The basic principles of building an analytics platform for the search of new scientific knowledge

V O Gladyshev¹, D A Bazlev² and A V Kayutenko¹
¹ Bauman Moscow State Technical University, Moscow, Russia
² Central Aerohydrodynamic Institute (TsAGI), Zhukovsky, Russia
E-mail: vgladyshev@mail.ru

Abstract. By analyzing the progress in the development of ideas and techniques for recording gravitational waves, the paper introduces a vector characteristic for assessing the future development of scientific knowledge \( \tilde{R}(t) \), taking into account new dynamic scientometric indicators. This characteristic is a function of time and depends on the priority assessment of the scientific problem \( P(t) \) of the divergence of the flow of new knowledge \( Z(t) \) stimulated by the solution obtained to this problem, or a new result in this area, as well as the time-varying scientometric vector potential \( N(t) \).

1. Introduction
The global development of information resources, the increasing speed of obtaining and accumulating information in various fields of knowledge have led to the need to analyze large amounts of data (Big Data) and to automatically single out previously unknown knowledge from these data, which can result in real benefits (Data Mining).

To a certain extent, this development is caused by the search for effective tools of managing the processes of obtaining new knowledge. Hence, bibliographic databases have been widely spread in recent years: Scopus, Ingenta, Google Scholar, Scholar.ru, Science Direct, SciNet - Science search, Science Research Portal, Scirus, Cite Seer Publications Research Index, and others. To varying degrees these databases use scientometric indices to assess the potential of authors (researchers) and organizations (Universities, etc.). As a rule, these indices are based on identifying the citation of the articles published in scientific journals.

The use of these indices has become the norm for scientific foundations when allocating funding for projects and research teams and is a powerful tool of affecting the development of science. If an article is published in a scientific journal or other publication, it is certain to have been reviewed and undoubtedly contains a new scientific achievement, i.e. new knowledge.

Moreover, the higher the ranking of the journal, the higher the level of the review and, consequently, the accuracy of the result is more reliable, i.e. its novelty, priority, authenticity are undeniable.

Furthermore, we often associate the ranking of the journal with the result level, i.e. its importance, significance, complexity of the research and analysis methods used. We also expect the effects the result can possibly produce on the development of scientific thought and its applications. It is a reasonable expectation, as in high-ranking journals the competition is higher as well.

However, it is obvious that even one journal publishes the articles that are distinctly important for scientific and technological progress. Thus, for example, theoretical journals can publish articles
containing some innovative ideas and hypotheses that may later turn out to be erroneous, but, perhaps, they will encourage researchers to find the right solution to the problem.

In this regard, it becomes obvious that the adopted system of assessing the significance of research and the potential of the research team or the author is far from being perfect, and it is advisable to further look for the distinctive signs of scientific research and their results.

2. The problem of assessment of acquiring new scientific knowledge
The history of the development of science and scientific discoveries makes it possible through the existing examples to evolve some distinctive characteristics of scientific ideas that led to the significant development of science and technology.

For example, one of the fundamental problems of the 20th century was the problem of detecting gravitational radiation from distant space objects.

The existence of gravitational waves (GW) was theoretically predicted by Albert Einstein in his work devoted to calculating the power of gravitational radiation [1]. In [1-2] Einstein established the transversality of GW and two degrees of polarization. Later Arthur S. Eddington showed that flat GWs, as well as waves coming from a point source, propagate at the speed of light [3]. In 1988, Academician Vladimir A. Fok was the first who drew attention to the possibility of using astrophysical catastrophes as a source of powerful gravitational radiation [4].

Further calculations showed that a possible change in the distance between test masses, due to the action of a gravitational wave of cosmic origin, is very small. Therefore, for a long time the problem of detecting gravitational radiation remained the subject of theoretical research.

The existence of gravitational radiation was first confirmed in experimental studies by Joseph H. Taylor, Joel M. Weisberg and others [5-6]. These research works investigated the effect of the slowing down period of the binary stellar system PSR 1913 + 16 due to the energy losses for gravitational radiation. The obtained results accurately coincided with the calculated values, obtained according to the solution of general relativity equations.

However, many years before, different laboratories began their experimental search of methods for recording gravitational waves from space radiation sources by terrestrial gravitational antennas.

Since the general theory of relativity was built, various methods for detecting gravitational radiation have been introduced, most of which have remained unrealized either because of the lack of the method sensitivity, or because of the complexity of its technical implementation.

The article [7] contains the main ideas and gives a classification of methods for recording gravitational waves, which makes it possible to systematize the basic physical principles of recording.

In general, the most promising and technically well-provided are the projects of laser interference gravitational antennas, already built, or being constructed, with sensitivity reserves.

The idea of using a pair of free mass-mirrors and a laser interferometer to record their small oscillations caused by GWs was first proposed by Vladislav I. Pustovoit and Mikhail E. Gertsenshtein in [8].

This type of broadband gravitational antennas contains a lot of possibilities concerning the methods for recording GW, signal extraction methods, the possibility of including gravitational antennas into the network.

The main element of the laser interference GA is Fabry-Perot multi-beam free-mass resonator, on the properties of which the sensitivity and noise immunity of GA is largely dependent [9].

In scientific literature one can find the articles describing the significant progress in the use of interferometers for recording gravitational waves in the projects of VIRGO (Italy, France), LIGO (USA), TAMA (Japan), CLIO (Japan), GEO-600 (Germany) and OGran (Russia).

The first direct detection of gravitational waves by the LIGO and VIRGO collaborations was announced on February 11, 2016. The results were published in the journal Physical Review Letters [10] and brought about an extensive discussion [11].

However, long before this success in the search for gravitational waves, experimental attempts were made to record gravitational waves using the idea suggested by Joseph Weber [12]. It is on record that
the first gravitational antennas were built as solid-state detectors. It took years of experiments and theoretical studies to come to the conclusion that the more promising is the idea of using displacement interferometers. This story gives some insights into the assessment of a promising scientific idea, which is closely related to the technological progress in the field of laser, vacuum, electronic, computer technology, manufacturing of high-quality optics and solving many other problems.

3. The elements of information analysis system of scientific knowledge

When the first gravitational antennas were built, computer technology was not used yet, and it impossible to comprehensively assess the future development of the method for recording gravitational waves. The use of computer technology in assessing scientific projects also leads to the need to introduce new vector or multivariate characteristics of the development of scientific knowledge that can analytically reflect the logic of assessing the prospects of a particular study. It would be safe to assume that such an assessment can be based on extrapolation from the available scientometric data. At the same time, the indices used are of an integral nature and do not take into account the dynamic pattern. In particular, this leads to a young scientist or research team losing the competition, other factors being equal, although their development could be rapid. As a result, there arises the need to develop new scientometric indices that take into account the dynamic behavior of the estimated results. The set of both adopted and new indices, can be represented by the vector potential \( \vec{N}_i(t) \) of the researcher or research team. Moreover, a drawback of the existing system for assessing scientific projects is the fact that not the scientific result itself is evaluated, but its indirect indicators (for example, citedness of the publication, or the impact factor of the journal).

To solve this problem, we need to create a system for a more accurate assessment of scientific knowledge obtained, i.e. theoretical or experimental result, technology, and give a unique digital code to it. For this purpose, it is necessary to develop a universal template of the scientific result, which would enable classification of new and known scientific knowledge in a semi-automatic mode. The creation of such an information system for classification of scientific knowledge is an inevitable step within the conditions of the rapid growth of scientific publications throughout the world. This will prevent the results from being duplicated, as well as reduce the number of works that do not contain new scientific results. Coding scientific work will essentially mean the certification of copyright on the obtained scientific result, similar to patenting. Editorial boards of scientific journals should reject papers that contain a result duplicating already registered scientific knowledge. Literature references should contain sources with the results codes, on the basis of which a new article is based. It is these works that should receive citation indices, as the new result would not be possible without these works.

Thus, according to one scientific result, a certain number of new scientific results can be obtained at a fixed point in time, which, in turn, can become the basis for obtaining other new results. All these works represent a certain set that characterizes the power of the initial result, which became the basis for the development of scientific knowledge. This value can be identified as a number of connections between the basic work and its derivatives, i.e. made not on its basis. Because the basic work is the source of new knowledge, its significance can be assessed by the divergence of the flow of new knowledge

\[
\ddot{Z}_i(t) = div F = \lim_{v \to 0} \frac{\Phi_F}{V}, \Phi_F = \oint_S \left( \vec{F}, d\vec{S} \right),
\]

where \( \Phi_F \) is the flow of the vector field \( \vec{F} \) through the spherical surface of the area \( S \), bounding the volume \( V \). The field is a vector field. Each new scientific work is characterized by a vector in the space of scientific knowledge.
The divergence of the flow of new knowledge for each work is also a function of time. Moreover, the assessment of the priority of the scientific task $\tilde{P}_i(t)$ depends on time, which is an important factor in determining whether it is reasonable or not to support a particular study.

The progress in the field of science can be defined as the non-zero vector of development $\vec{R}_i(t)$ (Figure 1). However, the development or progress does not result in a breakthrough, intermittent growth of scientific knowledge, for this purpose mutual reinforcing influence of various factors is necessary.

![Figure 1](image)

**Figure 1.** The vector of development $\vec{R}_i(t)$ does not necessarily end in the area where a breakthrough is possible, i.e. intermittent accumulation of knowledge and technology.

It is obvious that the expectation of a breakthrough in a particular field of knowledge should be determined not by an additive but by a multiplicative law, including the divergence of the flow of new knowledge $\vec{Z}_i(t)$, priority assessment $\tilde{P}_i(t)$ and the scientometric vector potential $\vec{N}_i(t)$. It should be noted that material-and-technical, organizational and other characteristics are not taken into account here. They are commonly used when comparing scientific projects and teams. This is due to the fact that our goal is to assess the scientific potential and its ability to acquire new knowledge, i.e. other factors being equal.

In the simplest case, the extrapolation operation can be introduced as follows:

$$\vec{R}_i(t) = \tilde{P}_i(t) \times \vec{Z}_i(t) \times \vec{N}_i(t).$$

Thus, a possible breakthrough is predicted by the superposition or intersection of three vector fields.

Let us note that currently in the field of applied and basic research, the priority assessment $\tilde{P}_i(t)$ is not used explicitly, forming a list of promising, breakthrough lines of research. Therefore, this field can be formalized according to the available data.

4. Conclusion

In general, the analysis leads to the conclusion that the significant progress in the field of scientific development will occur when an information analysis system relying on the four basic elements is introduced: the extension of the standard set of scientometric indices, introduction of dynamic indices;
introduction of a universal classifier of scientific knowledge; formation of the priority task database; development of an algorithm for estimating the feasibility and forecasting a breakthrough.

References
[1] Einshtein A 1965 Approximate integration of gravitational equations Coll. sc. pap. in 5 vol. 1 (Moscow) pp 514–523
[2] Einshtein A 1965 On gravitational waves. Coll. sc. pap. in 5 vol. 1 (Moscow) pp 631–646
[3] Eddington A 1992 Proc. Roy. Soc. A102 pp 268–282
[4] Fok V A 1955 Theory of space, time and gravitation (Moscow, GITTL Publ.) 504 p
[5] Taylor J H and Weisberg J M 1982 A new test of general relativity: gravitational radiation and the binary pulsar PSR 1913+16 Astrophysical Journal. 253 2 pp 908-920
[6] Taylor J H 1982 Gravitational radiation and the binary pulsar Proc. 2-th Marcel Gross. Meeting on Gen. Relativity 1979 in Amsterdam pp 15-19
[7] Gladyshev V O and Morozov A N 2000 Classification of Gravitational-Wave Antennas by the Methods of Gravitational Radiation Detection Measurement Techniques 43 9 pp 741-746
[8] Gertsenshtein M E and Pustovoit V I 1962 To the problem of detection of low frequency gravitational waves ZhETF - Journal of Experimental and Theoretical Physics 43 2 (8) pp 605–607
[9] Gladyshev V O and Morozov A N 1993 Low-frequency optical resonance in a multiple-wave Fabry-Perot interferometer Technical Physics Letters 19 7 pp 449-451
[10] Abbott B P et al (LIGO Scientific Collaboration and Virgo Collaboration) 2016 Observation of Gravitational Waves from a Binary Black Hole Merger Physical Review Letters 116 6
[11] Pustovoit V I 2016 On the direct detection of gravitational waves Phys. Usp. 59 pp 1034–1051
[12] Weber J 1977 Topics in Theoretical and Experimental Gravitation Physics (New-York, London, Plenum Press)

Acknowledgements
The work was supported by Russian Foundation for Basic Research grants no. 16-02-00488 A and no. 16-08-00618 A.