Experimental Study of Information Technology for Detecting the Electronic Mass Media PR-effect based on Statistical Analysis

Oleg Hatyan[1] [0000-0003-2754-6938], Myroslav Ryabyy[1] [0000-0002-9651-9135],
Andriy Fesenko[2][0000-0001-5154-5324], Vitaliy Kyschenko[3] [0000-0003-4281-7812]
and Madina Bauyrzhan[4] [0000-0002-8287-4283], Anton Petrov [3][0000-0003-3731-4276]

[1] Interregional Academy of Personal Management, Kyiv, Ukraine
[2] Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[3] National Aviation University, Kyiv, Ukraine
[4] Satbayev University, Almaty, Kazakhstan
oleg.hatyan@gmail.com, m.o.ryabyy@gmail.com,
aafesenko@gmail.com, vitaliy.kyschenko@ukr.net,
madina890218@gmail.com, anton.a.petrov@gmail.com

Abstract. Today the dynamics and amount of modern information flows created by the Internet media content make a significant load of work for an expert while preparing an analytical conclusion on certain subject, industry and general trends. In the paper, the estimation of the limit values for detecting the PR-effect of electronic mass media integrated index is experimentally substantiated by conducting comparative and statistical analysis of numerical sequences of linguistic statistics for sets of messages which are intended a priori for purely "unbiased" reporting and documents with vocabulary characteristical for information influence (direct action commercial texts).

Keywords: Numerical Sequence, Statistical Analysis, Limit Value, Information Influence, PR-effect Index.

1 Introduction

The dynamics and amount of modern information flows created by the Internet media content (news feeds, blogs, micro blogs, audio and video streams etc.), make a significant load of work for an expert while preparing an analytical conclusion on certain subject, industry and general trends. In addition, as proved by the works of scientists and specialists devoted to the information space [1] (including our own [2]), today the nature of presenting information is somewhat different from merely reporting. Significant amount possesses the features of emotivity and manipulation. In copywriting, articles with such properties are called "direct action commercial texts". Thus, the bias of information sources, with the use of modern text construction and sequencing technology (target is the influence level), is capable of carrying hidden objectives of PR-effect aimed at the target audience. Doubtlessly, this is a factor that greatly increases the error rate in preparation and decision making.

Thus, the use of automated detection of "PR-effect" by applying an automated algorithm for the initial analysis of a given text and/or a load of texts, which would allow to pre-evaluate and decompose the flow of messages into those representing collections of documents with the vocabulary characteristical for “PR-effect” and the others, which, in our opinion, is an essential tool for analytical work and information support of decision-making.

Such a technological method was proposed by us in [2]. The idea is to calculate the "PR-effect" index for the thematic information flow (TIF) within a certain time span via the mathematical expectation of the daily TIF "PR-effect" index estimation:

$$PRI(M_{TIP\Delta t}) = M\left[PRI(M_{TIP\Delta t})_i\right], \tag{1}$$

where TIF "PR-effect" index for the i-th day makes 1/3 of the number of messages with the features of "non-eventivity", "emotionality", "manipulativity" to the total number of messages of the thematic vector MTIP:

$$PRI(M_{TIP\Delta t})_i = \frac{(|M_{TIP\Delta ti}| - P_{EV}(M_{TIP\Delta ti})) + P_{EM}(M_{TIP\Delta ti}) + P_{MA}(M_{TIP\Delta ti})}{3 \times |M_{TIP\Delta ti}|} \tag{2}$$

$$P_r(M_{TIP\Delta ti}) = \sum_{\ell=1}^{|M_{TIP\Delta ti}|} P_r(m_\ell) \tag{3}$$

where r={EV, EM, MA} is the index for the characteristic with linguistic features, respectively {"eventivity", "emotionality", "manipulativity"}; |MTIPΔti| is the power of a set of TIF messages for the i-th day.

Based on the construction, the value of PRI(MTIPΔt) is always positive and falls within the range between 0 and 1.

The threshold value α is set as the criterion for making a decision on the presence of "PR-effect" in the TIF. Thus if

$$PRI(M_{TIP\Delta t}) > \alpha, \tag{4}$$

then such a TIF represented by the thematic vector MTIPΔt possesses features of "PR-effect".

Consequently, the given value α sets the basis for the automated decision-making on detecting the PR-effect. Therefore, the experimental substantiation of its estimate value is significantly relevant.

2 Analysis of recent research and publications

Nowadays, research and evaluation of information flows (IFs), including those created by electronic media, is mostly carried out within convention-analytical research. The respective criteria for assessing the quality of quantitative content analyses based on the requirements of national and foreign methodologists is suggested by

Ivanov O.V. [3]. Fedorenko R.M. considers the main objects, threats and negative factors of ensuring information security in the military domain, characterizes the existing level of information space monitoring automation, suggests the use of content monitoring methods in order to increase the efficiency of providing information security of Ukraine [4].

Among the approaches to solving the problem of assessing IF texts created by electronic media, the methodology for assessing the quality of messages' resonance should be noted, which is the original development of one of the leading national consulting companies NOKS FISHES [5], and can be divided into the next 4 steps .

Step 1. Encoding the text array and giving the subject of research resonance characteristics

Text characteristics

S – size

E – genre

R – distribution (advertizing or not)

Reference characteristics

I – subject/object

Tmc – indirect reference

Tmk – the nature of speaker's reference

Tme – the nature of independent experts' reference

Ta – author's reference tone value

Ts – event reference tone value

To – total reference tone value

C – IF relevance

Mass media characteristics

O – printing

V – type

P – periodicity

G – geography

M – maginality

Step 2. Intermediate characteristics of reference quality

Saturation - reflects the proportion of influence on the quality of the company's information drive of its subjective representation in the media and the representation of the company's speakers in the infodrive.

= (I + Tmk + Tme) * Tmc= (subjectivity + total speakers' and independent experts' reference characteristics) * indirect reference.

K(tl) Logical presence coefficient – reflects the proportion (%) of the company's infodrive logical presence in the text array. Includes the tone value and subjective saturation of reference for both the given company's and the other companies of the given trend

= (logical presence coeff. + emotional presence coeff) / 2

=[K(l)+K(t)] / 2

K(l) – subjective representation coeff. = D of the given object / Σ (D of all the objects in the publication)

K(t) – emotional representation coeff. = |To| of the given object / Σ (|To| of all the objects in the publication)

Pg – potential influence of mass media rate

Step 3. Integrated characteristics of the reference quality

W – media’s loyalty to IF and the demand for media’s IF

Includes the qualitative characteristics of media’s perception of the company's infodrive or shows the editorial policy vector for the company's infodrive

= image * abstract * distribution type * author’s tone value * (text array size + publication type) * reference type * saturation * logical presence coeff.

Y – Event saturation of the client’s media area

Includes qualitative characteristics of the text array’s event filling by the company’s infodrive

= image * abstract * eventivity * event tone value * (text array size + publication type) * reference type * saturation * logical presence coeff.

Z – Probable influence – informational drive (ID) towards the media’s audience

Determines the probability of reporting the company's information to the general audience

= marginality * eventivity * potential influence of mass media coeff. * (7 – publication size) * reference type * subjectivity * logical presence coeff.

Step 4. Rating

R – integrated resonance quality index

If $W < 0$, $Y > 0$, $|W|>|Y|$, $\beta = - \surd\, W2 - Y2 + Z2$

If $W < 0$, $Y > 0$, $|W|<|Y|$, $\beta = \surd\, Y2 - W2 + Z2$

If $W < 0$, $Y < 0$, $\beta = - \surd\, W2 + Y2 + Z2$

If $Y < 0$, $W > 0$, $|W|<|Y|$, $\beta = - \surd\, Y2 - W2 + Z2$

If $Y < 0$, $W > 0$, $|W|>|Y|$, $\beta = \surd\, W2 - Y2 + Z2$

If $Y > 0$, $W > 0$, $\beta = \surd\, W2 + Y2 + Z2$

However, these tools, as well as the previously mentioned areas of research, are based on methods of expert evaluation, which in the general case does not exclude the systemic error problem mentioned above.

In previous works, by analyzing the trends of the TIFs created while categorizing the general IF and the synthesis of linguistic statics (both standalone messages and calculated for a specific TIF load that had the characteristic features of PR-effect) [2] we: formulated and experimentally proved the hypothesis about the difference between the trends of thematic information flows of unbiased reporting and "PR-effect" in the characteristic space of linguistic features; developed a method for decomposing the general information flow into the constituent thematic streams; indicated a way of distinguishing thematic streams with linguistic features of "PR-effect"; empirically established a value for the "PR-effect" index of the thematic flow; presented the empirical studies data, which allow to assess the limit conditions for the decision-making on the "PR-effect" index of the thematic flow.

In this paper, we will focus on the detailed substantiation of the "PR-effect" index and the estimation of the criterion α for the decision-making on "PR-effect" in the TIF (4). Thus, the purpose of this paper is to formally present the technology for detecting the "PR-effect" of electronic media by: conducting a comparative analysis for sets of messages that are a priori intended for "unbiased" reporting, and documents with the vocabulary characteristic for the "PR-effect"; conducting statistical analysis of the numeric sequences of estimates of $PRI(MTIP\Delta t)i$ and experimental sets of data (doc-

uments) of both types; setting (substantiation) of the limit conditions for estimation of the value α of the method "PR-effect" detection, which is the task of the present work.
The subject of the study is two groups of information messages. One of them is to be focused on "unbiased" reporting, the other one contains vocabulary characteristic for the "PR-effect".
The object of the study is the boundary criterion α of the "PR-effect" index.

3 The main material of research study

To conduct a study to determine the limit value of the criterion α we performed the following:
1. In the system of information space analysis (SISA) [6] a data bank was created (conventional name SLOVAR). This way, for the created bank, we received all the necessary tools for calculating the estimates for the presence of the "PR-effect", namely trained by expert evaluation algorithms for calculating linguistic features, "eventivity", "emotionality", and "manipulativity" for each document are represented by a block diagram (Figure 1). The presented algorithm is a formalized basis of detecting the "PR-effect" of IF of electronic mass media. The algorithm is designed to evaluate texts in both Ukrainian and Russian.

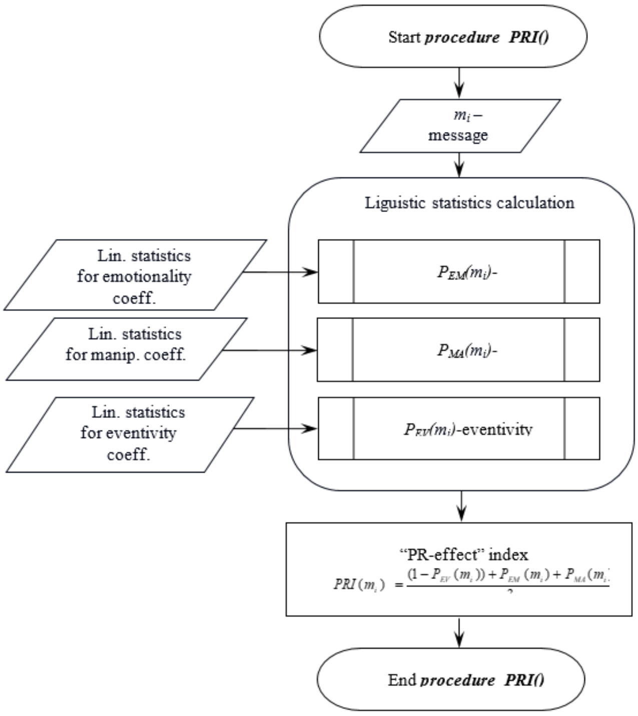

Start procedure PRI()
m_i – message
Liguistic statistics calculation
Lin. statistics for emotionality coeff.
$P_{EM}(m_i)$-
Lin. statistics for manip. coeff.
$P_{MA}(m_i)$-
Lin. statistics for eventivity coeff.
$P_{EV}(m_i)$-eventivity
"PR-effect" index
$PRI(m_i) = \frac{(1 - P_{EV}(m_i)) + P_{EM}(m_i) + P_{MA}(m_i}{?}$
End procedure PRI()

Fig. 1. Algorithm for the detection of the "PR-effect" based on determining the linguistic statistics of a given message

2. Two groups of messages were formed that correspond to specifics of "unbiased" reporting and containing the vocabulary characteristic for the "PR-effect". Hypothetically, texts correspondent with "unbiased" reporting should be articles of encyclopedic dictionaries. Therefore, we created the first group of documents based on the electronic version of the Great Encyclopedic Dictionary [7] (source: https://www.vslovar.ru/). The dictionary contains more than 80,000 articles, including about 20,000 biographical ones (further denoted as C1). At the same time, 58283 articles from the specified electronic resource, the text size of which is more than 77 characters, was collected as the entry to the data bank. Another group of documents that have the vocabulary characteristic for "PR-effect" was selected from the most popular Google search engine professional copywriter sites (examples of direct action texts from the portfolio and currently relevant as SEO texts on customer sites) of Ukrainian, Belarusian and Russian segments of the Internet with a total of 279 documents (denoted as C2):

- "Copywriter Dmitry Kot" - http: //www.mastertext.spb.ru/index.html;
- "Protect. Professional copywriting. " Company. http://protext.by/;
- "Marmore." Copywriter Marina Greben ". http://marmore-text.ru/;
- "PromoText. Copywriting. " Company. https://promotext.com.ua/

3. Both sets of documents were entered into the SLOVAR data bank; for each document the linguistic features of "eventivity", "emotionality", "manipulativity", and also an estimation of the presence of "PR-effect" - PRI(C1,2) (PRI(MTIPΔt)i were calculated using the SISA tools and the formula (2). Thus obtained individual numerical sequences for C1 and C2 were arranged in the form of CSV-tables suitable for further statistical analysis by the relevant software (MS Excel and Statistica 6.0).

Notes:

1. As for our choice of texts, which contain the vocabulary characteristic for the "PR-effect" of the product of professional copywriting. We believe that a quote describing the purpose of copywriting, taken from the section "Professional look" site of one of the Ukrainian copywriting companies, is quite revealing (source: http://cookiezz.com.ua/copyright), “While writing texts we try to make them as simple for the client to read as possible. At the same time, they should be informative and aimed at SEO site promotion. By ordering a single text item of a full site content development from our company, you can be certain in receiving the content of high quality, that will rise the site’s distribution and make already existing clients by from you.”

2. When constructing a numerical sequence of estimates of PRI(C1) (by the expression (2) - PRI(MTIPΔt)i) for C1, we came to the conclusion that among the preselected 58283 vocabulary articles of the Great Encyclopedic Dictionary (GED), a significant portion is made by short texts . This way, only 11.7% of the documents of the GED (6821 articles), are text of more than 512 characters. At the same time, the number of such documents is indicative for our study and they formed the basis for constructing a numerical sequence of estimates PRI(C2) (by the expression (2) - PRI(MTIPΔt)i) for C1.

Taking into account the above, the algorithm for the experimental determination of the limit values of the value α for the "PR-effect" detection can be represented by a block diagram (Figure 2).

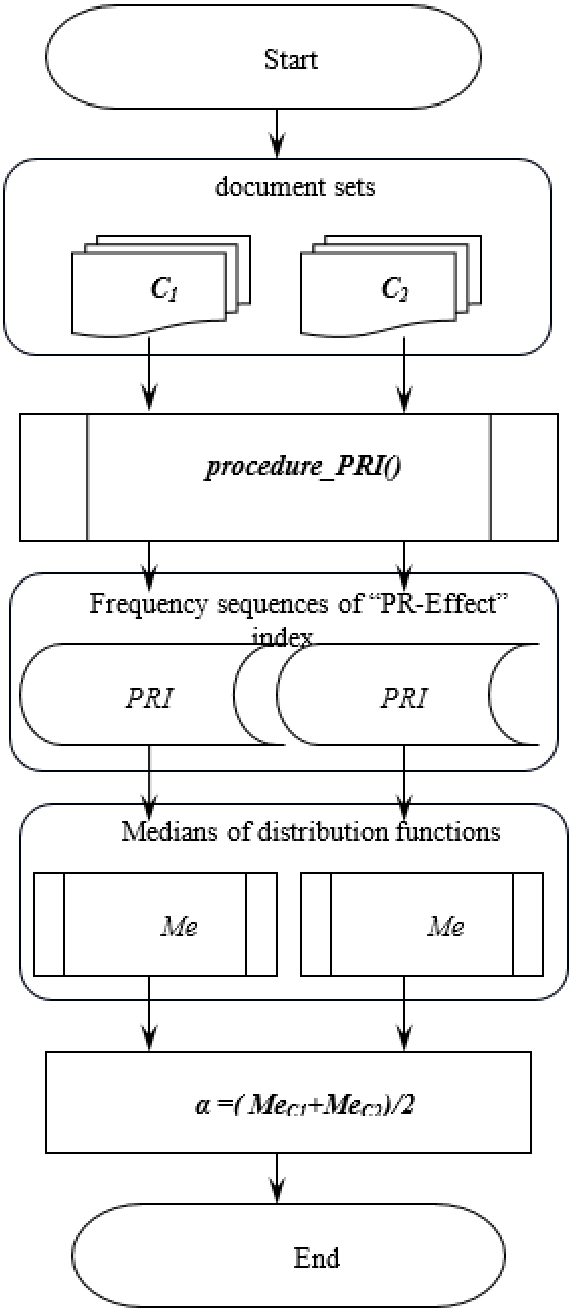

Start
document sets
C_1
C_2
procedure_PRI()
Frequency sequences of "PR-Effect" index
PRI
PRI
Medians of distribution functions
Me
Me
α =(Me_{C1}+Me_{C2})/2
End

Fig. 2. Algorithm for determining the limit value of α for "PR-effect" detection

Thus, for the purpose of visual comparison and estimation of the type and nature of the probability distribution FC1 of the random values obtained for C1 and FC2 for C2 frequency distribution histograms of estimates PRI(C1,2)i for each document in the corresponding group are constructed. Given that the construction of the set of values PRI(C1,2)i, at $i = \overline{1, n}$ where $n \to \infty$ is approximated by a continuous function in the range between 0 and 1, and the number of class intervals calculated according to the Sturges rule [8]: n=1+log2N, meaning that for the sequence C1: n=1+log(6821)=13,736 intervals (scale step 0,07), C2: n=1+log(277)=9,114 intervals (scale step 0,11). Resulting histograms with a scale step 0,1, constructed for sequences C1 and C2 using the software Statistica 6.0 are presented on the Figure 3.

However, to improve the detail of the qualitative estimation of the experimental data distribution, we set the scale step of 0.025 (40 intervals) for both sequences.

Grouping the values of the numerical sequences C1 and C2 in the established class intervals allowed to determine the frequencies of the corresponding estimates PRI(C1,2)i (Table 1).

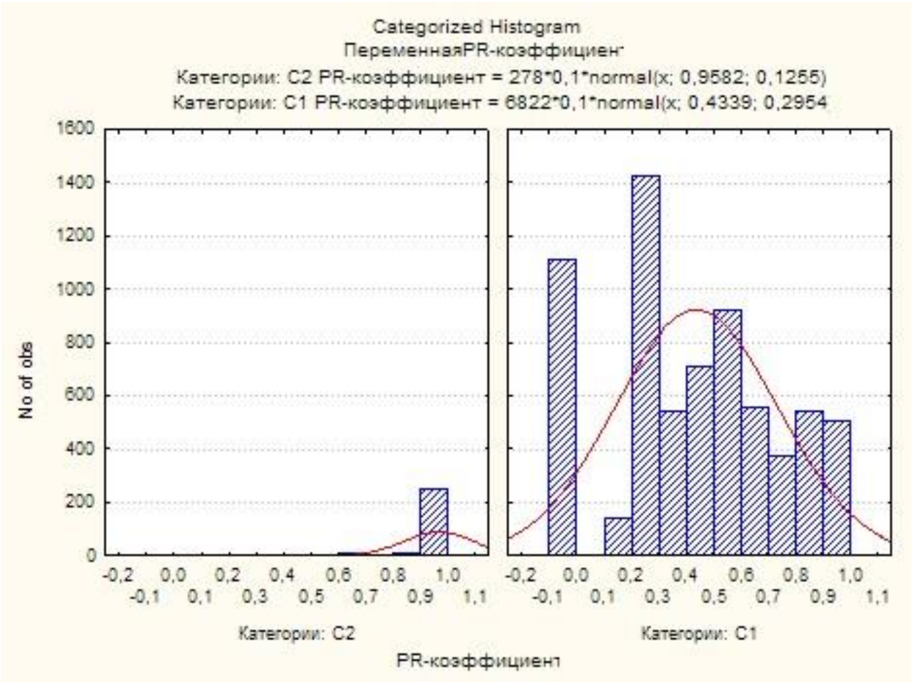
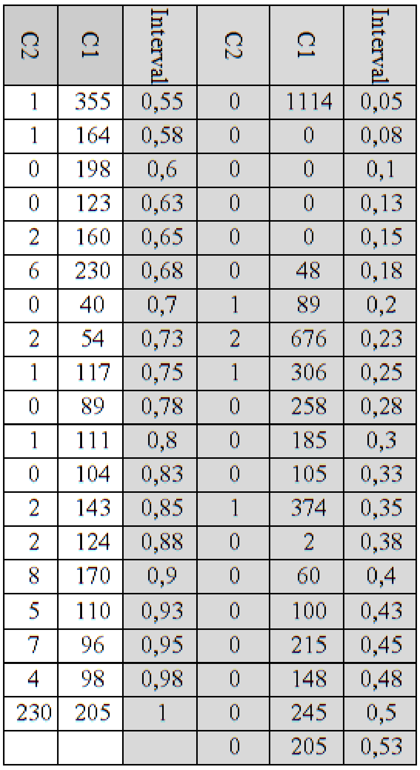


Fig. 3. Histograms of distribution functions for the sequences C1 and C2 using Statistica 6.0

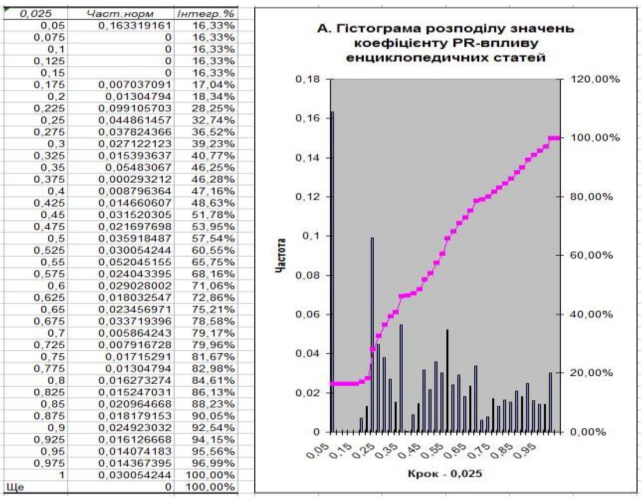
As the next step, the histograms of the frequency distribution of the values PRI(C1,2) are built. To do so, the resulting numerical series are normalized at the power of the corresponding sets of documents (samples C1 and C2).

Table 1. Frequency values of the PR-effect index for numerical sequences C1 and C2



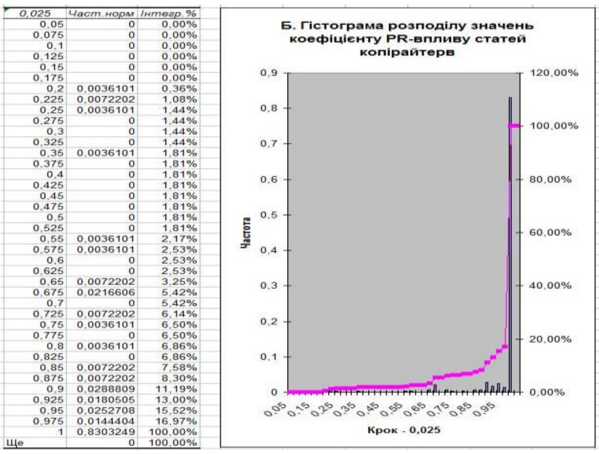
C2	C1	Interval	C2	C1	Interval
1	355	0,55	0	1114	0,05
1	164	0,58	0	0	0,08
0	198	0,6	0	0	0,1
0	123	0,63	0	0	0,13
2	160	0,65	0	0	0,15
6	230	0,68	0	48	0,18
0	40	0,7	1	89	0,2
2	54	0,73	2	676	0,23
1	117	0,75	1	306	0,25
0	89	0,78	0	258	0,28
1	111	0,8	0	185	0,3
0	104	0,83	0	105	0,33
2	143	0,85	1	374	0,35
2	124	0,88	0	2	0,38
8	170	0,9	0	60	0,4
5	110	0,93	0	100	0,43
7	96	0,95	0	215	0,45
4	98	0,98	0	148	0,48
230	205	1	0	245	0,5
			0	205	0,53

Resulting histograms are presented at Figures 4 and 5. The light color represents an integral distribution function of the coefficient PRI(C1,2)i.



0,025	Част.норм	Інтегр.%
0,05	0,163319161	16,33%
0,075	0	16,33%
0,1	0	16,33%
0,125	0	16,33%
0,15	0	16,33%
0,175	0,007037091	17,04%
0,2	0,01304794	18,34%
0,225	0,099105703	28,25%
0,25	0,044861457	32,74%
0,275	0,037824366	36,52%
0,3	0,027122123	39,23%
0,325	0,015393637	40,77%
0,35	0,05483067	46,25%
0,375	0,000293212	46,28%
0,4	0,008796364	47,16%
0,425	0,014660607	48,63%
0,45	0,031520305	51,78%
0,475	0,021697698	53,95%
0,5	0,035918487	57,54%
0,525	0,030054244	60,55%
0,55	0,052045155	65,75%
0,575	0,024043395	68,16%
0,6	0,029028002	71,06%
0,625	0,018032547	72,86%
0,65	0,023456971	75,21%
0,675	0,033719396	78,58%
0,7	0,005864243	79,17%
0,725	0,007916728	79,96%
0,75	0,01715291	81,67%
0,775	0,01304794	82,98%
0,8	0,016273274	84,61%
0,825	0,015247031	86,13%
0,85	0,020964668	88,23%
0,875	0,018179153	90,05%
0,9	0,024923032	92,54%
0,925	0,016126668	94,15%
0,95	0,014074183	95,56%
0,975	0,014367395	96,99%
1	0,030054244	100,00%
Ще	0	100,00%

Fig. 4. Table of the normalized frequencies of the coeff. PRI(C1)i for GED articles



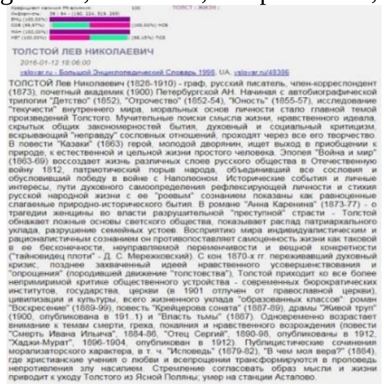
0,025	Част.норм	Інтегр.%
0,05	0	0,00%
0,075	0	0,00%
0,1	0	0,00%
0,125	0	0,00%
0,15	0	0,00%
0,175	0	0,00%
0,2	0,0036101	0,36%
0,225	0,0072202	1,08%
0,25	0,0036101	1,44%
0,275	0	1,44%
0,3	0	1,44%
0,325	0	1,44%
0,35	0,0036101	1,81%
0,375	0	1,81%
0,4	0	1,81%
0,425	0	1,81%
0,45	0	1,81%
0,475	0	1,81%
0,5	0	1,81%
0,525	0	1,81%
0,55	0,0036101	2,17%
0,575	0,0036101	2,53%
0,6	0	2,53%
0,625	0	2,53%
0,65	0,0072202	3,25%
0,675	0,0216606	5,42%
0,7	0	5,42%
0,725	0,0072202	6,14%
0,75	0,0036101	6,50%
0,775	0	6,50%
0,8	0,0036101	6,86%
0,825	0	6,86%
0,85	0,0072202	7,58%
0,875	0,0072202	8,30%
0,9	0,0288809	11,19%
0,925	0,0180505	13,00%
0,95	0,0252708	15,52%
0,975	0,0144404	16,97%
1	0,8303249	100,00%
Ще	0	100,00%

Fig. 5. Table of the normalized frequencies of the coeff. PRI(C_2)$_i$ for copywriting articles and distribution histogram.

As it can be seen from the figures (qualitative analysis of probability distribution for FC1 and FC2):

1. Probability distributions both of FC1 for C1 (three modes clearly expressed) and FC2 for C2 are not Gaussian (that is, they have a non-parametric form), and therefore the use of standard statistical estimates for the calculation of numerical characteristics (e.g. mean value or deviation) of the distribution function, as well as the use of Student's criterion to confirm the hypothesis regarding homogeneity of the samples, will not be correct;
2. 83,03% of C2 (set of texts by copywriters - Fig. 5) received an estimation of PR-effect index at PRI(C2)>0,975, while the same level of evaluation corresponded to only 3,01% of the documents in C1 (texts from GED - Fig. 4);
3. Certain number of documents (18.33%) from C1 (GED texts - Fig. 4) have an estimate of PRI(C1)>0,75, which is due to the literary nature of the information presentation in some of the encyclopedic articles. These texts include:

- biographies of outstanding artists, historians, and politicians, for example:



ТОЛСТОЙ ЛЕВ НИКОЛАЕВИЧ

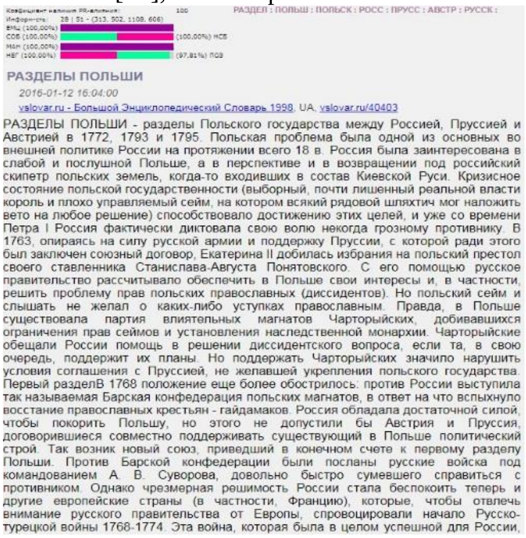
2016-01-12 18:06:00

vslovar.ru - Большой Энциклопедический Словарь 1998, UA, vslovar.ru/48386

ТОЛСТОЙ Лев Николаевич (1828-1910) - граф, русский писатель, член-корреспондент (1873), почетный академик (1900) Петербургской АН. Начиная с автобиографической трилогии "Детство" (1852), "Отрочество" (1852-54), "Юность" (1855-57), исследование "текучести" внутреннего мира, моральных основ личности стало главной темой произведений Толстого. Мучительные поиски смысла жизни, нравственного идеала, скрытых общих закономерностей бытия, духовный и социальный критицизм, вскрывающий "неправду" сословных отношений, проходят через все его творчество. В повести "Казаки" (1863) герой, молодой дворянин, ищет выход в приобщении к природе, к естественной и цельной жизни простого человека. Эпопея "Война и мир" (1863-69) воссоздает жизнь различных слоев русского общества в Отечественную войну 1812, патриотический порыв народа, объединивший все сословия и обусловивший победу в войне с Наполеоном. Исторические события и личные интересы, пути духовного самоопределения рефлексирующей личности и стихия русской народной жизни с ее "роевым" сознанием показаны как равноценные слагаемые природно-исторического бытия. В романе "Анна Каренина" (1873-77) - о трагедии женщины во власти разрушительной "преступной" страсти - Толстой обнажает ложные основы светского общества, показывает распад патриархального уклада, разрушение семейных устоев. Восприятию мира индивидуалистическим и рационалистичным сознанием он противопоставляет самоценность жизни как таковой в ее бесконечности, неуправляемой переменчивости и вещной конкретности ("тайновидец плоти" - Д. С. Мережковский). С кон. 1870-х гг. переживавший духовный кризис, позднее захваченный идеей нравственного усовершенствования и "опрощения" (породившей движение "толстовства"), Толстой приходит ко все более непримиримой критике общественного устройства - современных бюрократических институтов, государства, церкви (в 1901 отлучен от православной церкви), цивилизации и культуры, всего жизненного уклада "образованных классов": роман "Воскресение" (1889-99), повесть "Крейцерова соната" (1887-89), драмы "Живой труп" (1900, опубликована в 191..1) и "Власть тьмы" (1887). Одновременно возрастает внимание к темам смерти, греха, покаяния и нравственного возрождения (повести "Смерть Ивана Ильича", 1884-86, "Отец Сергий", 1890-98, опубликованы в 1912, "Хаджи-Мурат", 1896-1904, опубликован в 1912). Публицистические сочинения морализаторского характера, в т. ч. "Исповедь" (1879-82), "В чем моя вера?" (1884), где христианские учения о любви и всепрощении трансформируются в проповедь непротивления злу насилием. Стремление согласовать образ мысли и жизни приводит к уходу Толстого из Ясной Поляны; умер на станции Астапово.

Fig. 6. Example of biography

- description of historical events and countries that have directly participated in them or had geopolitical influence [11], for example:


Коэфициент наличия PR-влияния: 100
Информ-сть: 28 | 51 - (313, 502, 1108, 606)
ЕМЦ (100,00%)
СОБ (100,00%)
(100,00%) НСБ
МАН (100,00%)
НЕГ (100,00%)
(97,81%) ПОЗ

РАЗДЕЛ : ПОЛЬШ : ПОЛЬСК : РОСС : ПРУСС : АВСТР : РУССК :

РАЗДЕЛЫ ПОЛЬШИ

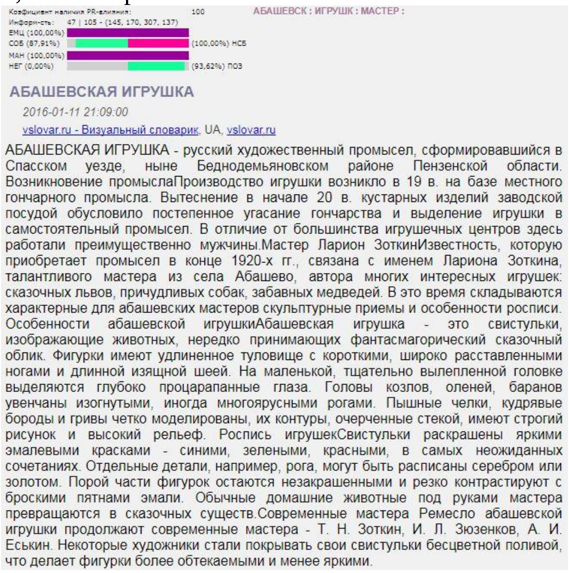
2016-01-12 16:04:00

vslovar.ru - Большой Энциклопедический Словарь 1998, UA, vslovar.ru/40403

РАЗДЕЛЫ ПОЛЬШИ - разделы Польского государства между Россией, Пруссией и Австрией в 1772, 1793 и 1795. Польская проблема была одной из основных во внешней политике России на протяжении всего 18 в. Россия была заинтересована в слабой и послушной Польше, а в перспективе и в возвращении под российский скипетр польских земель, когда-то входивших в состав Киевской Руси. Кризисное состояние польской государственности (выборный, почти лишенный реальной власти король и плохо управляемый сейм, на котором всякий рядовой шляхтич мог наложить вето на любое решение) способствовало достижению этих целей, и уже со времени Петра I Россия фактически диктовала свою волю некогда грозному противнику. В 1763, опираясь на силу русской армии и поддержку Пруссии, с которой ради этого был заключен союзный договор, Екатерина II добилась избрания на польский престол своего ставленника Станислава-Августа Понятовского. С его помощью русское правительство рассчитывало обеспечить в Польше свои интересы и, в частности, решить проблему прав польских православных (диссидентов). Но польский сейм и слышать не желал о каких-либо уступках православным. Правда, в Польше существовала партия влиятельных магнатов Чарторыйских, добивавшихся ограничения прав сеймов и установления наследственной монархии. Чарторыйские обещали России помощь в решении диссидентского вопроса, если та, в свою очередь, поддержит их планы. Но поддержать Чарторыйских значило нарушить условия соглашения с Пруссией, не желавшей укрепления польского государства. Первый разделВ 1768 положение еще более обострилось: против России выступила так называемая Барская конфедерация польских магнатов, в ответ на что вспыхнуло восстание православных крестьян - гайдамаков. Россия обладала достаточной силой, чтобы покорить Польшу, но этого не допустили бы Австрия и Пруссия, договорившиеся совместно поддерживать существующий в Польше политический строй. Так возник новый союз, приведший в конечном счете к первому разделу Польши. Против Барской конфедерации были посланы русские войска под командованием А. В. Суворова, довольно быстро сумевшего справиться с противником. Однако чрезмерная решимость России стала беспокоить теперь и другие европейские страны (в частности, Францию), которые, чтобы отвлечь внимание русского правительства от Европы, спровоцировали начало Русско-турецкой войны 1768-1774. Эта война, которая была в целом успешной для России,

Fig. 7. Example of historical event

- some definitions, for example:


Коэфициент наличия PR-влияния: 100
Информ-сть: 47 | 105 - (145, 170, 307, 137)
ЕМЦ (100,00%)
СОБ (87,91%)
(100,00%) НСБ
МАН (100,00%)
НЕГ (0,00%)
(93,62%) ПОЗ

АБАШЕВСК : ИГРУШК : МАСТЕР :

АБАШЕВСКАЯ ИГРУШКА

2016-01-11 21:09:00

vslovar.ru - Визуальный словарик, UA, vslovar.ru

АБАШЕВСКАЯ ИГРУШКА - русский художественный промысел, сформировавшийся в Спасском уезде, ныне Беднодемьяновском районе Пензенской области. Возникновение промыслаПроизводство игрушки возникло в 19 в. на базе местного гончарного промысла. Вытеснение в начале 20 в. кустарных изделий заводской посудой обусловило постепенное угасание гончарства и выделение игрушки в самостоятельный промысел. В отличие от большинства игрушечных центров здесь работали преимущественно мужчины.Мастер Ларион ЗоткинИзвестность, которую приобретает промысел в конце 1920-х гг., связана с именем Лариона Зоткина, талантливого мастера из села Абашево, автора многих интересных игрушек: сказочных львов, причудливых собак, забавных медведей. В это время складываются характерные для абашевских мастеров скульптурные приемы и особенности росписи. Особенности абашевской игрушкиАбашевская игрушка - это свистульки, изображающие животных, нередко принимающих фантасмагорический сказочный облик. Фигурки имеют удлиненное туловище с короткими, широко расставленными ногами и длинной изящной шеей. На маленькой, тщательно вылепленной головке выделяются глубоко процарапанные глаза. Головы козлов, оленей, баранов увенчаны изогнутыми, иногда многоярусными рогами. Пышные челки, кудрявые бороды и гривы четко моделированы, их контуры, очерченные стекой, имеют строгий рисунок и высокий рельеф. Роспись игрушекСвистульки раскрашены яркими эмалевыми красками - синими, зелеными, красными, в самых неожиданных сочетаниях. Отдельные детали, например, рога, могут быть расписаны серебром или золотом. Порой части фигурок остаются незакрашенными и резко контрастируют с броскими пятнами эмали. Обычные домашние животные под руками мастера превращаются в сказочных существ.Современные мастера Ремесло абашевской игрушки продолжают современные мастера - Т. Н. Зоткин, И. Л. Зюзенков, А. И. Еськин. Некоторые художники стали покрывать свои свистульки бесцветной поливой, что делает фигурки более обтекаемыми и менее яркими.

Fig. 8. Example of definition

The descriptive statistics given in Table 2 also indicate a non-parametric distribution of FC1 and FC2.

Table 2. Descriptive statistics of the distribution of PR-effect index for numerical sequences C1 (FC1) and C2 (FC2)

Statistics	Sequence	
	C1	C2
Mean value (μ)	0,433882	0,958077
Standard error (σ2)	0,003577	0,007551
Median (Me)	0,44	1
Mode (Mo)	0	1
Standard deviation (s)	0,295416	0,125679
Expected mean square (D)	0,087271	0,015795
Excess (Ex)	-0,9799	19,92078
ExCrit (α=0,05)	0,814	0,818
Assymetry (As)	0,167379	-4,26553
AsCrit (α=0,05)	0,130	0,230
Interval	1	0,8167
Minimun (min)	0	0,1833
Maximum (max)	1	1
Intervals (n)	6821	277
Reliability level (95,0%)	0,007012	0,014865

Thus, given that for C1 and C2 |AsC1| > Ascrit, and |Ex C1| > Excrit [9], this can be concluded on the zeroth hypothesis H(0) deviation: the distribution of the estimation of the PR-effect index PRI(C1,2) for C1 and C2 corresponds with the Gaussian (normal) distribution. In addition, the distinctive negative excess and the right-hand asymmetry of C1 and a significant positive excess and left-hand asymmetry of C2 is present.

In addition, the median value is at the level of 0.44 for C1 and 1 for C2, based on which we will subsequently obtain the desired limit value of α for detecting "PR-effect".

The hypothesis of the homogeneity of independent samples C1 and C2 is then checked. Theoretical distribution functions FC1 and FC2 are unknown, that is, they belong to the same general set.

Then, the verified zeroth hypothesis H(0) is presented as: FC1 = FC2 unlike the opposite hypothesis H(1): FC1 ≠ FC2, while FC1 and FC2 are considered continuous.

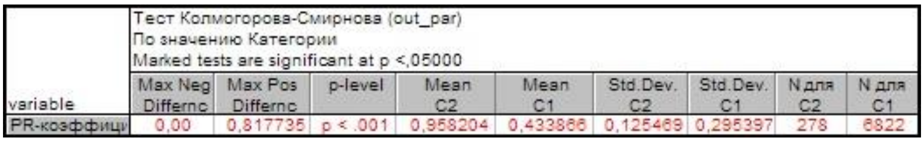
To test H(0) the statistics of the non-parametric Kolmagorov-Smirnov criterion is applied:

$$\lambda' = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \cdot \max\left|F_{C_1} - F_{C_2}\right|,$$

where FC1 and FC2 are empirical distribution functions with two samples of volumes n1 and n2.

Hypothesis H(0) is not confirmed, if the given statistic value λ′ is more than the critical λ′cr, ie λ′ > λ′cr, otherwise it is considered to be confirmed.

The hypothesis H(0) testing with the Kolmagorov-Smirnov criterion is conducted using the Statistica 6.0 software.



Тест Колмогорова-Смирнова (out_par)
По значению Категории
Marked tests are significant at p <,05000

variable	Max Neg Differnc	Max Pos Differnc	p-level	Mean C2	Mean C1	Std.Dev. C2	Std.Dev. C1	N для C2	N для C1
PR-коэффици	0,00	0,817735	p < .001	0,958204	0,433866	0,125469	0,295397	278	6822

Fig. 9. Results of testing the hypothesis H(0) with the Kolmogorov-Smirnov criterion

The Kolmagorov-Smirnov criterion has shown that the distribution functions FC1 and FC2 do not belong to the same general set (that is, the hypothesis H(0) deviates) with an error rate of 5% (Figure 9).

In other words, there is the 95% confidence in a significant deviation between the experimental data C1 (messages that correspond to the "unbiased" reporting) and C2 (documents that have characteristics of PR-effect vocabulary) PR-effect index values PRI(C1,2)i, as shown by the Box-and-whisker diagram (Figure 10) [10].

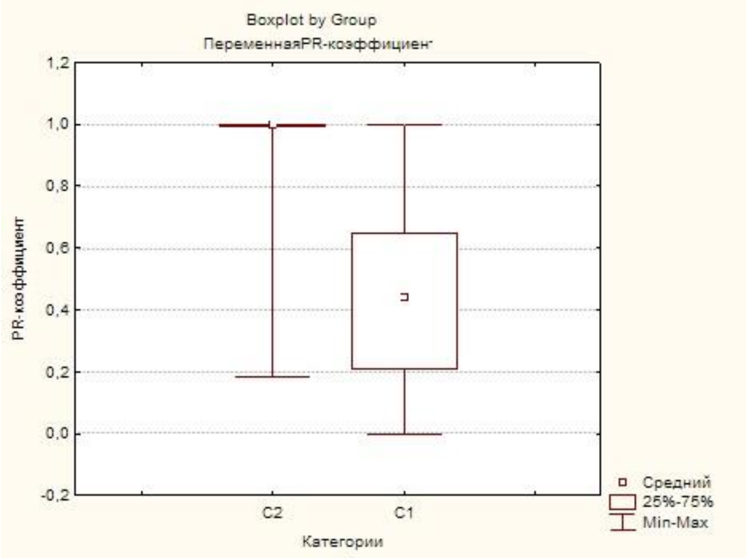


Fig. 10. Box–and–whisker diagram for experimental data from C1 and C2

Additionally, we obtained the updated median of experimental data sequences: for C1 MeC1 = 0,433866 and for C2, MeC2 = 0,958204, based on which we calculate the limit value of α for "PR-effect" detection as an average value of the medians. Then:

$$\alpha =(MeC1+MeC2)/2=(0{,}433866 + 0{,}958204)/2=0{,}696035.$$

At this, the value α = 0,696035 is taken as the limit value of the decision-making criterion for the presence of "PR-effect" (4), with an error rate of 5%.

4 Conclusions

To summarize, as a result of the study, we gave a formal statement for detecting "PR-effect" by electronic media (the algorithm is depicted as a flowchart in Figure 1). An experimental substantiation and qualitative analysis of the PR-effect index were performed, which showed the difference in the distribution of the probabilities of its frequencies for both experimental samples (C1 for the articles of the Great Encyclopedic Dictionary, C2 for articles by copywriters) from the Gaussian one (having a non-parametric form). By conducting statistical analysis of the numerical sequences of experimental data (C1 and C2), the hypothesis of the difference in the messages of "unbiased" reporting (C1) and those that contain characteristic vocabulary of PR-effect (C2) in the characteristic space of linguistic features was confirmed. The limit value of the decision-making criterion for the presence of "PR-effect" (4) $\alpha = 0.696035$, with an error rate of 5%, was obtained. At the same time, both qualitative and statistical analysis of the experimental material showed a certain difference in the value of the integral PR-effect index for the characteristic groups of documents. Therefore, as the tasks for further research, we will aim at developing an algorithm for automatic grouping of an arbitrary set of messages in the characteristic space of linguistic features based on the PR-effect index.

References

1. Zrazhevskaya, N.I., Mogilko, S.V.: The technique and methods of manipulliations in Internet publications (for example, Internet-newspapers "Press-Center", "Antenna"). Scientific Papers of the Institute of Journalism, T. 31, April - June, p.118-122 (2008).
2. Ryaby, M., Khatyan, O., Bagatsky, S.: The method of revealing PR-impact through the Internet media. Information security, T.21, No. 3, p. 294-300 (2015).
3. Ivanov, O.V.: Quantitative analysis of the text or the production of numeric artifacts: audit of content analytical research. Sociological sciences, Vol. 148, p. 11-15 (2013).
4. Fedorenko, R.M.: Content-monitoring of the information space as a factor in providing information to the state in the military sphere. Modern protection of information №2, p. 21-25 (2015).
5. Media research and reputation analysis of NOKs FISHES company. February 9, 2016 http://www.slideshare.net/mark_kanarsky/nok-presentation-new-03
6. Khatyan, O.A.: The algorithm for building the "day" of the Internet media. Information security of man, society, state, No. 2 (18), p. 110-123 (2015).
7. Prokhorov, A.M.: Large Encyclopedic Dictionary; St. Petersburg: No-rint, 1452 p. (1999)
8. Sturges, H.A.: The choice of a class interval. JASA. v.21. p. 65-66 (1926).
9. Koichubekov B.K. Biostatistics : training, Almaty: Evero, 154 p. (2014).
10. Tikhomirov, A., Kinash, N., Gnatyuk, S., Trufanov, A. et al: Network Society: Aggregate Topological Models, Communications in Computer and Information Science. Verlag: Springer International Publ, vol. 487, pp. 415-421 (2014).
11. Danik Yu., Hryschuk R., Gnatyuk S.: Synergistic effects of information and cybernetic interaction in civil aviation, Aviation, vol. 20, №3, pp. 137-144 (2016).