Scientific Visualization, 2025, volume 17, number 4, pages 87 - 98, DOI: 10.26583/sv.17.4.09

This paper is an extended version of a contribution presented
at the Graphicon 2025 conference.

Information density of objects in digital environment: theoretical foundations

N.A. Bondareva[1], A.E. Bondarev[2], S.V. Andreev[3], I.G. Ryzhova[4]

Keldysh Institute of Applied Mathematics RAS, Moscow, Russia

[1] ORCID: 0000-0002-7586-903X, nicibond9991@gmail.com
[2] ORCID: 0000-0003-3681-5212, bond@keldysh.ru
[3] ORCID: 0000-0001-8029-1124, esa@keldysh.ru
[4] ORCID: 0000-0003-1613-3038, ryzhova@gin.keldysh.ru

Abstract

The paper presents a theoretical concept for evaluating information density of objects in digital environment. An analysis of limitations of existing methods for quantitative assessment of information space, based predominantly on simple data volume metrics, has been conducted. The concept of an object's information field is proposed as an aggregate of all informational units containing mentions of the studied object in digital space.

The prospects for applying this methodology in information security, social media analysis, and data preparation for neural network training are examined. The proposed approach opens new opportunities for comprehensive evaluation of information resources and may find application in search engines, recommendation algorithms, and big data analysis systems.

Keywords: information density, digital environment, information space, data analysis, information analytics.

1. Introduction

The modern digital space is characterized by exponential growth in information volumes and a qualitative increase in the complexity of information flows. According to various estimates, over two and a half quintillion bytes of data are generated daily in the global information space [1], a significant portion of which represents unstructured information in the form of text content, multimedia materials, and metadata structures. Given this information abundance, researchers and practitioners are faced with the fundamental question of developing adequate methods for assessing the information completeness and qualitative characteristics of data available for real-world objects of interest.

Traditional approaches to the quantitative assessment of information resources, based on simple metrics of data volume, frequency of mentions, or statistical indicators of occurrence, demonstrate significant limitations when addressing the complex characterization of the information value of aggregated data. The existing arsenal of bibliometric and scientometric analysis methods [2], despite the developed mathematical apparatus for assessing the scientific impact of publications and citations, does not provide the tools for a comprehensive assessment of the information environment of arbitrary objects beyond the scientific and academic sphere. Similarly, the methodological arsenal of social network analysis, focused on assessing the popularity and influence of nodes in network structures, does not take into account the qualitative parameters of information content and does not address the problem of assessing the uniqueness of information units.

This issue is particularly pressing in the context of the rapid development of artificial intelligence systems and machine learning technologies. The performance of modern neural network architectures is determined by the qualitative characteristics of training data, including not only its quantitative completeness but also its informational reliability, representativeness, and absence of redundancy. However, currently dominant approaches to training dataset formation focus primarily on quantitative aspects of sample populations, such as sample size and the balanced representation of different object classes, neglecting the qualitative assessment of the information richness of the data relative to the entities under study.

Similar methodological challenges arise in related areas of information technology, including information retrieval and knowledge extraction systems, recommender systems, automatic natural language processing technologies, and the construction of semantic knowledge graphs. These systems, whose functional purpose is to aggregate, structure, and intelligently analyze information about real-world objects of various natures, including organizational structures, events, and abstract concepts, require the development of tools for adequately assessing the completeness and qualitative characteristics of available information.

Research works in the subject area of entity resolution and knowledge graph construction develop methodological approaches to linking and aggregating heterogeneous information about research objects [3-4], but focus mainly on the technical aspects of integrating heterogeneous data, without offering solutions for the problem of assessing the information density and qualitative characteristics of the resulting object representations.

Scientific research in the field of information retrieval and extraction has developed a sophisticated mathematical apparatus for assessing the relevance of documents to user queries and information needs, including classical metrics of search accuracy and completeness. However, existing methods do not solve the conceptually distinct problem of comprehensively assessing the information environment of an object as an integral phenomenon of digital space.

This paper proposes a conceptual model for analyzing information space based on the concept of the density of an object's information field. An object's information field is defined as the totality of all information units containing references, descriptions, or links to the object under study and available in open digital sources. A fundamental characteristic of such an information field is its density, which is an integral measure reflecting not only the quantitative parameters of available information but also its essential qualitative characteristics, including the uniqueness of the information content, its relevance to the object of study, the authority of information sources, and the temporal relevance of the presented data.

The primary objective of this study is to develop the theoretical foundations of the concept of information field density and formulate a methodological framework for its practical application in the analysis and comprehensive assessment of digital information resources. Achieving this goal requires addressing a set of interrelated research tasks, including formulating rigorous definitions of the basic concepts of an object's information field, information field density, and information quantum, and developing a mathematical model for calculating information field density, taking into account the multidimensional characteristics of information content.

2. Analysis of existing approaches to assessing information resources

Modern scientific literature contains a wide range of studies devoted to various aspects of the assessment and analysis of information resources; however, existing approaches demonstrate fragmentation and limited applicability to the task of comprehensively assessing the information environment of arbitrary real-world objects.

The fundamental theoretical foundations of quantitative information assessment were laid in Claude Shannon's classic works on mathematical information theory, where the information content of a message is determined through the entropic characteristics of the source and transmission channel. Shannon's concept of information entropy [5] provides a mathematical framework for measuring the uncertainty and information capacity of systems. However, this approach operates exclusively with the syntactic characteristics of information, ignoring the semantic content of messages and their pragmatic value for solving specific problems. The development of this approach in Andrey Kolmogorov's works on algorithmic information theory [6] introduces the concept of string complexity as the length of the shortest program capable of reproducing it, which allows us to formalize intuitive notions about the meaningfulness and non-randomness of information sequences.

A significant body of modern research is focused on the problem of assessing data quality [7], considered as a multidimensional characteristic of information resources. Conceptual models of data quality, developed in the works of researchers in this area, include such fundamental dimensions as the accuracy of information content, reflecting the correspondence of data to the actual state of the described objects, the completeness of representation, characterizing the degree of coverage of relevant aspects of the subject area, data consistency, determining the absence of internal contradictions in information structures, the relevance of information relative to the temporal context of their use, and the relevance of information in relation to the assigned analytical tasks. These studies propose a variety of metrics and algorithmic approaches to the automatic assessment of the qualitative characteristics of data, including statistical methods for identifying anomalies and inconsistencies, techniques for comparative analysis of multiple sources to verify actual accuracy, and heuristic algorithms for assessing completeness based on the structural features of subject areas. However, existing data quality studies demonstrate a methodological focus primarily on structured information resources, such as relational databases and formalized catalogs, without offering adequate solutions for assessing the qualitative characteristics of unstructured information that naturally aggregates around objects in open digital space.

Entity resolution and semantic knowledge graph construction methods aimed at integrating and structuring heterogeneous information about real-world objects have attracted significant research attention in recent decades. The task of entity resolution, which involves identifying and linking different mentions of a single object across heterogeneous information sources, is solved through a combination of string representation comparison methods, contextual analysis of mentions, and machine learning on labeled data corpora. Modern approaches to knowledge graph construction, such as those implemented in Google projects, Knowledge Graph,DBpedia,and Wikidata [8-10] demonstrate impressive results in aggregating structured information about millions of entities of various types, providing unified interfaces for accessing integrated knowledge. However, existing research in this area focuses primarily on the technical aspects of integrating and verifying structured data, without offering methodological solutions for assessing the information density and qualitative characteristics of unstructured information environments of objects, which cannot be adequately represented in the format of structured graph relations.

Classical research in information retrieval and extraction has developed a sophisticated mathematical framework for evaluating the effectiveness of search engines and the relevance of documents to user queries. Fundamental metrics of precision and recall, as well as their harmonic mean in the form of the F-measure, provide quantitative tools for assessing the performance of information retrieval systems under controlled experimental conditions. Developments in this field have led to the creation of complex relevance models that take into account multiple factors, including the textual similarity between the query and the document, the authority of information sources, the temporal characteristics of documents, user behavior, and the contextual features of information needs. Modern approaches to document ranking in web search, based on algorithms such as PageRank [11] and its modifications, demonstrate the effectiveness of taking into account structural characteristics of the information space, such as hyperlink and citation patterns, to assess the authority and significance of information resources. However, existing methods of information retrieval solve a conceptually different problem of assessing the compliance of individual documents with specific information requests, without offering approaches to a comprehensive assessment of the information environment of objects as holistic phenomena of digital space, characterized by specific patterns of information aggregation and qualitative diversity of sources.

Bibliometric and scientometric studies have developed sophisticated methods for assessing the scientific impact and significance of publications, based on the analysis of citation patterns and collaborative relationships between researchers. Classic indicators, such as the h-index and its many modifications, provide comprehensive assessments of the productivity and influence of scientists, taking into account both quantitative characteristics of publication activity and qualitative parameters reflected in the frequency of citations of works by the scientific community. Modern approaches to scientometric analysis include network methods for studying scientific collaborations,temporal analysis of the evolution of research areas, and interdisciplinary metrics reflecting the broad impact of scientific results. Despite the methodological sophistication of bibliometric approaches, their applicability is limited to the specific context of scientific and academic activity and does not extend to assessing the information characteristics of arbitrary objects unrelated to formal scientific citation systems and peer review.

Research in the field of the Semantic Web and ontology engineering offers formalized approaches to representing and structuring knowledge about subject areas, based on logical formalisms and standardized resource description languages. Ontological models provide expressive means for specifying the conceptual structures of subject areas, including hierarchies of object classes, properties and relationships between entities, integrity constraints, and inference rules [12–13]. Linked data technologies developed within the framework of the Semantic Web initiative demonstrate the practical effectiveness of integrating distributed information resources through standardized protocols and data representation formats. Methods for assessing the quality of ontologies include analysis of the logical consistency of conceptual models, assessment of the completeness of subject area coverage, and metrics of the correspondence of ontological structures to the real characteristics of the phenomena being described. At the same time, ontological approaches presuppose the presence of expertly developed conceptual models and do not solve the problem of assessing the qualitative characteristics of spontaneously formed unstructured information, which cannot be adequately described within the framework of pre-defined ontological schemes.

An analysis of existing research areas reveals a fundamental gap in the methodological arsenal of modern information technologies, stemming from the absence of conceptual approaches to the comprehensive assessment of information fields naturally forming around arbitrary objects in unstructured digital space. Existing methods, despite their development within specialized subject areas, do not provide adequate tools for the integrated assessment of the qualitative characteristics of heterogeneous information aggregated around objects of various natures in the information abundance of the modern digital space. This methodological vacuum necessitates the development of new conceptual approaches capable of providing a theoretical foundation for the comprehensive assessment of information density and the qualitative characteristics of object-oriented information fields.

3. Theoretical foundations of the concept of information field density

Developing a conceptual framework for analyzing the information fields of objects requires the formulation of new theoretical constructs that would overcome the limitations of existing approaches to assessing information resources. The proposed theoretical model is based on a synthesis of classical information theory, modern concepts of semantic space, and the principles of systems analysis of complex information structures. The central premise of this approach is the understanding that information about real-world objects does not exist in digital space as isolated discrete units, but rather forms complex interconnected structures that can be conceptualized as information fields with specific topological and qualitative characteristics.

The concept of an object's information field is based on the metaphorical transfer of physical concepts of field structures to the analysis of information phenomena. Just as physical fields are characterized by the distribution of energy or matter in space, an object's information field represents the distribution of information units in digital space, where each information unit possesses a certain "mass" or significance, and the aggregate of such units forms a complex topological structure with various zones of information concentration and sparseness. This conceptualization allows for the application to the analysis of information structures of the mathematical apparatus developed for the study of field phenomena, including the concepts of density, gradient, flow, and other characteristics describing the spatial distribution of physical quantities.

A formal definition of the information field of an object can be formulated as follows: the information field of an object O is the set of all information units $I = \{i_1, i_2, \ldots, i_n\}$ available in the digital space and containing direct or indirect references, descriptions, links, or any other forms of informational connection with this object. Mathematically, this can be expressed as $IF(O) = \{i \in I \mid R(i, O) > \theta\}$, where $R(i, O)$ represents the relevance function of the information unit i with respect to the object O, and θ is the threshold value determining the minimum degree of relevance for including the information unit in the object's field. It is important to note that the boundaries of the information field are blurry, since the relevance of information can vary widely, from direct references to an object to complex contextual associations, the establishment of which requires deep semantic analysis.

An object's information field is characterized by a number of fundamental properties that determine its structure and dynamic characteristics. Spatial heterogeneity of the field manifests itself in the fact that different areas of digital space contain an uneven distribution of information about the object. Some sources and platforms may accumulate significant volumes of relevant information, while other segments of the information space contain virtually no mention of the object. Temporal dynamism is another key property of information fields, as they continuously evolve under the influence of new information sources, updates to existing materials, changes in the object's popularity, and other factors influencing information activity around it. The multidimensionality of the information field is due to the fact that information about an object can exist in various formats and at various semantic levels, including factual data, analytical materials, multimedia content, meta-information, and contextual relationships.

The central characteristic of an information field is its density, which represents an integral measure of the concentration of qualitative information about an object within a unit of information space. Unlike simple quantitative metrics such as total data volume or frequency of mentions, information field density must take into account the qualitative parameters of information content, including the uniqueness of the information, its relevance to the object of study, the authority of sources, and the temporal relevance of the data. Conceptually, information field density can be understood as a measure of the information "mass" of an object in digital space, reflecting not only the amount of available information but also its ability to form a complete and accurate representation of the object of study. More generally, information field density characterizes the ability of a technology, such as artificial intelligence, to recreate an image of an object based on collected data [14], established patterns, and relationships.

It's worth noting that this concept represents an attempt to move away from simply quantifying information in bits or the number of tags, links, or mentions. Given the existence of modern language models, information acquires particular value when it can be interpreted and searched using not just words, but semantic constructs and contexts. In terms of quantitative measurement of information, the traditional assumption is that the more data available, the more accurately and completely a digital image of an object can be recreated. However, the heterogeneity of information concerning the same object of study, across different interpretations, makes the idea of creating a neutral model whose answer satisfies all parties virtually impossible. As the NewsGuard report [15-16] states, the growth of information coverage and the connection of web search to language models and chatbots has resulted in a noticeable deterioration in the accuracy of the results returned, and during events taking place “online”, chats more often reinforced false narratives, pulling in materials from dubious sources and making no distinction between authoritative publications and their propaganda lookalikes.

At the same time, an integrated assessment of the density of information, both true and false, surrounding an object can allow for an analysis of the volume of sources a model can potentially rely on, as well as what image of the object it is most likely to form, what patterns it will utilize, and what opinions it will prioritize. Given that, in light of the development of language models across political poles, the market will likely push AI services toward a more explicit position in an attempt to meet the expectations of their audiences, there is a need to create neutral and objective criteria for assessing information volumes in order to interpret the main trends in the development of digital images of certain phenomena that language models will focus on. In the near future, the development of such services may lead to different language models, limited by fundamentally different principles, providing diametrically opposed answers to the same question. In light of this, assembling a relatively objective information picture will become increasingly difficult with each passing year.

The density of an object's information field can be characterized by the number of conventional units (a conventional unit is a quantum of unique, word-for-word, non-repeating information containing a thought/analysis/research/fact/conclusion related to the object in question). Then, say, 1,000 generated thoughts related to the object will increase the density of its information field, while, for example, 1,000 reposts (1,000 mentions of the same term without the slightest change in wording) will not change the density level. In this case, the quantity of information in the traditional sense of computer science, i.e., measured in bits, may be only one parameter, but not the most significant one.

In this regard, we can introduce a name for such a conventional unit. Infon (from " information " + the suffix " on ") – in the primary definition, these are non-repeating units of information containing original thoughts, analysis, research, or facts about an object .

However, each conventional unit (quantum of information) must, in one way or another, have its own "weight" or "significance." For example, a scientific article with original research on a subject and a random comment on a social network, although both contain unique information, have different values for shaping the information field. Thus, "heavier" information units have a stronger influence on the information field. It is also important to consider the time component. Information tends to become outdated, and its significance can change over time. For example, a ten-year-old scientific article may have less weight than a more recent study, unless it is fundamental to the field.

An interesting addition could be the concept of "information resonances." When multiple independent sources confirm the same information in different ways (not simply copying), this can create a reinforcing effect—similar to how waves can reinforce each other during resonance. Such resonances can significantly and sharply increase the density of the information field at certain points in time. Visualizing such temporal fluctuations can significantly improve understanding of the development of a given phenomenon.

Another important aspect is the coherence of such information quanta. Individual fragments of information, linked by logical or cause-and-effect relationships, can form more stable and meaningful structures in the information field than isolated facts.

If we consider the infon from the perspective of these aspects, its definition becomes more precise: an infon is the minimal indivisible unit of unique information about an object, which cannot be reduced without loss of semantic content. An infon is the minimal unique unit of information containing a complete thought (a fact) about an object,which is not a direct repetition of existing information. This model is open to two interpretations: one can view infons as a unit of measurement, in which case they must be identical. This necessitates defining what constitutes the minimal unit of unique information.

Another approach to the model assumes that each infon is defined as a quantum of unique information and can have its own informational weight and relevance (Fig. 1). An infon is characterized by atomicity, meaning the impossibility of its further division without loss of semantic and informational integrity; uniqueness, implying the absence of exact duplicates of a given information unit in other sources; relevance, ensuring a direct connection between the content of the infon and the object of study; and verifiability, allowing for the verification of the factual accuracy of the information contained in the infon.

Figure 1 shows an example of a possible visualization of such a phenomenon, where the represented object is surrounded by heterogeneous particles graded in color and size. The size of each particle reflects the size of the information quantum, while the color demonstrates the degree of conceptual connection with the object, from directly related to the object to very distantly related.

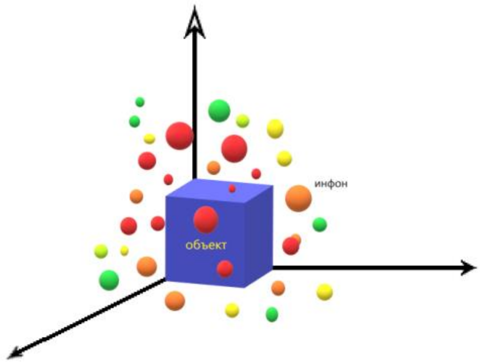

Fig. 1. Example of visualization of the density of the information field of an object

The process of identifying information from a general information field is a complex analytical task that requires a combination of automatic natural language processing methods and expert content analysis. Algorithmic identification can be based on text segmentation methods to identify semantically complete fragments, information novelty analysis through comparison with existing knowledge bases, assessment of actual uniqueness through plagiarism and duplication detection, and determination of relevance using semantic analysis and machine learning methods. Each identified information can be characterized by a set of quantitative parameters, including its uniqueness, its degree of relevance to the object, an assessment of the reliability of the information contained, and an indicator of information value for forming a holistic understanding of the object.

Since information in the modern world is presented in a wide variety of forms (visual, audio, video streams, text, social media comments, and numerical data sets), combining them into a unified structure becomes a distinct challenge. The key to unifying different types of information lies in the concept of a multidimensional information space. An information field can be imagined as a multilayered structure, where each type of data forms its own layer, but all are interconnected and influence each other. This is similar to how different modalities (text, images, sound) can be transformed into a single vector space in neural networks.

The concept of “information embeddings” is suitable for unifying different types of data – the transformation of any type of information into a universal vector representation [17]. Modern technologies already make this possible: CLIP can find connections between text and images, wav2vec converts audio into vectors, and large language models transform text into multidimensional representations. In this case, it is important to consider the "information density coefficient" for different types of data. For example, one second of video may convey more information about an object than a text description of the same duration, but a textual analytical article may contain deeper semantic information than a simple photograph.

The question of "cross-validation" between different types of data also arises. Validation between different types of data. If information from different sources and formats corroborates each other, this increases the credibility of each individual piece of information. For example, if a textual description of an event is supported by video footage and numerical data, the overall credibility of the information increases. In the context of artificial intelligence, this approach opens up new possibilities for creating multimodal systems capable of forming a holistic view of objects based on heterogeneous data. This could prove to be another data organization system useful for the development of general artificial intelligence systems, which must be able to process information holistically, similar to the human brain.

Currently, large generative artificial intelligence models tend to accumulate numerous individual algorithmic rules—specialized patterns for specific cases—that do not integrate into a coherent knowledge system. Such localized patterns often contradict each other, creating internal conflicts in the system's operation. Research attempts to find coherent conceptual representations in the model structure [18] reveal only disparate information fragments that do not form a single, coherent image. Nevertheless, such distributed rules have a certain practical value. The enormous parametric capabilities of language models allow for the storage of such patterns in large quantities, and the quantity often compensates for the lack of a clear structure. The ability to create verifiable internal representations opens avenues for combating artificial hallucinations, increasing the reliability of logical inferences, and ensuring greater transparency in the operation of intelligent systems.

The mathematical formalization of the density of the information field can be represented as a weighted sum of the information contributions of individual elements of the field, where each element (infon) is assessed according to multiple quality criteria. Let the information field (IF – Information The Field object is a multidimensional vector space. The basic formula might look like this:

$$IF = \sum (Q_i \times W_i \times T_i \times R_i) \times K_i$$

where *Qi* is a certain information quantum, *Wi* is a weighting coefficient reflecting the type and significance of the information unit i_i, *T(*i_i*)* takes into account the temporal relevance of the data, *R(*i_i*)* determines the relevance relative to the object, K is a certain normalizing coefficient.

The formula can also be supplemented with parameters such as *A (*i_i*),* which reflects the authority of the information source, and *U(*i_i*)*, which characterizes the uniqueness of the information content.

Each of these coefficients has its own calculation methodology based on objective parameters and existing data analysis methods. It makes sense to seek out and borrow some of these coefficients from existing research on big data. There are a number of related research areas: information field theory in physics [19], semantic spaces in linguistics [20 link to Klyshinsky /elastic maps], digital information ecology, and quantum information theory. These areas provide useful tools and methodologies that can be adapted to develop information field theory.

For example, if we consider Wi (the source weighting factor), the closest analogs would be entities such as the impact factor of scientific journals, Google 's PageRank, and the citation index. A possible calculation formula might be something like this:

$$W_i = (A_s \times C_s \times R_s \times V_s)/N_{max}$$

where *As* (Authority Score) is the authority of the source (0-1), *Cs* (Citation Score) is responsible for the citation index, Rs (Reliability Score) characterizes the reliability indicator based on historical data, *Vs* (Verification Score) is the possibility of verifying information, and Nmax is a certain normalizing maximum

Ti metric can also be used in real-world examples: scientific databases consider the "age" of publications when calculating their importance, and Netflix uses similar time-based factors to rank content. Metrics for assessing information relevance (*Ri)* are used in modern search engines and natural language processing (NLP) systems.

The normalizing coefficient is determined empirically for a specific subject area and can be calculated as:

$$K - 1/max(IF)$$

where max (IF) is the maximum possible value of the information field in a given area.

The interaction between the concepts of the information field, its density, and its constituent infons forms a theoretical model that enables the qualitative analysis of information structures in digital space. Infons act as individual particles of the information field; their totality defines the field's structural characteristics, while their qualitative parameters determine the overall density of an object's information field. This model provides a theoretical foundation for developing practical methods for assessing information resources, enabling a transition from intuitive notions of the "richness" or "poverty" of an object's information to more rigorous quantitative assessments based on an analysis of the qualitative characteristics of its information content.

4. Prospects for the application and development of the theory

The proposed approach to assessing information density in the digital environment opens up new opportunities for practical application and further development. Key areas of potential application and development prospects for the methodology include information security, social media and content analysis, and data preparation for neural network training.

In the field of information security, the methodology for assessing information density can be applied in several key areas. First, analyzing the density of the information field can help identify abnormal spikes in activity that may indicate targeted information campaigns or attacks. Second, assessing the qualitative characteristics of the information space will ultimately help identify sources of unreliable information and monitor the spread of disinformation. Early detection of information threats by analyzing the dynamics of changes in information density may also prove relevant.

In the context of social media analysis, the proposed methodology can provide tools for a deeper understanding of information processes. Assessing information density allows for identifying significant trends and separating them from information noise, which is especially important in the context of content overload on social media. Analyzing the qualitative characteristics of the information field helps determine the real impact of content and its authors, going beyond simple quantitative metrics such as the number of likes or shares.

In light of the development of this theory, the following directions for further research can be identified.

1. The development of mathematical apparatus is one of the key areas. It is necessary to develop more accurate models to describe the interactions of various components of the information field and to create methods for quantitatively assessing the qualitative characteristics of information. Particular attention should be paid to the creation of mathematical models that take into account the temporal dynamics of information processes and the nonlinear nature of the interactions between different types of information.

2. Experimental verification of the methodology requires conducting a series of studies in various subject areas. The effectiveness of the proposed methods for assessing information density must be confirmed using real data, and the results must be validated in various application contexts. An important aspect is the development of standardized experimental methods and criteria for evaluating the results.

3. The creation of practical tools is a necessary step for the widespread implementation of the methodology. This requires the development of software capable of automating information density analysis processes, the creation of user-friendly interfaces for working with data, and integration with existing information analysis systems.

The development of the proposed methodology could have a significant impact on several aspects of information technology. In the field of search engines, it will enable the creation of more accurate search ranking algorithms that take into account not only quantitative but also qualitative characteristics of information. In the field of artificial intelligence, the methodology could facilitate the development of more sophisticated natural language processing and data analysis systems. This methodology could also contribute to the development of personalized recommendation systems that can more accurately account for context and information quality.

5. Conclusions and Summary

This study is conceptual in nature and aims to formulate the theoretical foundations of a new approach to analyzing information space through the lens of the concept of information field density. The main result of this work is the introduction of a system of interconnected concepts, including an object's information field, its density, and its constituent information quanta (infons), which together form a holistic conceptual model for the qualitative assessment of information resources in digital space.

The mathematical formalizations proposed in this study are primarily illustrative in nature and serve to demonstrate the fundamental feasibility of quantitatively describing the qualitative characteristics of information structures. Further development of a rigorous mathematical framework will require extensive empirical research to determine the specific parameters of the uniqueness, relevance, authority, and other components of the proposed density model. Particular attention should be paid to the operationalization of the concept of infon, which requires the development of algorithmic procedures for automatically extracting information quanta from unstructured text arrays and qualitatively assessing them.

The theoretical significance of the proposed approach lies in its ability to overcome the limitations of existing methods for assessing information resources, which traditionally focus either on quantitative data characteristics or on highly specialized aspects of information quality. The practical significance of the developed concept lies in its potential applications in the development of next-generation artificial intelligence systems capable of generating more reliable and verifiable internal representations of real-world objects. The transition from statistical patterns extracted from uncontrolled text corpora to a systematic analysis of information density can significantly reduce the frequency of generating unreliable information in language models and increase the transparency of decision-making processes in intelligent systems. The presented conceptual model can thus serve as a starting point for the development of a new research paradigm in the field of information resource analysis and their qualitative characteristics.

References

1. Austin Harris, 2.5 quintillion bytes of data are produced by people every day, 2021, - URL: https://appdevelopermagazine.com/2.5-quintillion-bytes-of-data-are-produced-by-people-every-day/
2. Malakhov V.A., Bibliometric analysis as a method of science studies: possibilities and limitations // Science studies. 2022. No. 1. URL: https://cyberleninka.ru/article/n/bibliometricheskiy-analiz-kak-metod-naukovedcheskih-issledovaniy-vozmozhnosti-i-ogranicheniya (date of access: 09/08/2025).
3. Hogan A. et al. Knowledge graphs //ACM Computing Surveys (Csur). – 2021. – T. 54. – No. 4. – S. 1-37.
4. Kislitsyna M. Yu . Analysis of the Error Structure in Identifying the Author of a Text Using the Nearest Neighbor Graphs (2025). Scientific Visualization 17.2: 110–122, DOI: 10.26583/sv.17.2.08
5. Shannon C.E. A Mathematical Theory of Communication (1948)
6. Kolmogorov, A. N. (1965). "Three approaches to the quantitative definition of information"

7. Batini C. et al. Data and information quality //Cham, Switzerland: Springer International Publishing. – 2016. – T. 63.
8. Singhal A. Introducing the Knowledge Graph: things, not strings (2012) URL: https://blog.google/products/search/introducing-knowledge-graph-things-not/
9. Dbpedia : Global and Unified Access to Knowledge Graphs, URL: https://www.dbpedia.org/
10. Wikidata,URL: https://www.wikidata.org/wiki/Wikidata:Main_Page
11. Brin S., Page L. The anatomy of a large-scale hypertextual web search engine //Computer networks and ISDN systems. – 1998. – T. 30. – No. 1-7. - WITH . 107-117.
12. SI Chuprina,IA Labutin . A High-Level Adaptation Toolkit for Unified Integration of Software Systems with Neural Interfaces (2024). Scientific Visualization 16.4: 11–24, DOI: 10.26583/sv.16.4.02
13. SI Chuprina . Using Data Fabric Architecture to Create Personalized Visual Analytics Systems in the Field of Digital Medicine (2023). Scientific Visualization 15.5: 50–63, DOI: 10.26583/sv.15.5.05
14. NA Bondareva . The Impact of Input Data Density on the Performance of Graphic Neural Networks (2024). Scientific Visualization 16.5: 109 - 119, DOI: 10.26583/sv.16.5.08
15. Morrone M., Exclusive: Popular chatbots amplify misinformation, 2025, - URL: https://www.axios.com/2025/09/04/popular-chatbots-amplify-misinformation?utm_source=Securitylab.ru
16. AI False Information Rate Nearly Doubles in One Year, 2025, - URL: https://www.newsguardtech.com/ai-monitor/august-2025-ai-false-claim-monitor/
17. Li S., Guo H., Tang X., Tang R., Hou L., Li R., Zhang R. Embedding Compression in Recommender Systems: A Survey (2024), arXiv:2408.02304
18. Vafa K., Chen JY, Rambachan A., Kleinberg J., Mullainathan S. Evaluating the World Model Implicit in a Generative Model (2024) https://doi.org/10.48550/arXiv.2406.03689
19. Torsten A. Enßlin . Information theory for fields (2018). Annalen der Physik 2019, vol. 531, issue 3, p. 1800127 DOI: 10.1002/andp.201800127
20. Bondarev A.E., Bondarenko A.V., Galaktionov V.A., Klyshinsky E.K. Visual analysis of cluster structures in multidimensional volumes of text information / Scientific visualization, Vol. 8, No. 3, 2016, pp. 1-24