**Вісник Тернопільського національного технічного університету**
**https://doi.org/10.33108/visnyk_tntu**
Scientific Journal of the Ternopil National Technical University
*2025, № 4 (120) https://doi.org/10.33108/visnyk_tntu2025.04*
*ISSN 2522-4433. Web:* ***visnyk.tntu.edu.ua***

**UDC 004.82**



# THE SEMANTIC POWER OF TEXT CONTENT AS A FLOW OF A VECTOR FIELD OF EMBEDDINGS

**Viktor Stashkiv; Andrii Khamarchuk; Kyrylo Chornopyskyi; Vladyslav Shumeiko; Maksym Chorniak; Karina Yarosh; Valentyna Tserkovniuk; Oleh Pastukh**

*Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine*

***Abstract.*** *The growing volume of textual data demands advanced methods for evaluating both content effectiveness and semantic structure. While current Natural Language Processing (NLP) techniques offer powerful tools, they often lack metrics for quantifying intrinsic semantic intensity or conceptual coherence. This paper introduces «semantic power» – a novel quantitative measure designed to analyze the conceptual structure and semantic richness of texts, grounded in principles of field theory. The proposed methodology draws on the Ostrogradsky–Gauss theorem and the divergence operator, establishing a theoretical link between local semantic properties of a text (derived from LaBSE vector embeddings) and their global influence. The approach involves computing a semantic centroid, representing the point of highest meaning concentration, and measuring semantic power using a model that assumes an inverse-square decay of vector influence. For further analysis, Gaussian Mixture Model (GMM) clustering id applied, and Principal Component Analysis (PCA) is used for dimensionality reduction and visualization. Experiments on philosophical texts by key Early Modern thinkers – G. W. Leibniz, R. Descartes, and I. Kant – reveal distinct and meaningful variations in semantic power (0.6010, 0.5633, and 0.5787, respectively) and in the resulting clustering patterns (2, 7, and 2 clusters). These findings suggest that semantic power is not merely a numerical descriptor but one that correlates with established intellectual styles and methodological orientations of the authors. As such, semantic power emerges as a powerful and objective metric for assessing the deep cognitive and semantic dimensions of textual content, with potential applications in philology, cognitive science, and computational linguistics and related disciplines.*

***Key words:*** *text analysis, natural language processing, semantic power, vector embeddings, semantic space, divergence, clustering, field theory, large language models, transformers.*

*Submitted 19.08.2025*
*Revised 29.11.2025*
*https://doi.org/10.33108/visnyk_tntu2025.04. 110* *Published 27.01.2026*

## 1. INTRODUCTION

In today's social media landscape – encompassing platforms like Instagram, YouTube, and TikTok – content creators increasingly rely on engagement metrics such as likes, comments, and views to assess performance. These metrics not only indicate how widely content resonates with audiences but also influence monetization algorithms. Accordingly, creators often seek to maximize them to enhance revenue.

This gives rise to an optimization problem, where the objective function depends on variables such as the number of likes, views, and comments. Solving such a problem can offer practical insights into improving content reach and overall visibility.

Among the factors that shape engagement levels, textual data plays a significant role. This paper introduces a novel approach to evaluating the influence of text on content popularity by proposing the concept of semantic power – a measure rooted in field theory, particularly the idea of vector field flux.

The study aims to formalize the semantic power criterion using mathematical constructs from field theory. Specifically, it employs the divergence operator to quantify variations in

vector field flux associated with textual data, offering a new perspective for assessing the textual contribution to engagement outcomes.

## 2. LITERATURE REVIEW AND ANALYSIS

A broad spectrum of mathematical techniques and software tools is currently employed in modern textual data analysis. This field is undergoing rapid development, as evidenced by the growing body of scientific literature dedicated to semantic modeling, text vectorization, and the use of neural networks and statistical algorithms in classification, clustering, and semantic analysis tasks.

Specifically, a text classification method based on graph convolutional networks is proposed, where a single graph is constructed for a given document corpus, capturing token co-occurrence patterns and their associations with documents [1]. The authors construct a single graph for a given document corpus, after which joint training of vector representations for both types of nodes is performed. The method demonstrates high efficiency without the use of external embeddings.

The application of deep learning methods, particularly neural networks and graph structures, for analyzing large-scale corpora of scientific texts is discussed in [2]. The main focus is on constructing document vector representations that consider both semantic content and inter-document relationships. The advantages of such representations in studying scientific communication are also highlighted.

A combined approach using transformers and graph convolutional networks for text classification is investigated in [3]. The model considers word sequence and their parts of speech during graph construction. The incorporation of a transformer enables the model to retain contextual dependencies during classification. Comparative testing was conducted on five datasets.

An approach to semantic modeling based on representing linguistic units as vectors in a complex-valued Hilbert space is proposed in [4]. The method allows for non-linear composition of meanings and is implemented as a neural network for text classification tasks. Its advantages are demonstrated on several open datasets.

A method for automatic detection of semantic discrepancies in parallel texts using a neural architecture for bilingual semantic similarity estimation is presented in [5]. The model does not require manual annotation and demonstrates higher accuracy compared to approaches based on surface features. The obtained results are relevant for improving the quality of machine translation.

A multi-label text classification model based on a graph convolutional network considering semantic features is developed in [6]. A global graph is constructed that includes texts, words, and labels. Pre-training of an encoder for initializing text nodes is also proposed. Particular attention is paid to the model's ability to classify new texts not present in the training set.

An approach to assessing semantic similarity between textual descriptions of manufacturing line failures is explored in [7]. The method combines vector representations, particularly the LaBSE model, with clustering using Gaussian Mixture Models. Principal Component Analysis (PCA) is used for visual analysis of the semantic space structure. It is demonstrated that this approach enables effective grouping of similar incidents and supports timely detection of recurring problems.

Nevertheless, the collective findings presented in the aforementioned studies do not offer a fully comprehensive approach to textual information analysis.

## 3. CONSIDERATION OF THE CONCEPT OF “SEMANTIC POWER”

This research analyzes texts by Gottfried Wilhelm Leibniz, René Descartes, and Immanuel Kant. A central quantitative concept introduced here is «semantic power». Imagine

a semantic space where each point represents a meaning or context. To quantify how a meaning spreads or influences, a vector is assigned to each point, indicating both direction and strength. These vectors are created by first breaking down text into tokens. Then, a transformer-based large language models (LLMs) generates a vector representation (an embedding) for each token, capturing its context and meaning.

The concept of vector field flux is also employed, which is a scalar value that describes the intensity of a vector field passing through a given surface. For a closed surface $S$, the flux ($\varphi$) is calculated by the integral:

$$\varphi = \oiint_S \vec{a} \cdot \vec{n} \cdot dS,$$

where $\vec{a}$ is the vector field, and $\vec{n}$ is the unit normal vector to the surface $S$.

This formula brings us to the idea of divergence, which shows how much the field «spreads out» from each point in space.

A central mathematical tool in this model is the Ostrogradsky–Gauss theorem, which links the internal behaviour of a vector field (its divergence) to its external effect (the total flux through the surface):

$$\oiint_S \vec{a} \cdot \vec{n} \cdot dS = \iiint_V div\vec{a} \cdot dV,$$

where $div\vec{a}$ is the divergence of the vector field $\vec{a}$, and $V$ is the volume enclosed by surface $S$.

In the context of embedding vectors, this formula is interpreted as the relationship between the «semantic power» emanating from a text segment and the total divergence of the corresponding vector field within a closed surface $S$. This approach is widely used in physics, particularly in describing the electric field of a charged sphere, where the flux of the electric field strength vector through a spherical surface is proportional to the total charge inside that sphere.

This research extrapolates the presented physical principle to the space of semantic embedding vectors. Such a transfer allows for the interpretation of local vector variations in terms of their global impact on the semantic structure of the text, utilizing the mathematical apparatus of field theory.

Following the extrapolation of vector field theory's conceptual foundations to the space of semantic embedding vectors, this work draws an analogy between the «semantic power» of textual elements and the intensity of a physical vector field. Specifically, just as in an electrostatic field where force diminishes with distance from the source according to an inverse-square law, this model posits that the influence of an individual embedding vector decreases as its distance from the centroid increases.

To unify the scale, all vectors $\vec{v}^2$ are first normalized to unit length. Subsequently, their mean vector – the centroid $\vec{c}^2$ is computed, serving as the analog to the field source in classical physical models.

Building on this framework, a formula is introduced to calculate semantic power as the cumulative contribution of all embedding vectors within a given spherical region of radius $r$:

$$SP = \sum_i \frac{\|v_i - c\|}{4\pi r^2}.$$

In this context, the semantic power of an individual embedding vector is determined as a function of its Euclidean distance from the centroid, modeling an inverse-square decay. This

allows for the quantitative assessment of a particular word or phrase's «centrality» or representativeness within the author's semantic field. Specifically, the closer a vector is to the centroid, the higher its semantic significance within the text under consideration.

## 4. TEXT DATA PROCESSING ALGORITHM

In order to quantitatively characterize semantic power, software was developed for assessing and visualizing the semantic structure of textual data. This tool integrates natural language processing (NLP) techniques for text preparation, advanced vector representation models to illustrate semantic relationships, and machine learning algorithms for identifying semantic clusters. A core component of this algorithm is the computation of semantic power, which quantifies the intensity of a text's meaning.

The initial stage of the algorithm involves text preprocessing using the spaCy language model. Similar to approaches found in [8] and [9], this step performs lemmatization, converting lexemes to their base forms. Concurrently, elements lacking semantic content – such as prepositions, conjunctions, articles, and punctuation – are removed. This process yields a refined text, containing only lexemes crucial for subsequent semantic analysis.

Following preprocessing, the algorithm proceeds to vector space transformation. Using the Language-Agnostic BERT Sentence Embedding (LaBSE) model, consistent with methodologies in [10, 11], each word is assigned a multidimensional vector. This vector representation captures the word's contextual meaning. A critical aspect at this stage is the removal of the model's service tokens to prevent artifacts that could distort the semantic field's structure.

Next, the obtained embedding vectors are normalized, which allows us to move from the analysis of absolute values to the analysis of their orientations in space.

Using the NumPy library, the centroid $\vec{c}^2$ is calculated based on the normalized vectors as their average. In this model, the centroid functions as a source of semantic charge, representing the conceptual core of the text – the central vector around which all contextual meanings are grouped. It serves as the starting point for measuring the level of semantic deviation of individual elements.

The next stage includes measuring the distances to the centroid and making an integral assessment. For each vector, the Euclidean distance to the center is calculated, which fixes the degree of its semantic distance from the core of the text. After that, the conditional radius of the spherical area around the centroid is determined, and all distances are aggregated with subsequent normalization through the area of the spherical surface. This approach allows us to obtain a scalar value of the semantic power – a value that reflects the intensity of semantic ‘radiation’ outside the conceptual center.

In order to identify the internal structure of the semantic space, the vectors are clustered – similar to the approaches described in [12; 13] – using the Gaussian Mixture Model. To ensure high-quality grouping of vectors, it is necessary to determine the optimal number of model components. For this purpose, the Silhouette Score is calculated, providing a metric for evaluating the cohesion of elements within a cluster and their separation from other clusters. Figure 1 shows a graph of the dependence of the Silhouette Score on the number of clusters for the text of Rene Descartes.

Analyzing this graph helps us choose the optimal number of clusters, striking the best balance between internal homogeneity and inter-cluster difference. As a result, the field breaks down into localized clusters of values – sub-regions, each interpretable as a distinct semantic zone of influence.

In the final stage, the vectors' dimensionality is reduced to three dimensions using Principal Component Analysis (PCA). This method allows us to retain the most significant components of variation in the data while transitioning to a lower-dimensional space. PCA is

widely applied in various fields, including the spectral analysis of radio signals for extracting informative features [14].

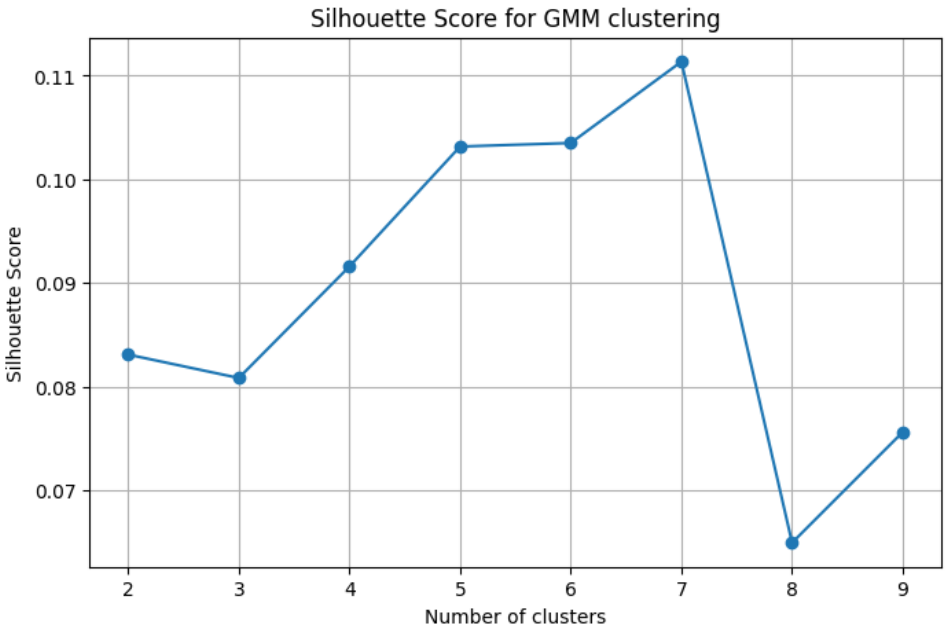


**Figure 1.** Dependence of Silhouette Score on the number of clusters in the Gaussian Mixture model

However, analogous to [15], within this study, PCA is used to visualize the structure of the semantic space: it enables observing the vectors' positions, their clustering around the centroid, cluster formation, and the overall spatial organization of the field within a three-dimensional plane.

The proposed algorithm combines deep natural language processing methods with a geometric interpretation of meaning. This allows us not only to quantitatively measure the semantic power of text but also to explore the spatial organization of meanings.

## 5. RESULTS OBTAINED AND THEIR ANALYSIS

To test the practical use of the proposed algorithm, an experimental analysis of the texts of three classics of Western philosophy was carried out: Gottfried Wilhelm Leibniz, René Descartes and Immanuel Kant. The analysis was based on the English versions of their texts.

Gottfried Wilhelm Leibniz's Letter to Foucher (1675) was the first text analyzed of Leibniz's perspectives on logic, truth, and metaphysics. The analysis revealed a calculated semantic power of 0.6010 for this text, marking it as the highest value among all texts examined in this study. In the realm of vector representations, this high semantic power manifested as the formation of only two compact clusters. This distinct clustering indicates a highly clear and concentrated conceptual organization. The text's content predominantly centers on a limited number of core concepts, which form closely interconnected semantic structures. A visual representation of this is provided in Figure 2, illustrating a cluster structure with a dense core and clearly defined boundaries.

The next fragment analyzed is Rene Descartes' Rule 7 from Regulae ad directionem ingenii, which describes the principle that complex ideas should be decomposed into simpler elements and that thinking should follow a clearly defined sequence. In contrast to Leibniz's letter, the analysis of Descartes' work revealed a completely different picture. Despite the lower

level of semantic power (0.5633), the vector structure of his text formed seven separate clusters, the largest number among the texts studied.

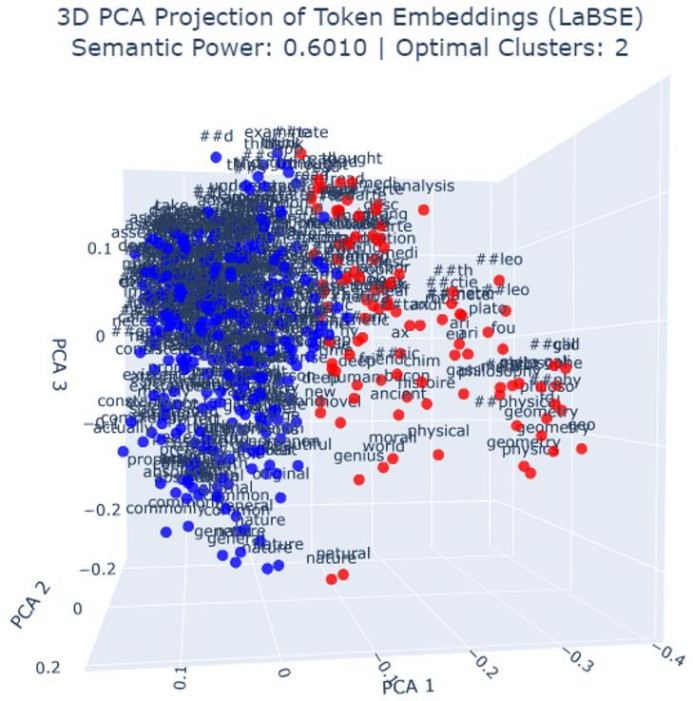


**Figure 2.** Visualization of clusters for Gottfried Wilhelm Leibniz (Semantic Power = 0.6010, clusters = 2)

This distribution demonstrates a distinct fragmentation of meanings, which are split into separate ideas and concepts, each of which has relative autonomy. The visualization shows the multiplicity of small clusters of vectors with minimal overlap between them (see Fig. 3).

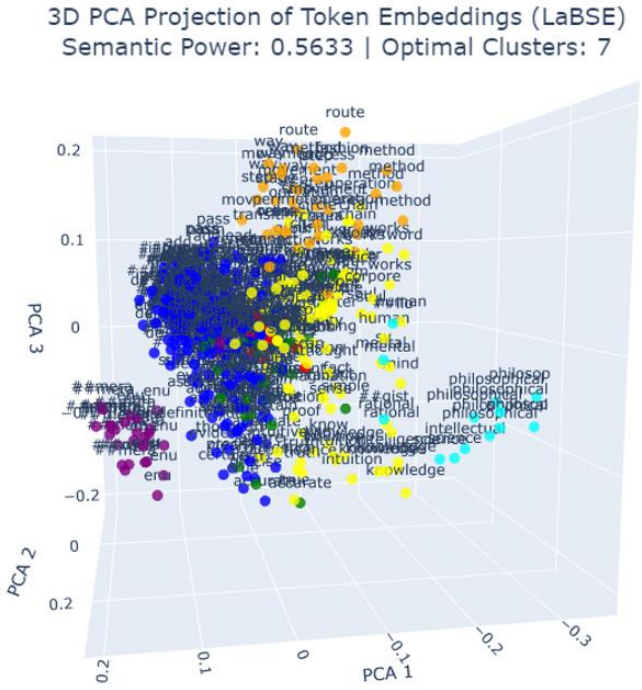


**Figure 3.** Visualization of clusters for René Descartes (Semantic Power = 0.5633, Clusters = 7)

The last fragment belongs to Immanuel Kant; it is from Critique of Pure Reason, chapter «The Transcendental Doctrine of Elements» (First Part). This fragment is the central part in which Kant outlines the foundations of his transcendental method, trying to answer the question: 'What can I know?' The analysis revealed certain structural similarities with the Leibnizian text. A two-cluster organization is also observed, but with a lower density and a lower semantic power value of 0.5787. The spatial visualization shows two clearly separated but internally more diffuse clusters, indicating a wider semantic dispersion, apparently due to the high level of abstraction of the concepts involved in the text. This pattern is shown in Fig. 4, where the internal fragmentation of each of the two main semantic cells can be observed.

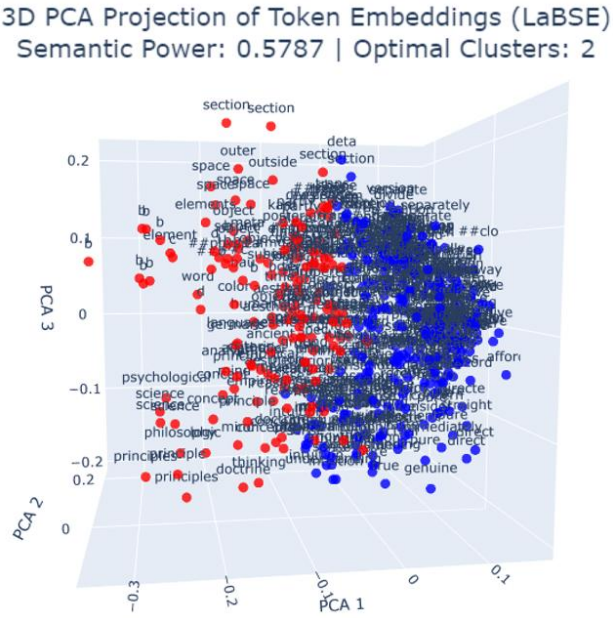


**Figure 4.** Visualization of clusters for Immanuel Kant (Semantic Power = 0.5787, Clusters = 2)

The variability in semantic power observed across the texts of the studied authors directly correlates with their unique intellectual methodologies and the logical-semantic organization of their discourse. The high semantic power found in G. W. Leibniz's texts is attributed to his propensity for systematic thinking, the formalization of concepts, and the use of a limited number of logically interconnected concepts. This contributes to the formation of compact clusters with high internal coherence.

In contrast, the maximum number of clusters observed in René Descartes' text reflects the analytical nature of his style, which relies on decomposing complex ideas into simpler components. This approach leads to a fragmented semantic structure with less centralization.

Meanwhile, the intermediate semantic power found in Immanuel Kant's text is due to his high level of abstraction and use of multidimensional transcendental categories. This results in decreased cluster density and increased semantic diffusion, even while maintaining a two-cluster structure.

Ultimately, the level of semantic power is closely linked to the conceptual specificities of each author's philosophical style. This confirms the relevance of the proposed model for analyzing the deep cognitive characteristics of text.

## 6. CONCLUSIONS

Within the scope of this research, the concept of semantic power was introduced, grounded in vector field theory and specifically applying the Ostrogradsky–Gauss theorem to

relate the local properties of a semantic field (divergence) to its global manifestations (flux). This framework models a semantic space where word embedding vectors–derived from transformer-based large language models (LLMs) – indicate the direction and magnitude of meaning propagation. Their influence, by analogy with physical fields, diminishes with distance from a semantic centroid according to the inverse-square law. Such an approach enables the mathematical formalization and quantitative assessment of a text's semantic load intensity, as well as the degree of centrality or representativeness of individual words and phrases within an author's semantic field.

To implement this concept in practice, a dedicated algorithm and software were developed, integrating sequential stages of text data processing. These stages include preprocessing with the spaCy language model for lemmatization and filtering out irrelevant elements; converting cleaned tokens into vector representations using the LaBSE model, followed by the removal of service tokens; normalizing the resulting embeddings; and computing the semantic centroid as a representation of the text's conceptual core. This implementation provides a reproducible method for quantifying semantic power, translating theoretical foundations into a functional tool for analyzing the deep semantic properties of text.

The algorithm further computes semantic power as an aggregated measure based on the Euclidean distances of individual vectors from the centroid, normalized by the area of a hypothetical spherical surface. It also incorporates clustering of embedding vectors using Gaussian Mixture Models, with the optimal number of clusters determined by the Silhouette Score, and visualizes the semantic space via Principal Component Analysis (PCA). These methods not only enable a scalar evaluation of a text's overall semantic intensity but also reveal its internal structure, identify thematic clusters, and offer an intuitive visualization of the spatial organization of meaning – thus enriching the interpretive understanding of value distribution within the text.

Experimental validation of the proposed method was carried out on philosophical texts by Gottfried Wilhelm Leibniz, René Descartes, and Immanuel Kant. The results quantitatively captured the semantic power and structural features of each work: Leibniz's text exhibited the highest semantic power (0.6010) with two compact clusters; Descartes' text the lowest (0.5633) with seven dispersed clusters; and Kant's text an intermediate value (0.5787) with two less dense clusters. These findings demonstrated a clear correlation with known features of each thinker's intellectual methodology and logico-semantic organization, thereby confirming the sensitivity and relevance of the proposed model in analyzing the cognitive depth of textual content.

In conclusion, this research successfully extrapolates the principles of vector field theory to the analysis of semantic spaces in text, resulting in a novel methodology for the quantitative determination and interpretation of semantic power and its associated cluster structure. The proposed approach opens promising directions for objective analysis of semantic richness and textual organization, authorial stylistic profiling, and comparative discourse studies, offering significant potential for application in philology, cognitive science, computational linguistics, and related disciplines.

**References**

1. Yao L., Mao C., & Luo Y. (2019) Graph convolutional networks for text classification. Proceedings of the AAAI Conference on Artificial Intelligence, 33, 7370–7377. https://doi.org/10.1609/aaai.v33i01.33017370
2. Kozlowski D., Dusdal J., Pang J., & Zilian A. (2021). Semantic and relational spaces in science of science: Deep learning models for article vectorisation. Scientometrics. https://doi.org/10.1007/s11192-021-03984-1
3. Liu B., Guan W., Yang C., Fang Z., & Lu Z. (2023) Transformer and graph convolutional network for text classification. International Journal of Computational Intelligence Systems, 16 (1). https://doi.org/10.1007/s44196-023-00337-z

4. Wang B., Li Q., Melucci M., & Song D. (2019). Semantic hilbert space for text representation learning. У The world wide web conference. ACM Press. https://doi.org/10.1145/3308558.3313516
5. Vyas Y., Niu X., & Carpuat M. (2018). Identifying semantic divergences in parallel text without annotations. У Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 1 (long papers). Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1136
6. Zeng D., Zha E., Kuang J., & Shen Y. (2024) Multi-label text classification based on semantic-sensitive graph convolutional network. Knowledge-Based Systems, 284, 111303. https://doi.org/10.1016/j.knosys.2023.111303
7. Tekgöz H., İlhan Omurca S., Koç K. Y., Topçu U., & Çelik O. (2022). Semantic similarity comparison between production line failures for predictive maintenance. Advances in Artificial Intelligence Research. https://doi.org/10.54569/aair.1142568
8. Premalatha M., Viswanathan V., & Čepová L. (2022) Application of semantic analysis and LSTM-GRU in developing a personalized course recommendation system. Applied Sciences, 12 (21), 10792. https://doi.org/10.3390/app122110792
9. Narendra G. O. & Hashwanth S. (2022) Named entity recognition based resume parser and summarizer. International Journal of Advanced Research in Science, Communication and Technology, 728–735. https://doi.org/10.48175/IJARSCT-3029
10. Venkatesh D., & Raman S. (2024). BITS pilani at semeval-2024 task 1: Using text-embedding-3-large and labse embeddings for semantic textual relatedness. У Proceedings of the 18th international workshop on semantic evaluation (semeval-2024). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.semeval-1.124
11. Feng F., Yang Y., Cer D., Arivazhagan N., & Wang W. (2022) Language-agnostic BERT sentence embedding. У Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: Long papers). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.62
12. Kesiraju S., Plchot O., Burget L., & Gangashetty S. V. (2020) Learning document embeddings along with their uncertainties. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 2319–2332. https://doi.org/10.1109/TASLP.2020.3012062
13. Hu C., Wu T., Liu S., Liu C., Ma T., & Yang F. (2024) Joint unsupervised contrastive learning and robust GMM for text clustering. Information Processing & Management, 61 (1), 103529. https://doi.org/10.1016/j.ipm.2023.103529
14. Chesanovsky I., & Levhunets D. (2017). Representation of narrow-band radio signals with angular modulation in trunked radio systems using the principal component analysis. Scientific Journal of the Ternopil National Technical University, 86 (2), 117–121. https://elartu.tntu.edu.ua/handle/lib/22368
15. Musil T. (2019). Examining structure of word embeddings with PCA. У Text, speech, and dialogue. Springer International Publishing. https://doi.org/10.1007/978-3-030-27947-9_18

**УДК 004.82**

# СЕМАНТИЧНА СИЛА ТЕКСТОВОГО КОНТЕНТУ ЯК ПОТІК ПОЛЯ ВЕКТОРІВ-ЕМБЕДИНГІВ

**Віктор Сташків; Андрій Хамарчук; Кирило Чорнописький; Владислав Шумейко; Максим Чорняк; Каріна Ярош; Валентина Церковнюк; Олег Пастух**

*Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль, Україна*

***Резюме.*** *Зростаючий обсяг текстової інформації вимагає передових методів оцінювання ефективності контенту та його семантичної структури. Існуючі техніки опрацювання природної мови (NLP) часто не надають метрик для вимірювання внутрішньої «семантичної інтенсивності» або концептуальної узгодженості. Ця стаття представляє «семантичну силу» – нову кількісну характеристику, розроблену для аналізу концептуальної структури та смислової насиченості текстів на основі принципів теорії поля. Методологія базується на теоремі Остроградського-Гауса та операторі дивергенції, встановлюючи зв'язок між локальними семантичними властивостями тексту (на основі векторних ембедингів LaBSE) та їхнім глобальним впливом. Підхід включає обчислення семантичного центроїда як точки найбільшої концентрації смислу та кількісну оцінку семантичної сили за допомогою*

*моделі, що враховує обернено-квадратичний спад впливу векторів. Для подальшого аналізу застосовуються кластеризація методом Gaussian Mixture Models та візуалізація за допомогою методу головних компонент (PCA). Експерименти, проведені на філософських текстах видатних мислителів Нового часу, таких як Готфрід Вільгельм Лейбніц, Рене Декарт та Іммануїл Кант, продемонстрували чіткі та значущі відмінності у значеннях семантичної сили (0.6010, 0.5633 та 0.5787 відповідно) та у сформованих патернах кластеризації (2, 7 та 2 кластери). Результати показують, що ці показники не лише є числовими характеристиками, а й корелюють з відомими особливостями інтелектуального стилю та методології кожного з авторів. Таким чином, «семантична сила» виступає як потужний і об'єктивний інструмент для оцінювання глибинних когнітивних та семантичних характеристик тексту, відкриваючи потенційні можливості для широкого спектру застосувань у філології, когнітивістиці, комп'ютерній лінгвістиці та інших суміжних галузях.*

***Ключові слова:*** *текстовий аналіз, опрацювання природної мови, семантична сила, вектори-ембединги, семантичний простір, дивергенція, кластеризація, теорія поля, великі мовні моделі, трансформери.*