

---

ADVANCED STUDIES IN ARTIFICIAL INTELLIGENCE  
AND MACHINE LEARNING

---

## Digital Ark of Knowledge

V. V. Goryachko<sup>a</sup>, A. S. Bubnov<sup>a</sup>, E. V. Rayevskii<sup>a</sup>, and Academician of the RAS A. L. Semenov<sup>a,b,\*</sup>

Received October 30, 2022; revised November 6, 2022; accepted November 8, 2022

**Abstract**—The general problem of creating and using encyclopedic knowledge in the digital civilization is considered. The creation and use of a digital encyclopedia are an example of intellectual activities of a collective expanded personality. The basic principles concerning the creation and functioning the “Digital ark of knowledge” at Lomonosov Moscow State University are described in the case of mathematical sciences. The arising problems and their solutions are discussed.

**Keywords:** encyclopedia, GRE, expanded personality, expanded consciousness, Wikipedia, collective knowledge, open public information resources, professional community, Lomonosov Moscow State University, mathematics

**DOI:** 10.1134/S1064562422060096

### 1. INTRODUCTION

It was noted in [1] that the development of human-kind can be characterized by information revolutions. These revolutions were determined by the appearance of

- consciousness,
- speaking,
- writing,
- automation of intelligent activities—artificial intelligence.

Each of the revolutions was accompanied by changes in the ways humans think, communicate, and act. We can say that each of them expanded the human consciousness and personality [2, 3]. Simultaneously, each of the revolutions led to a reduction, sometimes, disappearance of certain forms of activities and related human abilities, as was noted by Vygotsky [4–6]. In relation to the appearance of writing, this was mentioned by Plato in his dialogue *Phaedrus*, referring to (nonliterate) Socrates [7].

The concept of expanded personality is a fruitful metaphor for

- description of a modern human as a participant of productive, social, and economic relationships;
- design of educational objectives and contents and the assessment of educational results.

The last of the listed revolutions began about 100 years ago within mathematics. Specifically, mathematicians, having begun to construct formal mathematical systems describing mathematical reasoning and

mathematical communications, eventually constructed a universal formalization of intelligent activities as a whole [8]. Due to the achievements in natural sciences and technology, this formalization materialized in the computer created in the first half of the 20th century. Today, progressively growing segments of intelligent activities are passed by humans to computers, starting with rational elements, such as the solution of computational problems, including numerical simulation and computer algebra. In the 21st century, the soaring growth of digital technologies has led to the automation of intuitive human activities—machine learning—which is often identified with artificial intelligence as a whole. We will not get into a terminological discussion here.

Revolutionary events also occur within the revolution of artificial intelligence. The most important of them is the accumulation of information in a common space, its universal availability, and the organization of simple access to it. This “subrevolution” of knowledge (similar to Gutenberg’s subrevolution with the revolution of writing) was based on

- the Internet,
- mobile devices (smartphones),
- search engines.

Possibly, the next revolution will be based on a direct “brain–computer” interface. Today, this interface has been partially implemented in the form of an assistive listening device for some types of hearing loss (cochlear implants), visual prosthetics, and (still) primitive transfer of nervous pulses to electronic devices, in particular, for people with limited mobility.

The major, indeed revolutionary events of the 21st century also include the start of fundamentally new forms of collaborative intelligent activities, namely,

---

<sup>a</sup> Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Moscow, Russia

<sup>b</sup> Moscow Institute of Physics and Technology (National Research University), Dolgoprudnyi, Moscow oblast, Russia

\*e-mail: [alsemno@ya.ru](mailto:alsemno@ya.ru)

the creation of a collective expanded personality. The clearest and most convincing form of such activities is Wikipedia.

## 2. CREATION OF WIKIPEDIA BY A COLLECTIVE EXPANDED PERSONALITY

Collections of knowledge about the world have played a fundamental role in the development of science, culture, and education. It suffices to mention Komensky's *Orbis Pictus* [9], a defining textbook with pictures for children published in many languages, and *Encyclopedia of Diderot and D'Alembert*, which marked the Age of Enlightenment [10].

Wikipedia is a symbiosis of Humankind and Technology. It is a product of an expanded global human personality, and it has become possible due to digital technologies. Thanks to Wikipedia, human knowledge is now available to most of the world's population: in fact, this knowledge is used by much more people than all paper-based encyclopedic editions. Specifically, the rate of use of Wikipedia is higher by several orders of magnitude than that of paper encyclopedias.

The foundation for the social technology of Wikipedia consists in its rules—pillars, reasonable constraints: encyclopedic character, neutrality, availability, mutual respect, and the absence of strict rules.

Thus, Wikipedia is the most important mass response of AI to the impossibility for individuals and humankind to cope with the exponential information explosion.

Serious difficulties and shortcomings of Wikipedia that are critical to certain issues and communities (some of them are discussed below) have only a slight effect on its general importance.

## 3. USE OF WIKIPEDIA BY AN EXPANDED PERSONALITY

The most important effect of the Wikipedia project is its active use by the world's population. Otherwise, the project would be, of course, senseless. Today, when a scholar, writer, or cultural figure reads a lecture in a large audience and refers to a certain unexpected idea or little-known historic figure or uses a specific concept, there appears a movement in the audience: several people get out their mobile phones. This means that the audience includes expanded personalities, part of whose individual knowledge is encyclopedic articles of Wikipedia. This was not possible in the pre-digital era: did a curious young man in the middle of the 20th century often refer to the *Great Soviet Encyclopedia*? Today, in contrast, even an average school student, having received a task going beyond the scope of the textbook, immediately looks up in Wikipedia. Another important fact is that even elementary school students can create and use their own micro-wikipedia: collective knowledge about the surrounding

world (natural, anthropogenic, social) of students of class (and, then, of school). This approach is used in the draft course "Future world" for elementary school [11].

Thus, today the above-mentioned subrevolution of knowledge has covered a significant part of the world's population, which is much larger than the population that was able to write in the 19th century (recall that this fraction of the population in Russia did not exceed 15%). In fact, as was noted by numerous authors, the idea of a knowing person has changed, which drifts toward an expanded personality. Today, an example reflecting this migration in the scientific discourse can be seen even in Bachelor's degree works [12].

It has happened repeatedly over the last century that a technological achievement originated in the West was then mastered and interpreted by China. Baidu Baike, an alternative to Wikipedia created in China on an authoritative basis, has quickly left behind the global Wikipedia (see [13]).

Another phenomenon important for the global scientific community is open access publications and the culture of electronic preprints. An illustrative, yet unique, example is Grigori Perelman, who has received the highest mathematical awards for this-type publications, which were not peer-reviewed (or checked for plagiarism).

Finally, there is huge amount of professional literature on the Internet, often doubtful in terms of the quality of digital images and ownership.

## 4. SHORTCOMINGS AND PROBLEMS OF WIKIPEDIA

Below is an example of criticism of Wikipedia taken at random from the Internet:

"However, the main problem of Wikipedia is that its moderators and administrators, being specialists in wiki-labeling, are often not experts in the fields they write about. To put it simply, they only compile information from different sources in a single text. The Wikipedia rules do not forbid a mechanic from the state of Texas from writing articles about oncology achievements in Russia, as an example" [14]. The author of this publication pointed out not only the shortcoming of Wikipedia being "too democratic," but also the possibility of exploiting this shortcoming for improper, including personal or political, purposes in contradiction to the original idea and ideology of Wikipedia.

Remarkably, it is the original democratic idea that, in many respects, has ensured the proof of concept of this major sociotechnological project.

Today, the key problems of Wikipedia are as follows:

- differences in the level of motivation and qualification of some authors (those who can write professionally do not write),

- quality of articles,
- completeness of references and reference sources.

Despite these problems, it is clear that today Wikipedia has become a phenomenon that considerably supplements digitized versions of conventional encyclopedic resources.

## 5. ARK OF KNOWLEDGE AND FEATURES OF THE PROJECT AT MSU

In the pre-digital era, great progress was achieved in our country in the field of encyclopedic editions, such as the *Brockhaus and Efron Encyclopedic Dictionary* and the *Great Soviet Encyclopedia* (GSE) in three editions. The *Mathematical Encyclopedia* [15], which was created at the late Soviet period of development of mathematics in our country, became a unique phenomenon in the world. These words are supported by the fact that the *Encyclopedia of Mathematics* of the European Mathematical Society was created on its basis [16].

The *Great Russian Encyclopedia* (GRE) continued the tradition of GSE in paper format and, in the last decade, in digital format [17]. For the professional community, the collaboration with GRE provides an opportunity to address the whole society and receive state support. In what follows, the ongoing collaboration of the professional scientific community with GRE is described primarily by the example of mathematics and information technologies.

Note that the Russian-language Wikipedia develops in parallel with GRE, and today it is among the top five largest wikipedias in the world. The amount of open access publications is also growing.

When we began our cooperation with GRE, we faced the (worldwide) problem of low motivation in professionals regarding writing encyclopedic articles. The direct use of Wikipedia mechanisms would lead to a decline in the quality of the result.

The following approaches were developed to cope with this problem:

- The creation of articles is separated from the other functions (order, review, and passing to GRE).
- The quality of articles is ensured by the professional community, by analogy with scholar journals, dissertation councils, etc.
- An **Ark of Knowledge** is created, where working materials are prepared and preliminarily placed and, after their approval, are passed to GRE.
- Existing sources, primarily *Mathematical Encyclopedia*, are used as a basis.
- A convenient platform for collective work with documents (wiki-editing) is placed on servers in the Russian Federation and is technologically supported.
- Articles published in GRE are regarded as scientific publications in leading journals.

The structure of processes constituting the creation of the encyclopedia and the participants of these processes can be described as follows.

- A **professional** who has taken responsibility for some field of knowledge from the professional community, namely, a university department head (transfer of knowledge to new generations of professionals), department head at a research institute (new studies and results), chairman of an academic council (assessment of others' contributions), journal editor-in-chief (presentation of new results to the community), academy member, or RAS professor (elected by the professional community):

- Receives an invitation to take part in writing the Encyclopedia in his or her domain of expertise from a reputable scientific source (e.g., a RAS department); this participation does not assume writing articles in person, but means making a list of articles, choosing authors, and reviewing written articles.

- Rejection or total disregard for the invitation puts this professional outside his or her reference group. An analogy is that rather few experts decline to join the editorial board of an academic journal.

- Acceptance of the invitation means that the professional, together with colleagues, shares the responsibility for the completeness and quality of the representation of their field in the Encyclopedia.

- The invited professionals in each field make up an **editorial group** (EG) in that field (e.g., number theory or geneticists), into which they can coopt colleagues, as a rule, having academic status (scientific degree).

- Each EG member, unless there are any objections from the other EG members, can write an article himself or can invite any **author** to write an article on a subject from the EG field or put an open invitation to write an article on the proposed subject on the EG page.

- Everyone can write an article on a subject proposed by some EG, or can offer his own subject and submit an article to the Ark; moreover, this article can also be put, for example, in Russian Wikipedia.

- A submitted article is placed in the Ark after it has been approved by at least one EG member.

- Articles created in the Ark can be based (with a reference) on those from the *Mathematical encyclopedia*, GSE, *Mathematical encyclopedic dictionary*, or open sources. The EG can propose that an existing or revised article be placed in the GRE.

- **Scientific editorial board of the GRE** in a given domain of knowledge, for example, mathematics, sends existing articles from encyclopedic sources to EG for placing them in the Ark and revising them. If all EGs agree, an article is placed in the GRE, and a corresponding reference appears in the Ark.

- Any member of the editorial group can propose passing an article from the Ark to the GRE. An article

having no objections from EG members is passed to the GRE, a contract is signed with the author, the authorship can be indicated in the GRE, and the author gets permission from the GRE to publish the article with a reference to the original source in part or as a whole, in the original or as a translation into foreign languages, after publication of the article in GRE, with a reference to the GRE.

If an authoritative scholar is invited by EG members to write an article on the subject of his or her expertise, the work on article writing can be organized as follows:

- EG can make recommendations on the article to be written, including, for example, the size of the article version to be placed in the GRE, a list of questions to be covered in the article, and information on related articles in a possible network structure: more general articles, references to the given articles, more specialized articles, and neighboring articles with possibly overlapping contents. In the course of preparing an article, the author can significantly deviate from these recommendations, proposing alternatives.

- The author invites one or several colleagues and students to write an article and gives a talk-interview, using the list of questions proposed by EG. The listeners also ask questions of interest. The talk is recorder (if necessary, notes are made, formulas are written, etc.).

- The recorded talk is transcribed using, if necessary, the resources of the Ark.

- The transcript and the record are processed by the listeners, formulations are refined, formulas and references are inserted, etc.

- The processing result is presented to the author, who finishes the text; some of the listeners can be added as coauthors, etc.

Thus, the following conclusions can be drawn.

1. The traditional structure underlying the Ark is a **professional community** with its links, hierarchy, motivations, etc., which have been built continuously for centuries. Its members can lack resources or motivation for creating materials for the Ark on their own.

2. This structure is strengthened by a **wiki-community**, which is a neighboring structure expanding and supplementing the professional community. The wiki-community is dynamically formed by all people who are motivated enough to enter the project. This community or its individual members can lack competence or influence for creating an authoritative and qualitative source of knowledge.

Our scientific and educational community provides an example of a balanced combination of traditional and digital mechanisms.

The quality of materials presented in the Ark and GRE on a certain field of knowledge reflects the quality of scientific knowledge of professionals and their interest in maintaining this quality. The number and

size of materials also reflects the interest of the wide community in the given subject.

## 6. METHODOLOGY AND TECHNOLOGY OF THE ARK

The ark represents a structure intended for continuous network creation of network products for network users using professional hierarchies to ensure the required quality.

This structure can be regarded as a metaphor for machine learning implemented in the sociotechnological environment of a collective expanded personality. In the corresponding “neural network,” there is a layer of users of the developed contents, a layer of developers—authors and editors, a layer of experts—EG, and a customer layer—the GRE. We plan the process of learning and improvements of the interaction rules as part of functioning of the network.

The core of the platform is based on the wiki-system developed at the Faculty and Cybernetics and Computational Mathematics of Moscow State University. The choice of this type of software is explained by its following advantages:

- A large number of remote users can work simultaneously.

- A detailed history of all editions of user-formed information content is stored.

- Users can work with a thin client system, since the software and information are stored in the central kernel.

- A unified simple language of wiki labeling can be used to form texts, including formulas.

- Hypertext documents and templates and the organization of references are easy to create.

## 7. FURTHER DIRECTIONS OF DEVELOPMENT OF THE ARK

A further development of the Ark can include the following directions:

- placement of archives of preprints similar to foreign analogues;

- placement of open access materials, for example, educational multimedia courses and video hosting, including interviews and memoirs concerning the history of mathematics and the history of science;

- placement of materials with different rules of development and access;

- placement of learning management systems or integration with them;

- a system of video conferences with automated annotation and transcription.

EG can become active participants of publishing online academic journals and various forms of expertise.

## 8. CONCLUSIONS

The MSU scientific and educational project “Ark of knowledge” is based on a digital platform developed at Lomonosov Moscow State University. The platform ensures that a nearly unbounded number of specialists forming, developing, and supporting the unified network structure can work simultaneously with databases.

The interface of the project with the Russian and Russian-language community is the Great Russian Encyclopedia portal.

## ACKNOWLEDGMENTS

The authors are grateful to Academician V.A. Sadovnichii for suggesting the problem and to S. L. Kravets for helpful discussion and encouragement.

## FUNDING

Bubnov and Rayevskii acknowledge the support of the School of Mathematical Analysis Methods for Complicated Systems of Lomonosov Moscow State University. A.L. Semenov’s research was supported by the Russian Science Foundation, grant no. 22-11-00177.

## CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

## OPEN ACCESS

This article is distributed under the terms of the Creative Commons Attribution 4.0 International Public License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## OPEN ACCESS

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

1. A. L. Semenov and K. E. Ziskin, “Expanded personality as a basic subject and object of philosophical analysis: Implications for education,” *Humans and Artificial Intelligence Systems*, Ed. by V. A. Lektorskii (Yuridicheskii Tsent, St. Petersburg, 2022), pp. 172–200 [in Russian].
2. A. Clark, *Being There: Putting Brain, Body, and World Together Again* (MIT Press, 1998). <http://www2.econ.iastate.edu/tesfatsi/BeingThere.AClark1998.Entire-Book.pdf>. Accessed November 8, 2022.
3. M. Serres, *Petite Poucette* (Pommier, Paris, 2012).
4. L. S. Vygotsky, “The instrumental method in psychology,” *Collected Works* (Pedagogika, Moscow, 1982), Vol. 1 [in Russian]. [http://elib.gnpbu.ru/text/vygotsky\\_ss-v-6tt\\_t1\\_1982/go,108;fs,1](http://elib.gnpbu.ru/text/vygotsky_ss-v-6tt_t1_1982/go,108;fs,1). Accessed November 8, 2022.
5. L. S. Vygotsky, “The instrumental method in psychology” (1981). <https://www.marxists.org/archive/vygotsky/works/1930/instrumental.htm>. Accessed November 8, 2022.
6. L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes* (Harvard University Press, Harvard, 1980).
7. Plato, *Phaedrus* (Focus Philosophical Library, 2003). See also discussion at <https://cyberleninka.ru/article/n/mif-ob-izobrenii-pismennosti-v-dialoge-platona-fedr-sovremennoe-osmyslenie-idei-sokrata-i-platona-o-preimuschestve-ustnogo-slova-2>. Accessed November 8, 2022.
8. A. M. Turing, “On computable numbers, with an application to the Entscheidungsproblem,” *Proc. London Math. Soc.*, Ser. 2 **42**, 230–265 (1937). Erratum *Proc. London Math. Soc.*, Ser. 2 **43**, 544–546 (1937).
9. J. A. Komensky, *Orbis Pictus*. [https://en.wikipedia.org/wiki/Orbis\\_Pictus](https://en.wikipedia.org/wiki/Orbis_Pictus)
10. *Encyclopedia of Diderot and D’Alembert: Triumph of Enlightenment*, virtual exhibition. <https://press-libfl.tilda.ws/enciklopediya-didro-i-dalamberta-triumf-prosveshcheniya>. Accessed November 8, 2022.
11. A. L. Semenov, E. I. Bulin-Sokolova, A. A. Muranov, et al., “Digital technologies in primary school: Introduction to the future world,” *Proceedings of the 6th International Scientific Conference on Informatization of Education and Techniques for Electron Learning: Digital Technologies in Education, Krasnoyarsk, September 20–23, 2022* (Krasnoyarsk. Gos. Ped. Univ. im. V.P. Astaf’eva, Krasnoyarsk, 2022), Part 2, pp. 325–329.
12. G. Choi, “The internet-extended mind: The psychological ramifications and philosophical implications of cognitive offloading,” *Scripps Senior Theses*, 1675 (Claremont Colleges Library, USA, 2021). [https://scholarship.claremont.edu/scripps\\_theses/1675](https://scholarship.claremont.edu/scripps_theses/1675). Accessed November 8, 2022.
13. Baidu Baike. [https://en.wikipedia.org/wiki/Baidu\\_Baike](https://en.wikipedia.org/wiki/Baidu_Baike). Accessed November 8, 2022.
14. S. Kuznetsov, “Who is the editor? What is wrong with Wikipedia entries?” *Portal RuPosters*, March 26, 2020. <https://ruposters.ru/news/26-03-2020/oruzhie-informatsionnoi-voini>. Accessed November 8, 2022.
15. *Mathematical Encyclopedia*, Ed. by I. M. Vinogradov (Sovetskaya Entsiklopediya, Moscow, 1977) [in Russian].
16. *Encyclopedia of Mathematics* (European Mathematical Society, Springer-Verlag). <http://encyclopediaofmath.303.si>. Accessed November 8, 2022.
17. *Great Russian Encyclopedia*, electron version. <https://bigenc.ru>. Accessed November 8, 2022.

Translated by I. Ruzanova