

<https://doi.org/10.5281/zenodo.17534427>

Волошин В'ячеслав,

доктор технічних наук, професор,

Приазовський державний технічний університет,

ORCID: 0009-0005-6809-6779, e-mail: vsvlshn52@gmail.com

МЕТОДИКА АНАЛІЗУ ДОСТОВІРНОСТІ ІНФОРМАЦІЇ В МЕРЕЖІ ІНТЕРНЕТ

АНОТАЦІЯ. Робота присвячена методам розпізнавання неправдивої та недостовірної інформації в Інтернеті шляхом її формальної оцінки з розподілом за показниками структурної когерентності, банальності інформації, ступеня її випадковості та пов'язаної з цим ентропії інформаційної системи, як формалізованого показника її об'єктивності. Запропоновані три критерії у порівнянні з ентропією Шеннона для текстової інформації дозволяють за допомогою розробленого алгоритму оцінити рівень достовірності або хибності наданої інформації в першому наближенні. На прикладах реальної текстової інформації, витягнутої з Інтернету, показано можливості та ефективність запропонованої методології виявлення недостовірної інформації. Такі методи можуть бути корисними, зокрема, в наукових дослідженнях, при рецензуванні наукових статей та книг, при оцінці соціальної інформації на предмет її об'єктивності.

КЛЮЧОВІ СЛОВА: дезінформація, недостовірна інформація, розпізнавання фейків, структурна когерентність, ентропія Шеннона, інформаційна система, текстовий аналіз, верифікація інформації, алгоритм оцінювання, академічна доброчесність, рецензування, інтернет-контент.

ВСТУП. Актуальність роботи підтверджується відомими обсягами недостовірної та вочевидь неправдивої інформації в мережі «Інтернет», що перешкоджає її використанню в наукових цілях і є способом прямої чи непрямої дезінформації суспільства. Тому надійні методи розпізнавання неправдивої інформації можуть бути важливим інструментом як у наукових працях, так і в суспільстві.

МАТЕРІАЛИ ТА МЕТОДИ, що використовуються в роботі, включають обробку первинної та вторинної інформації шляхом її диференціювання за ентропією з використанням методів розрахунку ентропійних показників текстових матеріалів [1].

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ. У роботі запропоновано звернути увагу на такий об'єктивний показник інформації, як її ентропія, міра невизначеності в текстовому повідомленні. Сама по собі ентропія не визначає, правдивою чи хибною є інформація, однак вона впливає на якість інформації, пов'язаної зі структурною передбачуваністю, яка може бути використана як ознака правдоподібності [2]. Ентропія впливає на структурну когерентність інформації [1, 3]. Це також не є абсолютним критерієм істини, але зміни в структурній когерентності можуть бути показником потенційної неправди, особливо в разі порівняння з іншими структурами повідомлення.

Розроблений алгоритм визначення неправдивості інформації в текстових повідомленнях, заснований на відношеннях функцій когерентності $N(H)$, банальності $B(H)$ і хаотичності $C(H)$ до ентропії текстового повідомлення, дозволяє оцінити об'єктивність інформації з певним ступенем достовірності. Алгоритм включає етапи збору статистичних даних по текстовому повідомленню з розбивкою тексту на окремі слова, словосполучення за змістовним значенням із визначенням частоти слів у групах. Розраховується ентропія структури тексту, оцінюються параметри когерентності, банальності й хаотичності текстової структури в порівнянні з їх ентропією, визначається непрямий ступінь достовірності інформації.

Наведено приклади аналізу реальних (текст повідомлення зі стрічки новин українського контент-Оboz.ua за 01.06.2025) та гіпотетичних (варіанти) повідомлень, вихідні дані та результати розрахунку за якими наведені в таблиці 1.

Результати аналізу рівнів достовірності повідомлень, змодельованих за допомогою радар-графіків (рис. 1), на яких зображені всі чотири варіанти наших прикладів, очевидні. Чим більша «площа» фігури на графіку, тим різноманітніша та збалансованіша інформація в повідомленні. Результатом такого аналізу є судження про об'єктивність кожного з текстів.

Таблиця 1. Результати оцінки якості гейміфікованих зображень на базі комбінації методологій байєсівського та фрактального аналізу

Показник	Варіант 1	Варіант 2	Варіант 3	Oboz.ua
Структура груп слів	12×2;5×3; 5×4;2×5	25×2;1×3; 1×4;2×5	25×2;25×3	128×1;34×2;8×3; 7×4;4×5;3×6;1×13
Кількість унікальних слів	59	61	3	185
Ентропія Шеннона (H)	6,14	6,33	5,66	7,228
Когерентність, N(H)	6,64 ($H_{opt}=6,5$; $C=7$)	6,83	6,16	0,6789
Банальність, B(H)	0,00216	0,00178	0,00345	7,5314
Дисперсія імовірностей D	0,60	0,55	0,30	0,0272
Хаотичність повідомлення, C(H)	-3,684	-3,480	-1,698	-0,000149
Інтерпретація достовірності	Помірно правдоподібне, частково шаблонне	Найбільш природнє, збалансоване	Малоймовірне, шаблонно-синтетичне	З імовірн. 90% повідомлення об'єктивне і не фейк

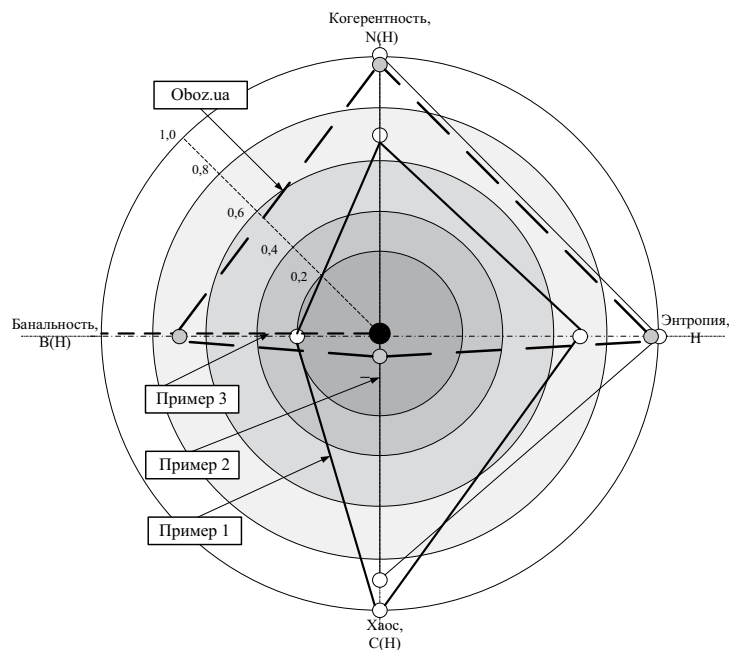


Рис. 1. Радар-графік для оцінки стану формалізованих повідомлень, що складаються з 128 слів у різних групах початкових сигналів (слів), а також одного реального повідомлення

ВИСНОВОК. Запропонована методика оцінки рівнів об'єктивності текстових повідомлень за прямими та опосередкованими показниками, незважаючи на її формалізацію, дозволяє, у першому

наближенні, більш об'єктивно підходити до інформації в мережі «Інтернет», яка тривалий час обходилася без критики та цензури й має стійку тенденцію до розвитку хибності цілих інформаційних потоків. Такі підходи можуть бути корисні, зокрема, в наукових дослідженнях, для рецензування наукових статей і книг.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Gray, R. M. (1990). Entropy and Information Theory. Springer. <https://studfile.net/preview/393707/>
2. Peter Grunwald P., & Vitányi P. (2004). Shannon Information and Kolmogorov Complexity. ArXiv, cs/0410002. <https://doi.org/10.48550/arXiv.cs/0410002>
3. Tishby, N., Pereira, F. C., & Bialek, W. (2020). The Information Bottleneck Method. ArXiv, physics/0004057. <https://doi.org/10.48550/arXiv.physics/0004057>

Voloshyn Vyacheslav,

Professor, D.Sc. in Engineering

Azov State Technical University,

ORCID: 0009-0005-6809-6779, e-mail: vsvlshn52@gmail.com

METHODOLOGY FOR ANALYZING THE RELIABILITY OF INFORMATION ON THE INTERNET

ABSTRACT. The work is devoted to the methods of recognizing false and unreliable information on the Internet by its formal assessment with its distribution according to the indicators of structural coherence, banality of information, the degree of its randomness and the associated entropy of the information system, as a formalized indicator for its objectivity. The proposed three criteria in comparison with Shannon's entropy for textual information allow, with the help of the developed algorithm, to assess the level of reliability or falsity of the information provided in the first approximation. Using examples of real textual information extracted from the Internet, the possibilities and effectiveness of the proposed methodology for identifying unreliable information are shown. Such methods can be useful, in particular, in scientific research, in reviewing scientific articles and books, in assessing social information for its objectivity.

KEYWORDS: disinformation, unreliable information, fake detection, structural coherence, Shannon entropy, information system, text analysis, information verification, evaluation algorithm, academic integrity, peer review, internet content.