

1. Przygotowano 4 pliki zawierające tekst:
  - a. literacki w języku polskim,
  - b. informatyczny w języku polskim,
  - c. literacki w języku angielskim,
  - d. informatyczny w języku angielskim.

W Tabeli 1. zamieszczono szczegółowe informacje na temat wybranych tekstów.

*Tabela 1. Szczegółowe informacje o wybranej literaturze*

L.p.	Autor	Tytuł	Rodzaj tekstu	Język	Liczba znaków
1.	H. Sienkiewicz	Potop	literacki	polski	10085
2.	J.Albahari, B.Albahari	C# 6.0 w pigułce	informatyczny	polski	12934
3.	J.K.Rowling	Harry Potter and the Philosopher's Stone	literacki	angielski	10016
4.	Andrew M. St. Laurent	Understanding Open Source and Free Software Licensing	informatyczny	angielski	10486

2. Wykonano program dla 4 przygotowanych plików. Pliki badano pod względem:
  - a. rodzaju występujących znaków,
  - b. częstości występowania znaków,
  - c. entropii binarnej,
  - d. ilości informacji.

Ilość informacji wyznaczono ze wzoru:

$$I(E) = \log_2 \frac{1}{P(E)} \text{ bitów},$$

gdzie  $E$  to pewne zdarzenie, które zachodzi z prawdopodobieństwem  $P(E)$ .

Entropię binarną wyznaczono ze wzoru:

$$H(S) = \sum_S (P(S_i) I(S_i)) \text{ bitów},$$

gdzie  $S$  - źródło, to skończony zbiór wiadomości elementarnych  $S = \{S_1, S_2, \dots, S_q\}$

W Tabeli 2. zamieszczono wyniki, które uzyskano po zbadaniu tekstów. Należy wziąć pod uwagę, że liczba znaków w *Tekst n.* była różna.

*Tabela 2. Zestawienie badanych elementów tekstów*

L.p.	Znak	Liczba wystąpień				Prawdopodobieństwo				Ilość informacji [bitów]			
		Tekst 1.	Tekst 2.	Tekst 3.	Tekst 4.	Tekst 1.	Tekst 2.	Tekst 3.	Tekst 4.	Tekst 1.	Tekst 2.	Tekst 3.	Tekst 4.
1.	a	846	693	691	563	0,070783	0,05358	0,056408	0,053691	3,820453	4,222162	4,147956	4,219176
2.	ą	111	0	0	0	0,009287	0	0	0	6,750572	0	0	0
3.	b	124	139	135	104	0,010375	0,010747	0,01102	0,009918	6,590745	6,539922	6,503732	6,655735
4.	c	429	363	180	312	0,035894	0,028066	0,014694	0,029754	4,800113	5,155033	6,088629	5,070773

5.	ć	41	0	0	0	0,00343	0	0	0	8,187576	0	0	0
6.	d	297	268	449	270	0,024849	0,020721	0,036653	0,025749	5,330668	5,592763	4,769925	5,27934
7.	e	694	1031	1181	994	0,058066	0,079712	0,096408	0,094793	4,106163	3,649059	3,374703	3,399076
8.	ę	129	0	0	0	0,010793	0	0	0	6,53376	0	0	0
9.	f	12	81	152	198	0,001004	0,006263	0,012408	0,018882	9,960025	7,31893	6,332586	5,726845
10.	g	127	90	219	177	0,010626	0,006958	0,017878	0,01688	6,556258	7,167112	5,805671	5,888541
11.	h	133	57	606	422	0,011128	0,004407	0,049469	0,040244	6,489662	7,825987	4,337331	4,635082
12.	i	906	724	534	631	0,075803	0,055976	0,043592	0,060175	3,721601	4,159048	4,519793	4,054692
13.	j	198	263	10	7	0,016566	0,020334	0,000816	0,000668	5,915631	5,619962	10,259143	10,547864
14.	k	313	221	78	70	0,026188	0,017087	0,006367	0,006676	5,25495	5,870957	7,295171	7,2268
15.	l	266	347	410	294	0,022256	0,026829	0,033469	0,028037	5,489662	5,220063	4,901031	5,156524
16.	ł	251	42	0	0	0,021001	0,003247	0	0	5,573398	8,266677	0	0
17.	m	274	271	170	161	0,022925	0,020953	0,013878	0,015354	5,446934	5,576699	6,171057	6,025242
18.	n	482	724	609	486	0,040328	0,055976	0,049714	0,046348	4,632074	4,159048	4,330204	4,431349
19.	o	709	689	637	759	0,059321	0,05327	0,052	0,072382	4,075313	4,230533	4,265345	3,788225
20.	ó	87	0	0	0	0,007279	0	0	0	7,102044	0	0	0
21.	q	0	0	5	5	0	0	0,000408	0,000477	0	0	11,259143	11,033723
22.	p	236	233	157	244	0,019746	0,018015	0,012816	0,023269	5,662296	5,794658	6,28591	5,425447
23.	r	316	519	587	672	0,026439	0,040127	0,047918	0,064085	5,241189	4,639283	4,383288	3,963869
24.	s	415	359	614	469	0,034722	0,027756	0,050122	0,044726	4,848006	5,171057	4,318412	4,482742
25.	ś	66	0	0	18	0,005522	0	0	0,001717	7,500593	0	0	9,185894
26.	t	286	632	835	823	0,023929	0,048863	0,068163	0,078486	5,385096	4,355114	3,874867	3,671421
27.	u	211	273	299	162	0,017654	0,021107	0,024408	0,015449	5,823861	5,566135	5,356502	6,016343
28.	w	460	289	226	132	0,038487	0,022344	0,018449	0,012588	4,699485	5,483969	5,760314	6,311807
29.	x	0	31	14	39	0	0,002397	0,001143	0,003719	0	8,704554	9,772959	8,07087
30.	y	345	289	240	164	0,028865	0,022344	0,019592	0,01564	5,114535	5,483969	5,673592	5,998616
31.	z	600	336	5	4	0,050201	0,025978	0,000408	0,000381	4,31614	5,266566	11,259143	11,357921
32.	ż	113	0	0	0	0,009454	0	0	0	6,724859	0	0	0
33.	ź	8	0	0	0	0,000669	0	0	0	10,545706	0	0	0
34.	A	10	17	14	12	0,000837	0,001314	0,001143	0,001144	10,222485	9,571819	9,772959	9,771697
35.	Ą	0	0	0	0	0	0	0	0	0	0	0	0
36.	B	31	1	6	5	0,002594	0,000077	0,00049	0,000477	8,590606	13,664782	10,994931	11,033723
37.	C	5	48	3	11	0,000418	0,003711	0,000245	0,001049	11,224209	8,073976	11,994931	9,89677
38.	Ć	0	0	0	0	0	0	0	0	0	0	0	0
39.	D	7	13	69	0	0,000586	0,001005	0,005633	0	10,736812	9,958589	7,471881	0
40.	E	0	75	7	2	0	0,005799	0,000571	0,000191	0	7,42998	10,774222	12,35414
41.	Ę	0	0	0	0	0	0	0	0	0	0	0	0

42.	F	2	3	3	7	0,000167	0,000232	0,000245	0,000668	12,547864	12,073588	11,994931	10,547864
43.	G	11	19	4	0	0,00092	0,001469	0,000327	0	10,086079	9,41095	11,578422	0
44.	H	10	1	45	3	0,000837	0,000077	0,003673	0,000286	10,222485	13,664782	8,088825	11,771697
45.	I	1	141	25	6	0,000084	0,010901	0,002041	0,000572	13,539251	6,519396	8,936508	10,771697
46.	J	5	5	2	0	0,000418	0,000387	0,000163	0	11,224209	11,335379	12,58284	0
47.	K	15	8	2	0	0,001255	0,000619	0,000163	0	9,638097	10,657773	12,58284	0
48.	L	14	36	2	4	0,001171	0,002783	0,000163	0,000381	9,738043	8,489143	12,58284	11,357921
49.	Ł	0	0	0	0	0	0	0	0	0	0	0	0
50.	M	12	16	55	3	0,001004	0,001237	0,00449	0,000286	9,960025	9,658939	7,799069	11,771697
51.	N	4	18	7	2	0,000335	0,001392	0,000571	0,000191	11,543551	9,488625	10,774222	12,35414
52.	O	8	16	8	6	0,000669	0,001237	0,000653	0,000572	10,545706	9,658939	10,580629	10,771697
53.	Ó	0	0	0	0	0	0	0	0	0	0	0	0
54.	Q	0	3	0	0	0	0,000232	0	0	0	12,073588	0	0
55.	P	23	5	26	3	0,001924	0,000387	0,002122	0,000286	9,021675	11,335379	8,88036	11,771697
56.	R	19	14	2	7	0,00159	0,001082	0,000163	0,000668	9,296758	9,852084	12,58284	10,547864
57.	S	15	9	7	8	0,001255	0,000696	0,000571	0,000763	9,638097	10,488625	10,774222	10,356029
58.	Ś	0	0	0	0	0	0	0	0	0	0	0	0
59.	T	9	59	26	21	0,000753	0,004562	0,002122	0,002003	10,375063	7,776118	8,88036	8,963622
60.	U	2	0	0	7	0,000167	0	0	0,000668	12,547864	0	0	10,547864
61.	W	24	8	13	6	0,002008	0,000619	0,001061	0,000572	8,960025	10,657773	9,88036	10,771697
62.	X	0	0	0	0	0	0	0	0	0	0	0	0
63.	Y	0	0	4	0	0	0	0,000327	0	0	0	11,578422	0
64.	Z	5	10	0	0	0,000418	0,000773	0	0	11,224209	10,337244	0	0
65.	Ż	0	1	0	0	0	0,000077	0	0	0	13,664782	0	0
66.	ź	5	0	0	0	0,000418	0	0	0	11,224209	0	0	0
67.		1867	1345	2234	1516	0,156208	0,103989	0,182367	0,144574	2,67846	3,265497	2,455083	2,79012
68.	.	73	90	224	64	0,006108	0,006958	0,018286	0,006103	7,355084	7,167112	5,773117	7,356266
69.	,	211	81	127	120	0,017654	0,006263	0,010367	0,011444	5,823861	7,31893	6,591858	6,449265
70.	?	0	0	14	0	0	0	0,001143	0	0	0	9,772959	0
71.	!	0	3	10	0	0	0,000232	0,000816	0	0	12,073588	10,259143	0
72.	(	4	65	1	6	0,000335	0,005026	0,000082	0,000572	11,543551	7,636374	13,574017	10,771697
73.	)	4	65	1	6	0,000335	0,005026	0,000082	0,000572	11,543551	7,636374	13,574017	10,771697
74.	-	8	7	41	6	0,000669	0,000541	0,003347	0,000572	10,545706	10,852084	8,222916	10,771697
75.	;	11	61	7	0	0,00092	0,004716	0,000571	0	10,086079	7,728221	10,774222	0
76.	:	4	23	1	4	0,000335	0,001778	0,000082	0,000381	11,543551	9,135529	13,574017	11,357921
77.	[	0	9	0	0	0	0,000696	0	0	0	10,488625	0	0
78.	]	0	9	0	0	0	0,000696	0	0	0	10,488625	0	0

79.	{	0	38	0	0	0	0,002938	0	0	0	8,41095	0	0
80.	}	0	38	0	0	0	0,002938	0	0	0	8,41095	0	0
81.		0	0	61	0	0	0	0,00498	0	0	0	7,649639	0
82.	`	0	0	0	0	0	0	0	0	0	0	0	0
83.	%	0	0	0	0	0	0	0	0	0	0	0	0
84.	&	0	0	0	0	0	0	0	0	0	0	0	0
85.		3	4	48	0	0,000251	0,000309	0,003918	0	11,960025	11,660106	7,995667	0

*Tabela 3.* prezentuje najwyższe wyniki osiągnięte w *Tekst n.* pod względem liczby wystąpień i prawdopodobieństwa danego znaku.

*Tabela 3. Top 5 - Liczba wystąpień, prawdopodobieństwo*

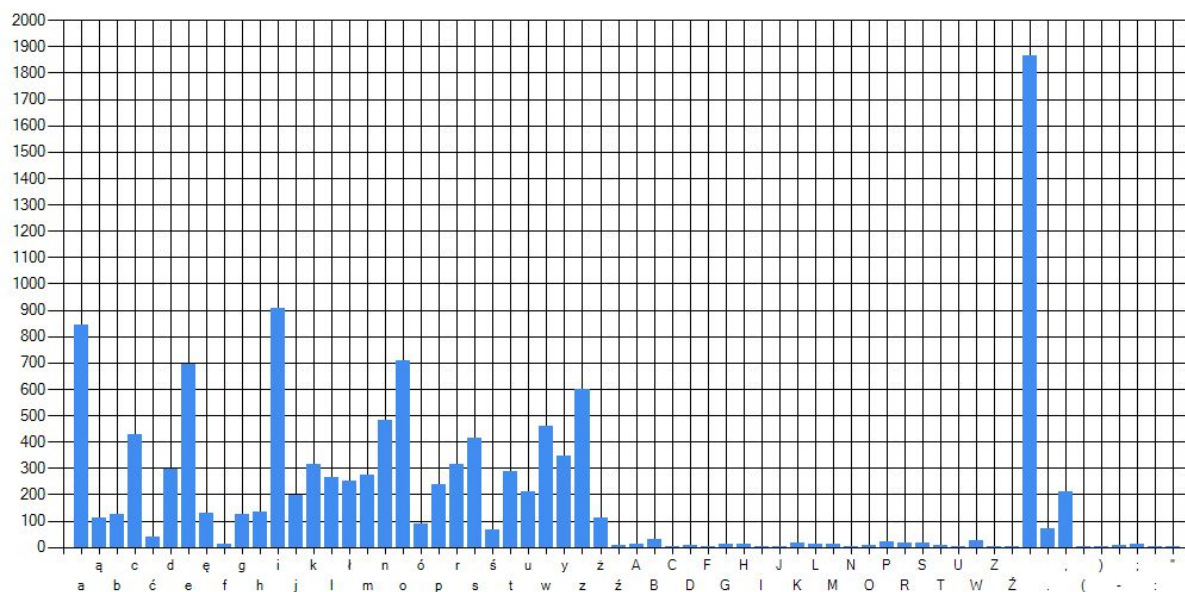
L.p.	Liczba wystąpień Prawdopodobieństwo							
	Znak	Tekst 1.	Znak	Tekst 2.	Znak	Tekst 3.	Znak	Tekst 4.
1.		1867 0,156208		1345 0,103989		2234 0,182367		1516 0,144574
2.	i	906 0,075803	e	1031 0,079712	e	1181 0,096408	e	994 0,094793
3.	a	846 0,070783	n	724 0,055976	t	835 0,068163	t	823 0,078486
4.	o	709 0,059321	i	724 0,055976	a	691 0,056408	o	759 0,072382
5.	e	694 0,058066	a	693 0,05358	o	637 0,052	r	672 0,064085

*Tabela 4.* prezentuje najwyższe wyniki osiągnięte w *Tekst n.* pod względem ilości informacji.

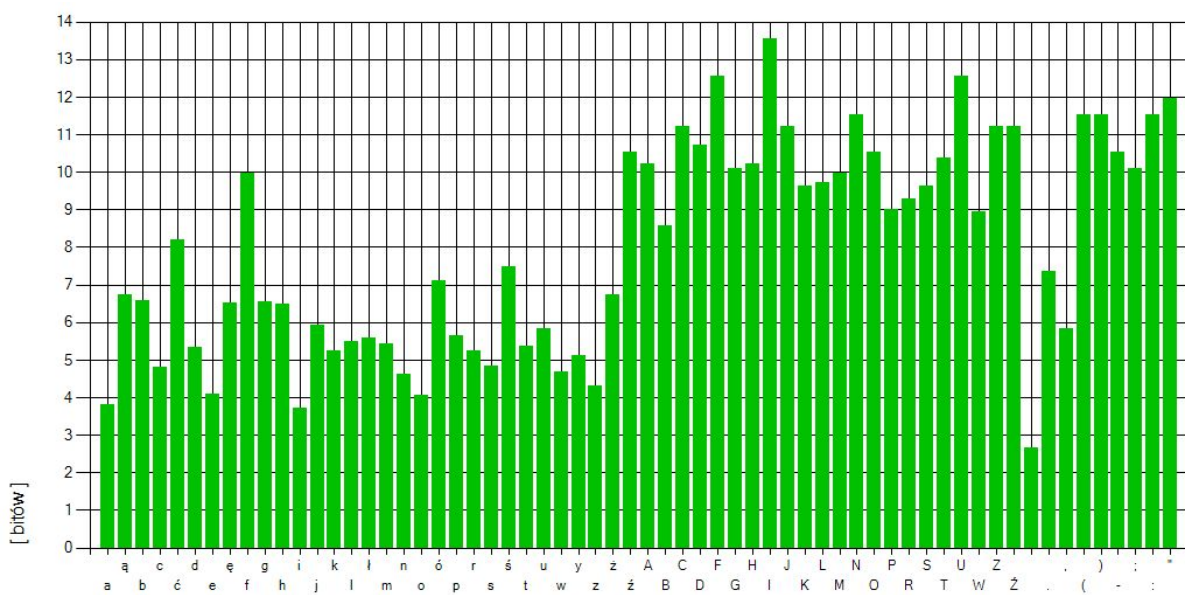
*Tabela 4. Top 5 - Ilość informacji*

L.p.	Ilość informacji							
	Znak	Tekst 1.	Znak	Tekst 2.	Znak	Tekst 3.	Znak	Tekst 4.
1.	I	13,539251	B	13,664782	(	13,574017	E	12,35414
2.	F	12,547864	H	13,664782	)	13,574017	N	12,35414
3.	U	12,547864	Ż	13,664782	:	13,574017	H	11,771697
4.	\n	11,960025	F	12,073588	J	12,58284	M	11,771697
5.	N	11,543551	Q	12,073588	K	12,58284	P	11,771697

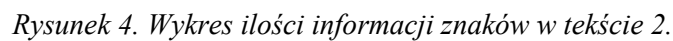
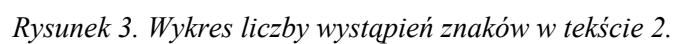
Rysunki 1-8. prezentują wykresy liczb wystąpień danych znaków oraz ilości informacji z nimi związanymi.

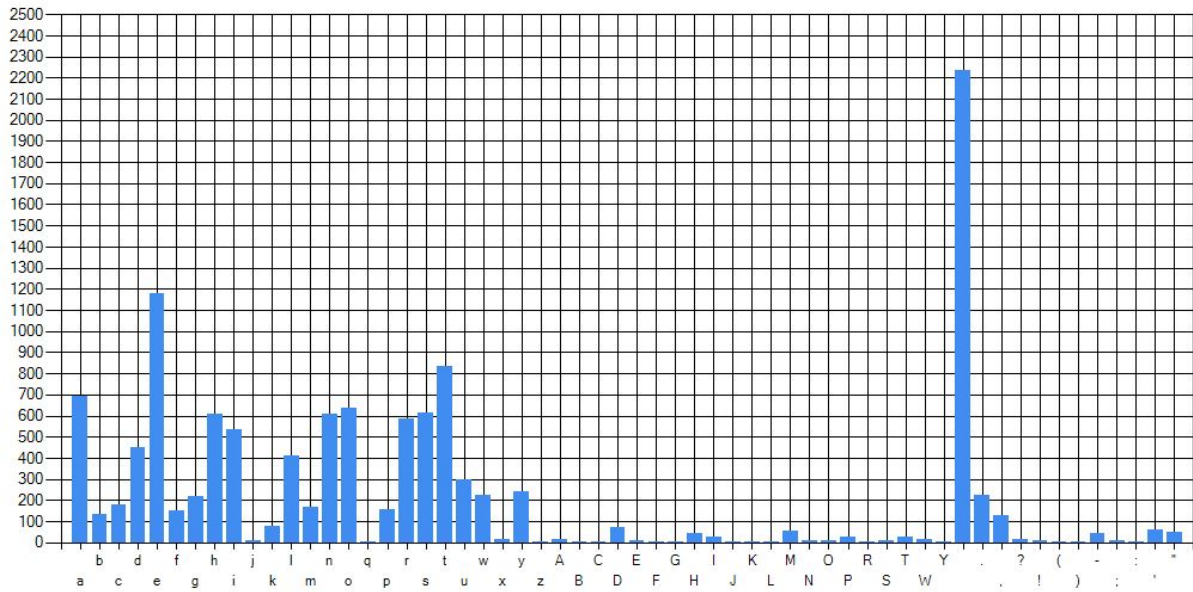


Rysunek 1. Wykres liczby wystąpień znaków w tekście 1.

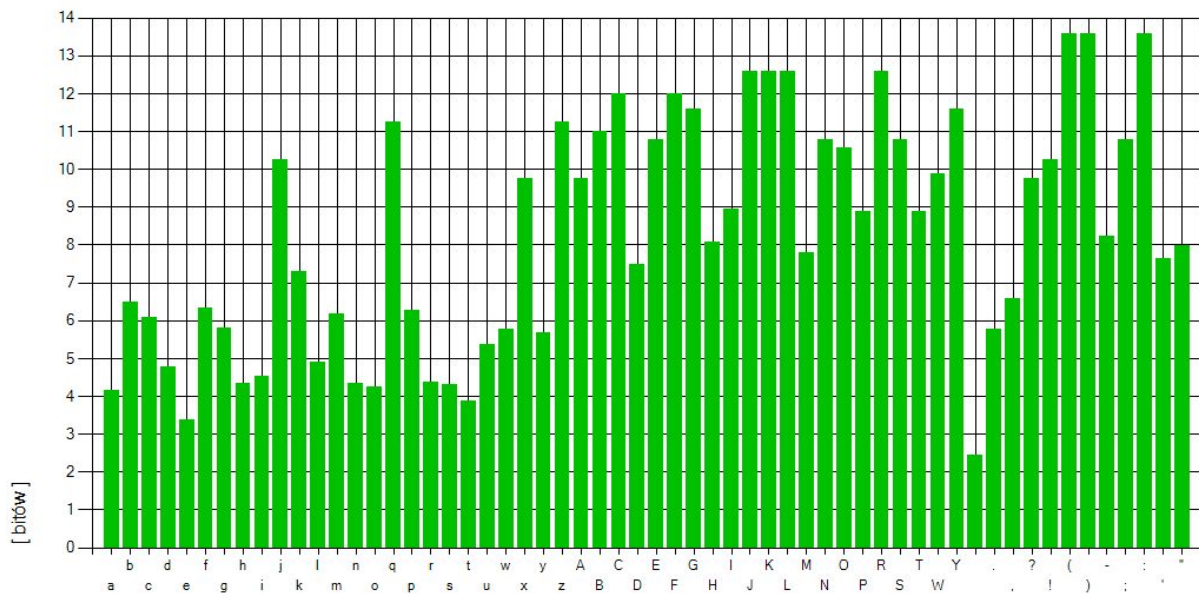


Rysunek 2. Wykres ilości informacji znaków w tekście 1.



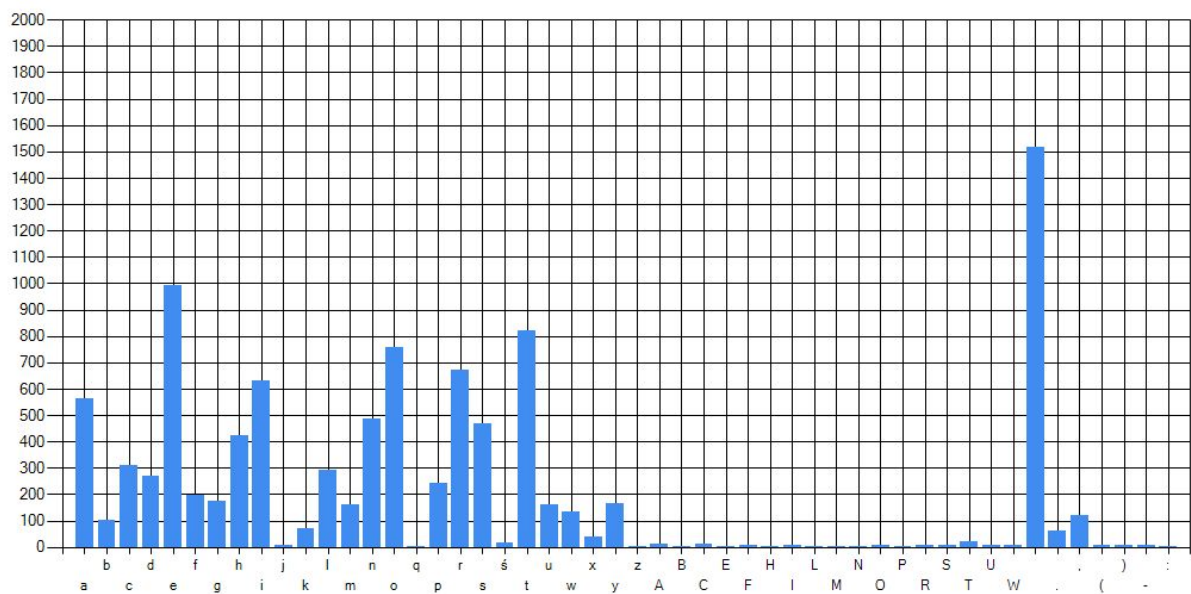


Rysunek 5. Wykres liczby wystąpień znaków w tekście 3.

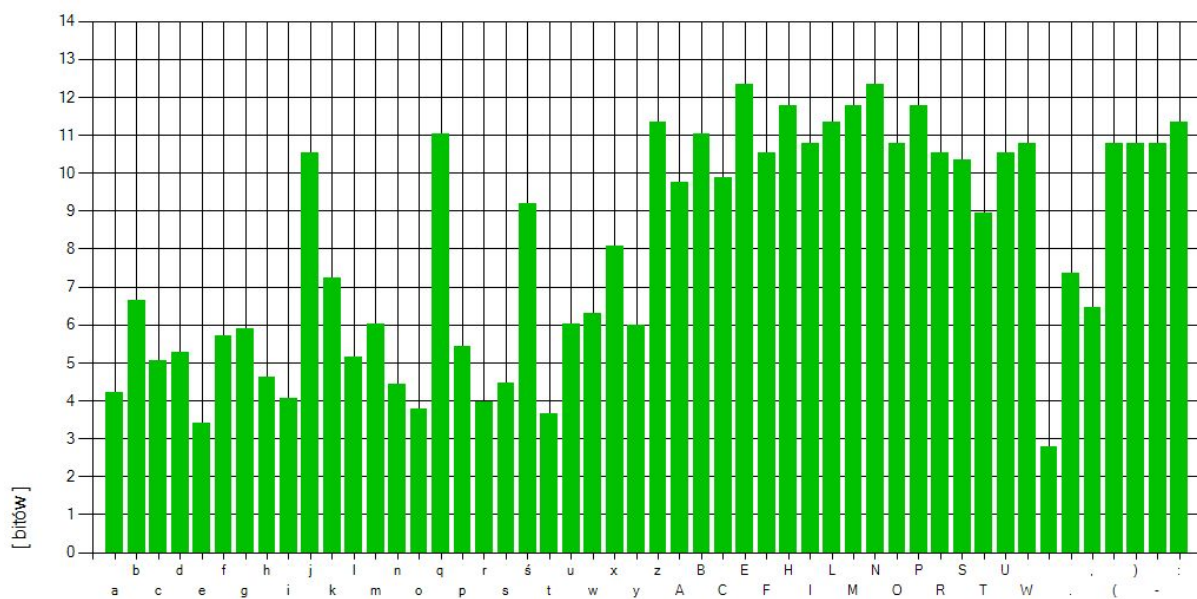


Rysunek 6. Wykres ilości informacji znaków w tekście 3.





Rysunek 7. Wykres liczby wystąpień znaków w tekście 4.



Rysunek 8. Wykres ilości informacji znaków w tekście 4.