The quantum-mechanical approach to construction of quantitative assessments of some documentary information properties (on example of nuclear knowledge)

A A Lebedev¹, N V Maksimov² and E V Smirnova²

¹ Lebedev Physical Institute of Russian Academy of Sciences, 119991, Russia, Moscow Leninskiy pr., 53
² National Research Nuclear University MEPhI, 115409, Russia, Moscow, Kashirskoe highway, 31

E-mail: lebedevalex@live.ru

Abstract. The paper presents a model of information interactions, based on a probabilistic concept of meanings. The proposed hypothesis about the wave nature of information and use of quantum mechanics mathematical apparatus allow to consider the phenomena of interference and diffraction with respect to the linguistic variables, and to quantify dynamics of terms in subject areas. Retrospective database INIS IAEA was used as an experimental base.

1. Introduction
To present day there is no generally accepted constructive definition of a phenomenon of information and issue of its physical nature is still open that does not allow to quantify its properties. Among the works devoted to this problem [1-4] the common is following: information exists objectively and itself is an objective physical quantity among others – such as mass, energy, momentum, etc. In article [5] the analogy between information and light was considered. The concept of information quantum was proposed as a kind of signal having the meaning of a simple piece of information, which arises due to the transition of a system from one stable state to another.

In article [6] a physical nature of information was considered from the perspective of physical fields variety, their superposition and nonlinear interactions, the concept of information as a physical field. It is shown that all information generated and perceived is a physical field or superposition and interaction of the fields.

2. Description of the model
The objects of this study are terms of natural-scientific language of documentary database INIS IAEA. Each term initially has single meaning, single interpretation. But using of the original term in combination with others in various circumstances alters the perceptual meaning of the term.

Changing of term context perception can be attributed to the presence of non-point (probabilistic) nature of its existence in space of meanings. [7] Additionally, the context change takes place in the space of patterns (meanings) as a result of information objects activity (objects that have the ability to reflect and change state of another physical object or its relationships), but this change observing (registration) takes place by means of fixed (in the form of documents) knowledge. Thus, the projection of term space on the space of fixed knowledge will be determined by a distribution function, which is represented by description of all meanings contained in the term. By determining of
term meaning through a description of the circumstances in which it is used, we come to the different 
subject areas where the term is have a sense.

This approach is similar to probabilistic description of particle characteristics, which is used in 
quantum mechanics [8]. It considers particles wave functions instead of material particles considering.
Although the wave function has no physical meaning (meaning has only square of the wave function, 
defining the probability density of quantum system observing in different states), the wave properties 
of quantum systems help to explain some of phenomena.

2.1. Basic concepts and definitions

As subject area (SubjAr) we will assume a set of objects (and / or state) of natural or artificial origin, 
existing either in form of sustainable natural formation exists in result of evolution or allocated to 
some entity in accordance with its activity goals.

In informational practice the SubjAr is traditionally identified by area headings. The rubrics in 
documentary systems denote attachment of a document to one or more classes defined by 
classification system. That is each SubjAr acts as a fully autonomous entity and a person may hold a 
mental border between two non-identical SubjAr. Meanings that terms express associated with a single 
SubjAr, but rubrics only indicates the SubjAr where the meaning likely takes place.

As information objects (InfObj) will named those objects of SubjAr that have property (ability) to 
reflect and, subject to certain conditions, change the state of another physical object feature or 
relationships, and that will have the role of objects-originals patterns.

Informational interaction (InfInt) will be called processes in which the pattern objects are used as 
operating. Unlike physical interactions occurring between individual bodies (originals) at a particular 
time, informational interactions will involve their patterns – objects that reflect properties of some of 
the originals, but do not possess them. InfInt differs from the physical by the fact that we are not 
interested in the specific state of an information object, but in probability state of SubjAr, changing as 
a result of interaction [1].

Information - a superposition of possible states of InfObj in \( n \) SubjAr, characterized by function:

\[
\Lambda = c_1 \psi_1 + c_2 \psi_2 + \cdots + c_n \psi_n
\]

where \( \psi_i \) – InfObj distribution function for the \( i \)-th subject area, in particular, those may be language 
terms, representing a linguistic variables; \( c_i \) – probabilistic measure of SubjAr refering to \( i \)-th subject 
area. It should be noted that the SubjAr can be represented as InfObj, consisting of a set of information 
objects forming some integrity given by rubric. I.e. in general case rubric can act both as artificially 
given set of terms, and as a separate SubjAr.

Then it is possible to characterize the function of informational interaction:

\[
\Omega = c_1 \psi_1 \Theta_i^T + c_2 \psi_2 \Theta_i^T + \cdots + c_n \psi_n \Theta_i^T
\]

where \( \Theta_i \) is the distribution function for InfObj, which interacts with the original InfObj, for the \( i \)-th 
SubjAr. I.e. the information by interacting with specifically SubjAr, in fact, takes the only one state of 
many possible. Such information state resulting of the selection (InfInt) recorded in the form of context caused information object should be called knowledge [1].

Further, for term "information" using one needs to understand that we are talking about InfObj, 
which has a characteristic duality state: before interaction – it is some integral object, and during the 
interaction, it’s macroobject "quanta". Moreover, these related by certain way microobjects, at the same 
time can be used as elements of other patterns (image), present or future, the original objects. For 
the text form the information quanta essence are concepts, to be exact - denoting them language terms.

The term wave function (by analogy with the particles in quantum mechanics) determines its state 
at any time, to within a phase factor. By analogy with [9,10] we assume that the state of the term is 
defined by the wave function:

\[
\psi(x, t) = A(x, t) \cdot \exp(-i \cdot \varphi \cdot t)
\]

so that:
\[ |\psi(x,t)|^2 dx = P(x,t) \]  

(4)

where: \( x \) – coordinate of term in the semantic space, which is defined by rubricator (named, sometimes, as Subject index headings) of SubjAr,

\( t \) – time; \( A(x,t) \) – probability amplitude; \( \varphi \) – term phase; \( P(x,t) \) –probability that the term has a meaning corresponding to a neighborhood \( dx \), given by rubric \( x \) (at the present moment we propose to evaluate this value as identification part of documents with the term which belong to particular rubric \( x \) - Subject index field of INIS record).

The wave function is complex and represents the probability wave that describes the objectively existing capabilities of a particular result.

Coordinate of term is equal to \( x \) if document to which it belongs, is related to the rubric \( h_x \) from set \( \{h_1, h_2, \ldots, h_m\} \ x=1, \ldots, m \) ( \( m \) – the number of rubrics).

Term phase \( \varphi \in [0; \pi/2] \). The phase is associated with the presence of the term opportunity to join other terms. In general, it is a cyclic quantity that is characteristic of the wave process of term perception. In this paper, cyclic recurrence is not taken into account, as the value of the period is great in comparison with considered time intervals.

The magnitude of phase in particular time depends on the condition of its life cycle where term is, whether it is a structural unit which can be coupled with other terms to form a new meaning. The value depends on the phase relation between the real and imaginary part of the term, that is, how it is specific. The more real is part of the term, the more specific it is. Since the used database does not allow to identify roles and relationships between terms, the phase of the wave function was evaluated as quotient of \( L_1 \) and \( L_{12} \), where \( L_1 \) is a number of documents that contain the term and related to rubric \( h_x \) at time \( t \); \( L_{12} \) is a number of documents that contain the term and at the same time belong to the rubric \( h_x \) and to any of rubrics set \( \{h_1, \ldots, h_{x-1}, h_{x+1}, \ldots, h_m\} \) at time \( t \).

Thus, the probability distribution of term \( S_i \) gets in rubric \( h_x \) at time \( t \):

\[ P_i(x,t)=|A(x,t)|^2=(Re \psi(x,t))^2+(Im \psi(x,t))^2 \]  

(5)

Based on foregoing, InfInt process can be represented as a movement of some "information wave" from source to receiver, consisting of oscillating particles, which are terms of incoming message, constituting the wave front.

Each wave is characterized by a spectrum – descriptor distribution by subject areas (rubrics).

The set of terms representing descriptors environment (associated with them) are secondary wave front, etc.

Primary wave \( W_1=\{S_1, S_2, \ldots, S_n\} \), where \( S_1, S_2, \ldots, S_n \) - the terms included in the message.

Secondary wave \( W_2=\{S_{i1}, \ldots, S_{im}, S_{21}, \ldots, S_{2b}, \ldots, S_{n1}, \ldots, S_{nl}\} \), where \( S_j \) – terms that constitute the \( S_i \) term environment, i.e. they are with him in the same document in the same rubric. For the term \( S_i \) \( j=1, \ldots, m \).

2.2 Interference and diffraction of information

The foregoing idea of "information wave" in form of classical waves of mechanical nature is a good tool for macrophenomena illustrations, for example: meanings overlay, rounding, outrunning, etc.

By "interference of information" we assume this type of InfInt, in which there is mutual increase or decrease of resultant amplitude of two or more information waves due application to each other.

By "diffraction of information" we mean a type of InfInt at which a penetration of the waves in an area of "information shadows" takes place, that is, the information distribution in the source affects the information distribution to the receiver.

In the quantum-mechanical description of "information wave" it is a wave of probability, describing the objective existing possibilities of one or another result.
For the primary wave $W_1$ (Fig. 1), the resulting probability distribution is defined as the total distribution for all terms $S_1,S_2,...,S_n$: $P_{W1} = \sum_{i=1}^{n} P_i(x, t)$.

For the secondary wave $W_2$ (see Fig. 2): $P_{W2} = \sum_{i,j=1}^{n,m} P_{ij}(x, t)$.

Thus, information interference is $\text{InfInt}$ in which there is a relative increase or decrease resulting probability of meaning finding in a given subject area.

Considering the "wave of information" as the flow of "quantum information" (a pair of terms), we obtain:

$$P_1 = |A_1|^2, \quad P_2 = |A_2|^2, \quad P_{12} = |A_1 + A_2|^2$$

(6)

$$P_{12} = P_1 + P_2 + 2\sqrt{P_1P_2}\cos(\delta - \frac{\pi}{2})$$

(7)

where $\delta = \phi_1 - \phi_2$ – the difference of terms phases; $P_{12}$ – the probability of finding the meaning which get into a neighborhood $\delta x$ near $x$ coordinate of semantic space at joint use of these two terms.

From (7) it follows that the greater the difference between the terms phases, the interference will be observed stronger. This is generally consistent with the practice of terms-phrases constructing when one is a metaconcept term and another term denotes a particular object or property.

The phenomenon of diffraction (fig.4) can be treated as a special case of interference, namely, as the interference of secondary waves.

At the "overlay" of text, carrying information it is possible to restore original meaning of the message on the remaining part. I.e. "wave", which carries the information, like goes around an obstacle and gets in area behind him.
Consider an example that illustrates the outlined phenomenon.

For example, in message "NUCLEAR REACTION YIELDS" term "REACTION" will be dropped out. The terms included in the environment of terms "NUCLEAR" and "YIELDS", i.e. co-occur with them in a same document in the same category (secondary waves), interfere with each other. Let us construct the distribution of probability that the meanings get in specific SubjAr, resulting in the interference of secondary waves, and on the same graph shows the actual distribution for term "NUCLEAR REACTION YIELDS" (see Fig. 5). Based on how close these distributions are one can judge on the possibility of message with omission of original meaning restoring.

![Figure 5](image.png)

**Figure 5.** Comparative schedule of the probability distribution for entering in semantic area for messages without omission (P(fact)) and for case of waves interference for message with omission (P(inter)).

Resulting distributions are close to each other, get into the same SubjAr, which indicates that the meaning of the message "NUCLEAR REACTION YIELDS" can be restored with omission of term "REACTION" due to the interference of secondary waves (generated by terms associated with the terms "NUCLEAR" and "YIELDS").

3. Conclusion

A model of a documentary information properties (in particular, for terminology systems), based on the assumption that term of language may be defined as a physical particle, to which basic laws of physics can be applied was proposed. With using of wave and quantum concepts the phenomena of documentary information interference and diffraction was described. On the INIS IAEA database array a probability distribution of meaning in the subject areas as a result of terms interference and diffraction that experimentally confirm adequacy of the model was constructed.

The proposed approach for documentary information properties definition allows to improve the quality of information request, in particular through a more effective request expansion procedures and dynamically definable issuance thresholds.

References

[1] Maksimov N V 2010 *Automatic Documentation and Mathematical Linguistics* 44(4) pp 177-186
[2] Brilloin L 1956 *Science and Information Theory* (New York: Academic Press)
[3] Colin K K 2010 *Alma mater (High Scholl Herald)* 1 pp 1-8
[4] Ursul A D 2010 *Nature of information: a philosophical essay* (Chelyabinsk: Chelyab. State Academia of Culture and Arts)
[5] Romanenko V N, Nikitina G V 2011 *Modern high technologies* 2 pp 1-4
[6] Vildyaykin G F 2010 *Scholarly Notes of Komsomolsk-na-Amure State Technical University* III-1(3) pp 46-56
[7] Nalimov V V 1989 *Spontaneity of consciousness. Probabilistic Theory of meaning and the Meaning architectonics of personality* (Moscow: Prometheus)

[8] Landau L D, Lifshitz E M 1989 *Theoretical physics, quantum mechanics (non-relativistic theory)* (Moscow: Nauka)

[9] Lebedev A A, Maksimov N V and Smirnova E V 2016 *Automatic Documentation and Mathematical Linguistics* 1 pp 26-34

[10] Gabora L. and Aerts D 2002 *Journal of Experimental and Theoretical Artificial Intelligence* 14 pp 327-358