[0,5416; 0,7949)
hijo	0,98	[0,6426; 0,7578)	0,96	[0,9456; 1,0886)	0,93	[0,4796; 0,5377)

Информация о размерах фрагментов в словах и символах

Номер фрагмента	1	2	3	4	5	6	7	8	9
Число слов	7000	3500	1700	900	450	250	130	60	20
Число символов	40000	20000	10000	5000	2500	1200	600	300	100

Расстояния между цифровыми портретами произведений из коллекции текстов и фрагментами, извлеченными из *“начала”* романа С.Айни *“Дохунда”*

Авторы (произв.)		Длины фрагментов (в символах)								
		40000	20000	10000	5000	2500	1200	600	300	100
АФ	Р&С	3,3783	4,9715	6,7872	6,9954	7,1174	7,1174	7,0964	7,0964	9,0921
	Б&М	3,4286	5,0218	6,8375	7,0457	7,1677	7,1677	6,9105	6,9105	9,0923
ЦР	ММ1	3,7737	5,3669	7,1825	7,3907	7,5127	7,5127	7,2551	6,9441	9,2609
	ММ2	3,9322	5,5254	7,3411	7,5493	7,6713	7,6713	7,3256	7,0731	9,3478
АС	Д1	1,6159	3,2091	5,0495	5,6676	5,4033	5,9929	5,9929	5,9929	8,0034
	Д2	1,6925	3,2655	5,1699	5,7881	5,5237	6,1133	6,1133	6,1133	8,0427
СТ	Н	2,2554	3,8487	5,7194	6,3374	6,0731	6,6627	6,6627	6,6627	8,6085
	ПКР	2,1139	3,6101	5,5036	6,1217	5,8573	6,4471	6,4471	6,4471	8,5379
СА	О	1,8822	0,9487	1,8043	2,4224	2,1581	2,7477	2,9876	3,3602	5,6145
	АД	1,0943	2,0211	3,1963	3,5291	3,5265	3,8711	4,3428	3,8543	6,9466
	Д	1,0951	0,9983	2,6248	3,2429	2,9785	3,5682	3,5682	3,5682	5,9816
	МС	1,1901	1,3745	2,6615	3,2386	2,9917	3,5639	3,8691	3,5639	6,3826

Расстояния между цифровыми портретами произведений из коллекции текстов и фрагментами, извлеченными из “середины” романа С.Айни “Дохунда”

Авторы (произв.)		Длины фрагментов (в символах)								
		40000	20000	10000	5000	2500	1200	600	300	100
АФ	Р&С	4,0421	3,9059	4,8551	5,2593	5,3341	4,8101	4,4169	7,9547	7,9547
	Б&М	4,0924	3,9563	4,8591	5,2633	5,3382	4,8141	4,4211	7,9587	7,9587
ЦР	ММ1	4,4374	4,3013	5,0532	5,4575	5,5473	4,9459	4,5529	8,0906	8,0906
	ММ2	4,5961	4,4599	5,2117	5,6161	5,7059	5,1421	4,7491	8,2867	8,2867
АС	Д1	2,4126	2,2341	3,2282	3,6326	3,7074	3,1833	2,7902	6,3281	6,3281
	Д2	2,5229	2,3443	3,3385	3,7428	3,8177	3,2936	2,9005	6,4383	6,4383
СТ	Н	2,9931	2,8144	3,8087	4,2131	4,2879	3,7638	3,3707	6,9084	6,9084
	ПКР	2,8855	2,5889	3,4821	3,8511	3,9261	3,4019	3,0088	6,5465	6,5465
СА	О	1,0901	1,3311	0,8312	1,2726	1,1678	2,5293	3,3155	3,8195	3,8195
	АД	1,0011	0,9379	1,4863	2,3843	2,2794	1,2381	2,4173	4,2576	4,2576
	Д	0,2542	0,5841	0,7267	1,3751	1,2702	1,7772	2,5633	3,8264	3,8264
	МС	0,4261	0,6716	0,8598	1,7709	1,6661	1,5033	2,2895	3,8991	3,8991

Расстояния между цифровыми портретами произведений из коллекции текстов и фрагментами из “окончания” романа С.Айни “Дохунда”

Авторы (произв.)		Длины фрагментов (в символах)								
		40000	20000	10000	5000	2500	1200	600	300	100
АФ	Р&С	4,7589	4,5362	4,7926	4,5627	4,4791	4,1811	4,3751	5,5892	7,9477
	Б&М	4,8092	4,6254	4,8789	4,6932	4,5689	4,1851	4,3753	5,5071	7,8655
ЦР	ММ1	5,1811	5,0052	5,2587	5,0729	5,0141	4,3171	4,5439	4,7694	7,1279
	ММ2	5,3187	5,1428	5,3963	5,2106	5,1147	4,5131	4,6308	4,7171	7,0755
АС	Д1	3,0187	2,8223	3,0769	2,9433	3,0137	2,5544	3,2864	2,0492	4,4077
	Д2	3,1077	2,9318	3,1853	3,0041	3,1341	2,6647	3,3258	2,2036	4,2854
СТ	Н	3,6932	3,5173	3,7708	3,6222	3,6836	3,1348	3,8915	1,5331	3,9898
	ПКР	3,5259	3,3573	3,6278	3,4129	3,5202	2,9746	3,8209	1,7486	4,1071
СА	О	0,5874	1,0162	1,0039	0,7948	0,9653	1,7235	2,6669	5,0254	7,3839
	АД	1,6258	1,4884	1,5808	1,3033	1,9484	1,7994	2,2296	4,2521	6,6105
	Д	0,7024	0,5651	0,6566	0,7606	0,9338	0,9594	1,9028	4,2613	6,6198
	МС	1,1033	0,9658	1,0388	0,7064	1,3155	1,1665	2,0086	4,3671	6,7256

Расстояния между цифровыми портретами произведений из коллекции текстов и фрагментами, извлеченными из “*начала*” поэмы А.Фирдауси “Рустам ва Сӯҳроб”

Авторы (произв.)		Длины фрагментов (в символах)								
		40000	20000	10000	5000	2500	1200	600	300	100
АФ	Р&С	0,2446	0,2731	0,3562	0,6964	1,0645	1,5623	1,6398	2,5832	2,3376
	Б&М	0,4305	0,3892	0,4319	0,6153	0,8806	1,7033	1,4559	2,3993	2,5235
ЦР	ММ1	0,9901	1,0436	0,8228	0,7992	0,8037	1,7378	1,1974	2,1408	2,4899
	ММ2	1,1172	1,1709	0,9501	0,9614	0,9311	1,6598	1,2309	2,1128	2,3611
АС	Д1	4,0087	4,0623	3,8414	3,8825	3,8224	4,2462	4,1241	4,1241	5,0263
	Д2	4,0895	4,1431	3,9223	3,9395	3,9032	4,3271	4,2854	4,2854	5,1791
СТ	Н	4,5136	4,5672	4,3463	4,1863	4,3273	4,7511	4,3043	3,9898	5,6702
	ПКР	4,2691	4,3228	4,1019	3,9419	4,0829	4,5067	4,0598	3,9578	5,3268
СА	О	5,2231	5,3725	5,4883	5,7978	6,0394	5,7729	6,2695	6,2695	6,6863
	АД	3,8487	3,9701	3,9471	4,2565	4,4981	4,2317	4,7282	4,7282	5,5797
	Д	4,3117	4,4612	4,5771	4,8865	5,1281	4,8617	5,3582	5,3582	5,8658
	МС	4,2165	4,3661	4,4818	4,7913	5,0329	4,7665	5,2631	5,2631	5,8701

Расстояния между цифровыми портретами произведений из коллекции текстов и фрагментами, извлеченными из “*середины*” поэмы А.Фирдауси “Рустам ва Сӯҳроб”

Авторы (произв.)		Длины фрагментов (в символах)								
		40000	20000	10000	5000	2500	1200	600	300	100
АФ	Р&С	0,3255	0,5782	0,7907	0,6964	1,1081	1,8881	1,2893	2,3376	3,0772
	Б&М	0,3047	0,5357	0,6607	0,8725	1,0259	1,7041	1,4753	2,5235	3,2611
ЦР	ММ1	0,9739	0,8091	0,7627	0,8391	0,6793	1,4456	1,7241	2,4899	3,5196
	ММ2	1,1011	0,9363	0,8899	0,7901	0,7263	1,4177	1,8513	2,3611	3,5476
АС	Д1	3,9925	3,8278	3,7813	3,6814	3,7762	3,7762	4,7428	4,7428	5,4129
	Д2	4,0734	3,9086	3,8622	3,8527	3,9706	3,9706	4,8236	4,8236	5,5345
СТ	Н	4,4974	4,3327	4,2862	4,1863	3,8326	3,7581	5,2477	5,2477	6,0731
	ПКР	4,2531	4,0882	4,0418	3,9419	3,5881	3,5137	5,0032	5,0032	5,7501
СА	О	4,9925	5,1048	5,3831	5,4634	5,5813	5,1967	5,6381	6,6863	6,6863
	АД	3,6281	3,7565	4,0347	4,1151	4,2331	4,0901	4,5315	5,5797	5,5797
	Д	4,0812	4,1877	4,4659	4,5462	4,6642	4,3762	4,8176	5,8658	5,8658
	МС	3,9866	4,1152	4,3933	4,4737	4,5916	4,3805	4,8219	5,8701	5,8701

Расстояния между цифровыми портретами произведений из коллекции текстов и фрагментами, извлеченными из “окончания” поэмы А.Фирдауси “Рустам ва Сүхроб”

Авторы (произв.)		Длины фрагментов (в символах)								
		40000	20000	10000	5000	2500	1200	600	300	100
АФ	Р&С	0,1697	0,3001	0,6383	0,6929	0,7327	2,5832	0,8071	2,6795	8,6943
	Б&М	0,3536	0,4833	0,8242	0,6367	0,8725	2,3993	0,6911	2,8205	8,7274
ЦР	ММ1	1,0127	0,7868	0,7906	1,1345	1,7241	2,1408	1,7241	2,8551	9,1419
	ММ2	1,1399	0,9141	0,8552	1,2617	1,8513	2,1128	1,8513	2,7771	9,2124
АС	Д1	4,0314	3,8054	3,7466	4,1531	4,7428	4,7428	4,7428	4,7428	7,6941
	Д2	4,1122	3,8863	3,8275	4,2341	4,8236	4,8236	4,8236	4,8236	7,6671
СТ	Н	4,5362	4,3103	4,2515	4,6581	5,2477	5,2477	5,2477	5,2477	8,1859
	ПКР	4,2918	4,0659	4,0071	4,4136	5,0032	5,0032	5,0032	5,0032	8,2392
СА	О	5,2333	5,3321	5,5663	5,6861	5,5681	5,7978	5,2213	4,3129	4,8744
	АД	3,7071	3,8157	4,1216	4,1727	4,0548	4,2565	3,6801	2,9355	6,2296
	Д	4,3221	4,4208	4,6551	4,7847	4,6668	4,8865	4,3101	3,8711	5,3493
	МС	4,2268	4,3257	4,5598	4,6412	4,5555	4,7913	4,2148	3,4692	5,7558

О РАСПОЗНАВАНИИ АВТОРА ТЕКСТА НА УЗБЕКСКОМ ЯЗЫКЕ С ПОМОЩЬЮ СИМВОЛЬНЫХ N-ГРАММ

Модельная коллекция текстов:

- **А.Навои** “Лайли ва Мажнун” (АН, Л&М, 162 Кб) и “Фарход ва Ширин” (АН, Ф&Ш, 340 Кб);
- двумя произведениями **З.М.Бобура** “Ғазалиёт” (БЗМ, Ғ, 35Кб) и “Маҳрами асрор топмадим” (БЗМ, МТ, 120Кб);
- текстами **А.Кодирий** “Меҳробдан чаён” (АК, М, 801Кб) и “Ўткан кунлар” (АК, Ў, 1215Кб); прозой С.Айни “Судхўрнинг ўлими” (СА, С, 406Кб) и “Эсдаликлар” (СА, Э, 1131Кб).

О РАСПОЗНАВАНИИ АВТОРА ТЕКСТА НА УЗБЕКСКОМ ЯЗЫКЕ С ПОМОЩЬЮ СИМВОЛЬНЫХ N-ГРАММ

Границы оптимального полуинтервала $\gamma^{\text{опт}}$ и π^{max} для различных ЦП

№	ЦПТ	γ	π (%)
1	Униграмма	[0,0612; 0,0794)	100
2	Биграмма	[0,3608; 0,4711)	100
3	Триграмма	[2,1352; 2,8311)	100
4	Униграмма с уч. проб.	[0,0539; 0,0643)	100
5	Биграмма с уч. проб.	[0,3227; 0,3920)	100
6	Триграмма с уч. проб.	[1,9436; 2,3490)	100

Спасибо за внимание!!!