Analysis of Results of the Rating of Volunteer Distributed Computing Projects

Vladimir N. Yakimets[1,2][0000-0003-4953-2932] and Ilya I. Kurochkin[1][0000-0002-0399-6208]✉

[1] Institute for Information Transmission Problems of Russian Academy of Sciences, Moscow, Russia
[2] The Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia
iakimets@mail.ru, kurochkin@iitp.ru

Abstract. Volunteer distributed computing (VDC) is a fairly popular way of conducting large scientific experiments. The organization of computational experiments on a certain subject implies the creation of a project of volunteer distributed computing. In this project, computing resources are provided by volunteers. The community of volunteers is about several million people around the world. To increase the computing power of the volunteer distributed computing project, technical methods for increasing the efficiency of computation can be used. However, no less important are methods of attracting new volunteers and motivating this virtual community to provide computing resources. The organizers of VDC projects, as a rule, are experts in applied fields, but not in the organization of volunteer distributed computing. To assist the organizers of the VDC projects authors conducted a sociological study to determine the motivation of volunteers, created a multiparameter method and rating for evaluating various VDC projects. This article proposes a method for assessing the strengths and weaknesses of VDC projects, based on the approach. The results of multiparameter evaluation and rating of projects can help the organizers of the VDC projects to increase the efficiency of computations, and the community of volunteers to provide a tool for comparing the various VDC projects.

Keywords: Volunteer distributed computing (VDC)· BOINC· The VDC project· Crunchers· Volunteers· Evaluation index of the VDC project· Characteristics for assessing the quality of VDC projects.

1 Introduction

The use of distributed computing systems for high-performance computing is an alternative to calculations on supercomputers and other multiprocessor computer systems. Distributed computing systems or grid systems have a number of features, such as heterogeneity of computing nodes, their geographical distance, unstable network topology and high probability of disconnection of a computing node or communication channel. But even with such features, the computing potential of the grid system can be huge because of the large number (hundreds of thousands) of compute nodes.

There are software platforms for organizing distributed computing, such as HTCondor [1], Legion [2], BOINC [3]. At the moment, the most common platform for organizing distributed computing is BOINC (Berkeley Open Infrastructure for Network Computing) [4]. Public grid systems attract the computing power of volunteers. As a rule, the organizers of public grid systems are scientific or educational organizations. A public computing grid system is called a volunteer distributed computing project.

Volunteer distributed computing projects (VDC projects) are rather heterogeneous, both in terms of topics and in their organization. There are long-lived projects that remain active and successfully conduct experiments for almost 20 years (SETI@home, Folding@home, Einstein@home). There are projects that were created for one large computational experiment or scientific problem, after several years, they stop working (POGS@home, Poem@home). Some projects simultaneously run several independent computing experiments (SZTAKI, LHC@home). Developers of some projects can pay great attention to the design of the project site (POGS@home, Acoustics@home), and other projects are limited to using the standard template of the BOINC platform (Climate@Home, Optima@home[11]). Some projects are considered only on the CPU, others use CPU + GPU.

Interaction with volunteers in projects has general principles[13]:

- Attracting computing resources to help the scientific team;
- Publication of experimental results;
- Publication of scientific and popular scientific publications on the project website;
- Interaction with volunteers;
- Organization of competitions for volunteers within the project;
- Motivation of volunteers with virtual prizes and certificates.

However, for each VDC project there are specific features of interaction with the community of volunteers. Some project is popular, as it was organized by scientists with a global name (LHC@home). Another project has a clear and high goal - the search for a cure for cancer (Rosetta@home). If a media project has been told about a project, this can give an impetus to the project even in the medium term - attracting new volunteers for several months to 1 year. The participation of the project in virtual sports (Formula BOINC, BOINC Pentathlon) can also attract significant computational resources to the project.

The websites of a number of VDC projects based on BOINC platform [3] were studied as well as information about them on the site boinc.ru. The list of VDC projects includes:

- SETI@home – one of the first projects on the basis of which the BOINC platform was developed. This project of a group of scientists from Berkeley University for processing data from radio telescopes.
- Asteroids@home project is organized by the Astronomical Institute of Charles University, Czech Republic. Project aimed at determining the shape and parameters of the rotation of asteroids according to photometric observations In the process of project implementation, the public database DAMIT is filled.

- POGS@home project is organized by the International Centre for Radio Astronomy Research, Australia. This project aimed at building a multispectral atlas (from near infrared to ultraviolet radiation), as well as at determining the rate of star formation, the stellar mass of galaxies, the distribution of dust and its mass in the galaxy and etc.;
- SAT@home project is organized by the Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences, Russia. Project associated with searching for solutions to such complex problems as inversion of discrete functions, discrete optimization, bioinformatics, and etc., which can be effectively reduced to the problem of the feasibility of Boolean formulas[9];
- Rosetta@home project is organized by Institute for Protein Design, University of Washington, USA. Project associated with solving one of the biggest problems in molecular biology - the calculation of the tertiary structure of proteins from their amino acid sequences;
- MilkyWay@home project is organized by Rensselaer Polytechnic Institute, USA. This project studies the history of our galaxy by analyzing the stars in the Milky Way galaxy's Galactic Halo;
- Some other VDC projects – Folding@home[5], Einstein@home[6], LHC@home[7][10], Gerasim@home[8]and etc.

It should be noted that in English and Russian literature a majority of publications and articles on Internet resources provide a description of the individual VDC projects [8][9][10]. Significantly less common are scientific papers which characterize organization of activities and the involvement of crunchers [11][14]. And it is quite rare to see works which examine various aspects of citizen participation in VDC projects [12, 13]. One of the first attempts of more or less systematic study of the conditions and characteristics of crunchers participating in VDC projects in Russia, as well as their motivations and preferences has become a sociological study among 650 Russian crunchers [15].

2 Description of the Methodology for the Analysis of VDC Projects

The methodology for the index evaluation of VDC projects includes the following main stages:

1. Creating an index model;
2. Collect information to describe the most important elements of VDC projects.
3. Conducting surveys of participants of selected VDC projects to calculate the values of the YaK-index and estimates.
4. Perform calculations and visualize the results.

The YaK-index was developed to assess the quality of the VDC projects implementation, which was applied to a number of the above-mentioned projects in order to identify their "weak" sides, develop proposals to increase their efficiency, increase

attractiveness for people interested in the VDC, and providing comparable information for the organizers of VDC projects.

2.1 Model of the YaK-index of the VDC Projects

We introduce the notation:

$i = \overline{1,n}$ - the ordinal number of the important characteristics of the VDC project (hereinafter referred to as characteristics),it is assumed that *n* equal to 7-9, that is, from 7 to 9 estimated characteristics of each VDC project will be taken into account;

$s = \overline{1,S}$ - the ordinal number of the VDC project, *S=34* (for this paper);

R^s - YaK-index of the VDC project *s*;

x_i^s - availability of characteristic *i* of the VDC project *s* (0 – if not present;1 – if available);

K^s - number of respondents for the VDC project *s;*

K - number of respondents for all VDC projects*;*

α_i^s - the mean weighting factor (significance) of characteristic *i* of the VDC project *s* from K^s respondents, $0 \le \alpha_i^s < 1$, $\sum_{i=1}^{n} \alpha_i^s = 1$;

ρ_i^s - the mean experts assessment of the quality of *i*-th characteristic of *s*-th VDC project,

The scale values vary from -2 to 2. A linguistic interpretation of these values is given in the questionnaire. If necessary $\rho_i^s \in \{-2,-1,0,1,2\}$ is converted into the set $\rho_i^s \in \{1,2,3,4,5\}$.

Identically for all *n* characteristics it is mapped one-to one in a numerical set from the set of possible linguistic estimates. The maximum value of the numeric scale is *m* . In our case, it is assumed that *m = 5*. The index values vary from 1 to 5. But it is possible to normalize the index values so that they vary from 0 to 1.

There are two possibilities for calculating the value R^s:

1. When the weights of characteristics for a VDC project are individual and independent of what such are such weights for all other VDC projects.
2. When for all VDC projects the same vector of weights is defined.

In the first case, the index R^s is calculated as follows:

$$R_1^s = \frac{\sum_{i=1}^{n} \alpha_i^s \bullet x_i^s \bullet \rho_i^s}{n_1^s m} \quad . \qquad (1)$$

Here n_1^s is the number of characteristics for VDC project *s*, $n_1^s \le n$.

In the second case

$$R_2^s = \frac{\sum_{i=1}^{n} \alpha_i \bullet x_i^s \bullet \rho_i^s}{n_1^s m}. \qquad (2)$$

2.2 Collect Information for Describing the Most Important Elements of VDC Projects

For the survey of volunteers, a preliminary list of characteristics for project evaluation was created. At conferences and round tables with representatives of the volunteer community and experts, discussions were held on this list and some corrections were made. As a result, 9 characteristics were selected to assess the features of VDC projects[16]:

1. The clear concept and vision of the project;
2. Scientific component of the project;
3. The quality of scientific and scientific-popular publications on the topic of the project;
4. Design of the project (site, certificate, screensaver);
5. Informativity of materials on the project site;
6. Visualization of the project results (photo, video, infographic);
7. Organization of feedback (forums, chat rooms, etc.);
8. Stimulation of the cruncher participation in the project (competitions, scoring system, prizes);
9. Simplicity of joining the project (there are no barriers and organizational or technical difficulties).

The assessment of each characteristic was supposed to be 5-point. The questionnaire was initially focused on the Russian-speaking audience, so the following assessments were proposed (with the following linguistic interpretation):

1. "+2" is excellent;
2. "+1" is good.
3. "0" is normal;
4. "-1" - it is necessary to improve;
5. "-2" is bad.

This interpretation of the estimates allows you to get rid of the unnecessary connotation with school grades (from 1 to 5) and makes it possible to use the entire range of estimates.

Due to the heterogeneity of the projects, each characteristic has a different impact on the evaluation of the project. Therefore, it was decided to assess the significance of each characteristic for a particular project. Weight characteristics could vary from "0" - not significant, to "10" - very significant. In addition to evaluating the project, several questions were asked in the questionnaire on the degree of involvement of the

respondent in this project: determining the respondent's status, determining the duration of the respondent's interaction with the project.

2.3 Conducting Surveys of Participants of VDC Projects

The questionnaire was implemented using Google Docs in Russian and English [21]. Information about the questionnaire was distributed by the administration of the site BOINC.ru on the profile forums and sites of the community of volunteers.

As a result, estimates were received from 259 respondents for 34 projects. However, for subsequent processing, only those projects for which more than 10 questionnaires were filled were taken (Fig. 1).

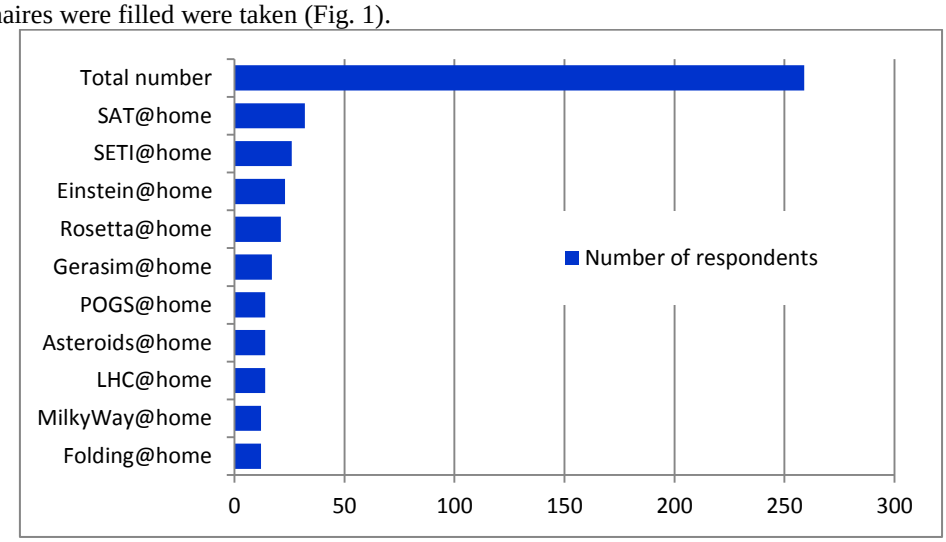

Fig. 1. Distribution of the number of respondents among projects.

The results were obtained in 2016-2017, therefore new projects such as RakeSearch, XANSONS for COD, etc., are not listed in the rating.

2.4 Visualization of Survey Results

To determine the averaged weights of each characteristic, questionnaires were used from all 259 respondents (for all projects). When comparing the average weight of a characteristic for individual projects, the values were not significantly different from the average values for individual projects. As an example, the project characteristics are ranked by weight according to the average values for all projects (Fig. 2).

7

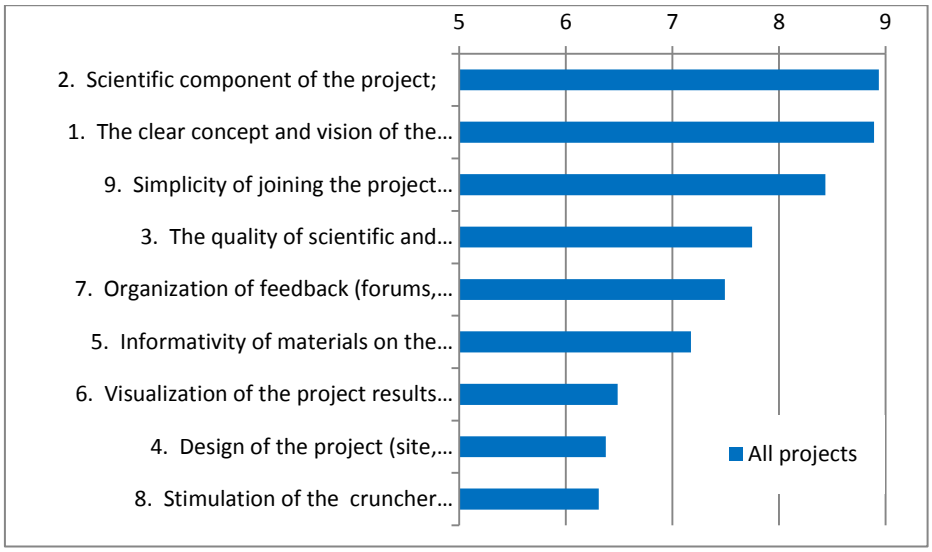

Fig. 2. Ranking of characteristics by weight.

For the visual presentation of the projects characteristics a radar diagram was used (Fig. 3). The characteristics are arranged in descending order of average weight.

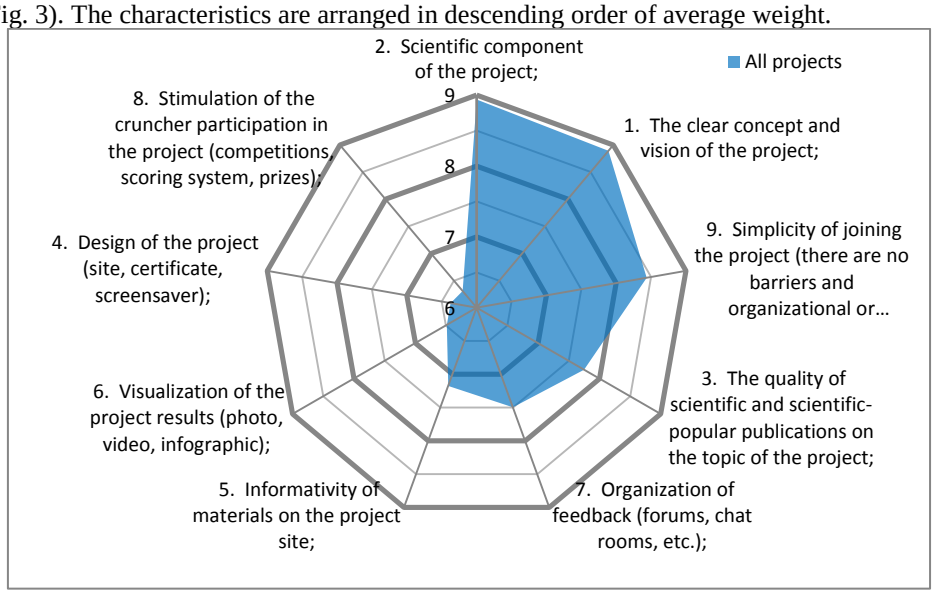

Fig. 3. Ranked weights of characteristics in radar diagram

The average estimations for all projects on the radar diagram (Fig.4) are given, while the ranking of the characteristics in descending order of importance is preserved (Fig. 2 and Fig.3).

8

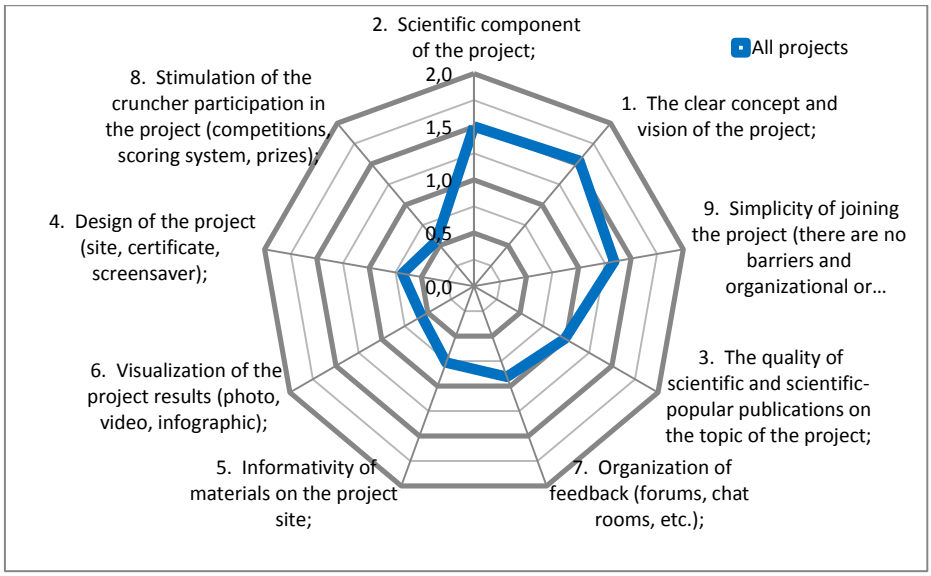

Fig. 4. Estimates of the characteristics of all projects, taking into account their ranking by weight

3 Results for Selected VDC Projects

Let's present estimates of large international VDC projects: Folding@home, Einstein@home, LHC@home and Russian VDC project Gerasim@home.

3.1 Folding@home

Folding@home is a distributed computing project for disease research that simulates protein folding, computational drug design, and other types of molecular dynamics [10]. The project started in 2000 and is one of the most successful international projects that began to use grid systems from personal computers. The audience of the project is more than 1 million users and continues to grow. For several years, the organizers of Folding@home have maintained compatibility with the BOINC client. This allowed the volunteers, who use a BOINC client, to participate in the project. But, despite this, the community of volunteers is very different from the volunteer community of VDC projects on the BOINC platform. This can be seen from the various estimates of the significance of the characteristics (Fig.5). The success and attractiveness of the project for volunteers can be seen at Fig.6, Where project estimates for each characteristic are better than average values for all projects.

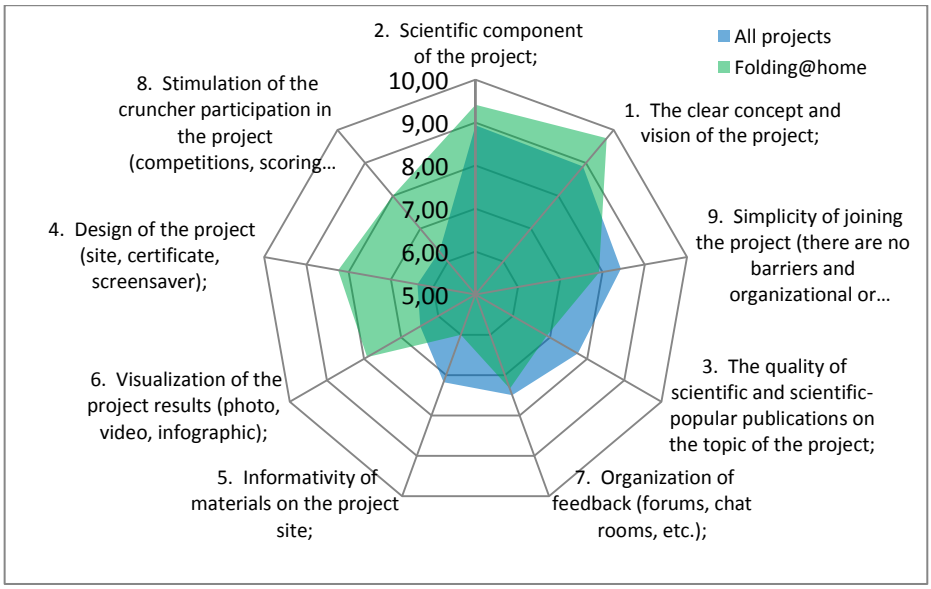

Fig. 5. Weights of characteristics for the project SETI@home

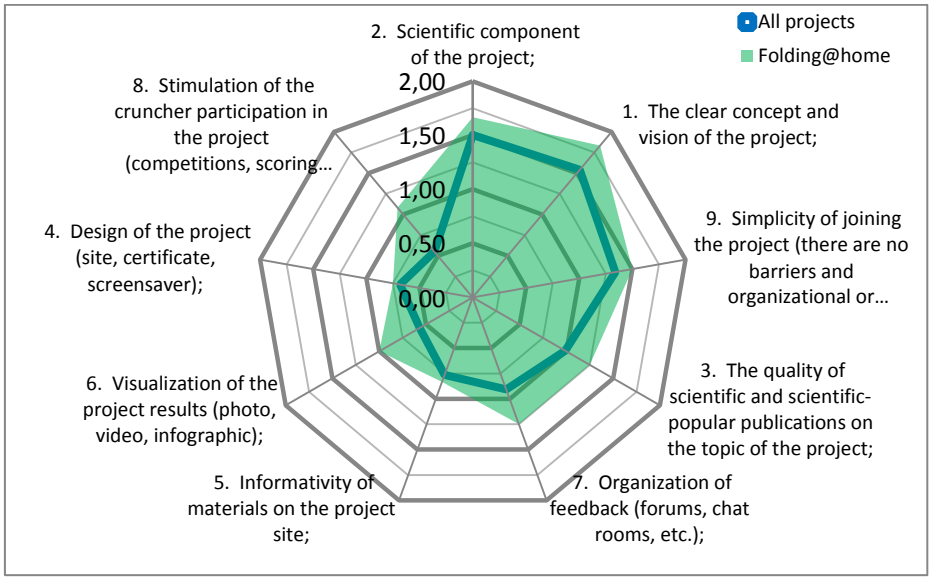

Fig. 6. Evaluation of the characteristics of the project Folding@home

3.2 Einstein@home

Einstein@home project is supported by the American Physical Society, the US National Science Foundation, the Max Planck Society. Einstein@home organized to

search for weak astrophysical signals from spinning neutron stars using data from the LIGO gravitational-wave detectors, the Arecibo radio telescope, and the Fermi gamma-ray satellite [11].

The project Einstein@home is one of the largest international VDC projects on the BOINC platform. The project is active now. About 500 thousand users were connected to the project. In fact, the Einstein@home project is a standard project, as the evaluation of the significance of the characteristics weights coincides with the average values for all projects (Fig.7) and the project estimates by characteristics are higher than the values of the averaged weights for all projects except for one - stimulation of the participation of the crunchers in the project (Fig.8).

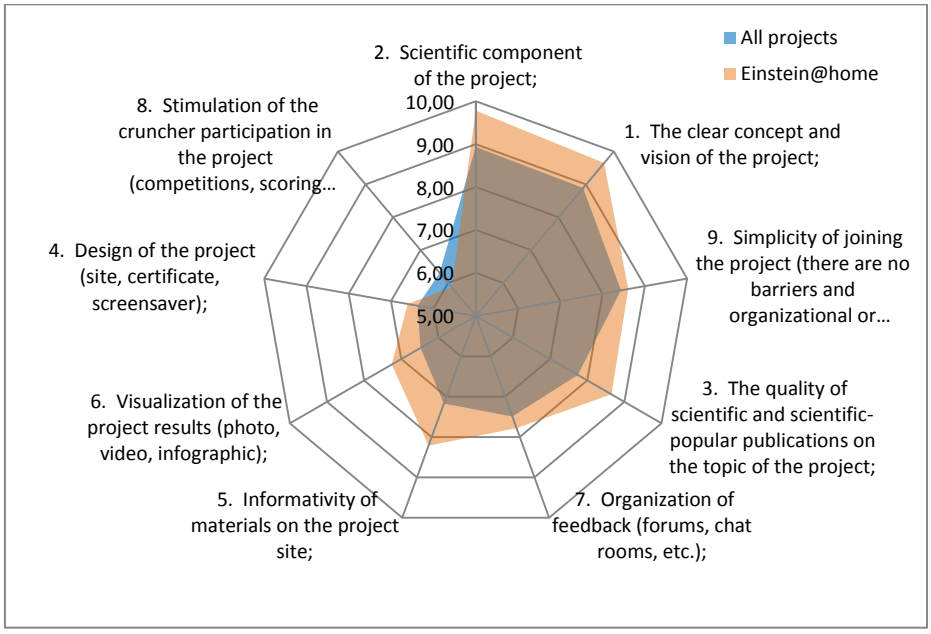

Fig. 7. Weights of characteristics for the project Einstein@home

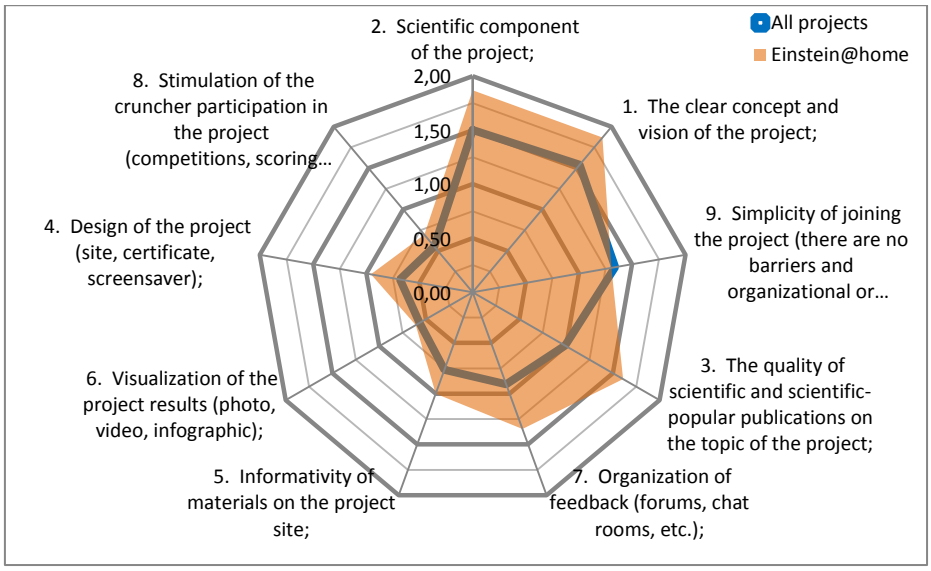

Fig. 8. Evaluation of the characteristics of the project Einstein@home

3.3 LHC@home

LHC@home project help physicists from CERN compare theory with experiment, in the search for new fundamental particles and answers to questions about the Universe [12]. In this project, several different computing tasks are being solved. The project has a large audience of about 160,000 volunteers. At present the project successfully functions and develops. The scale and importance of the scientific problems being solved even reduce the requirements of the volunteer community for such a characteristic as the stimulation of the participation of the crushers in the project (Fig.9). Most of his estimates are above the average for all projects (Fig.10).

12

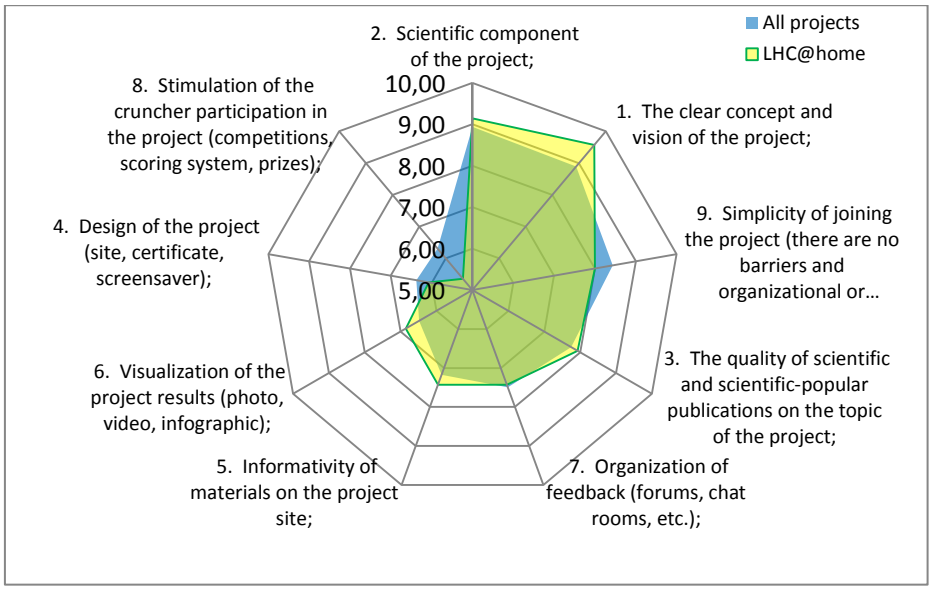

Fig. 9. Weights of characteristics for the project LHC@home

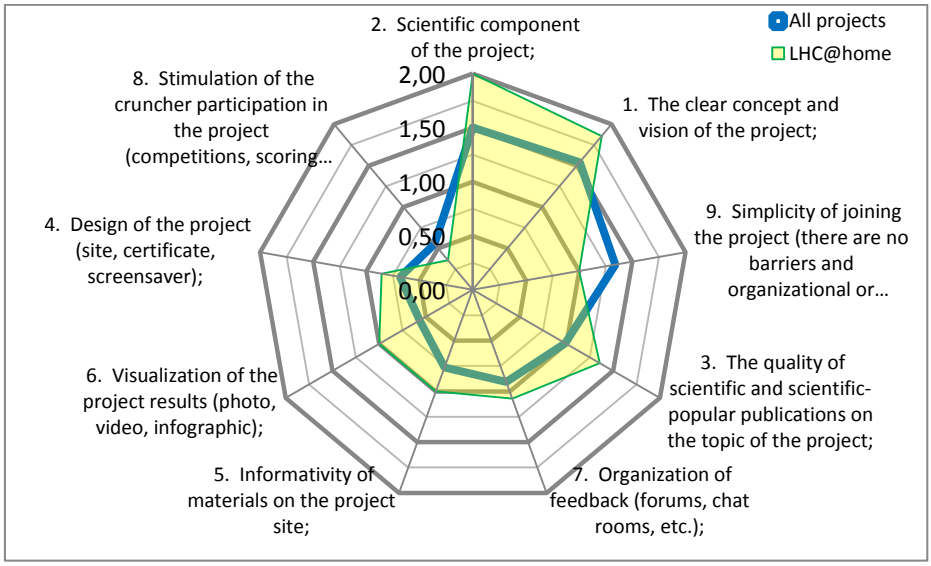

Fig. 10. Evaluation of the characteristics of the project LHC@home

3.4 Gerasim@home

The Russian project Gerasim@home is notable for the fact that the server part of the project is implemented for the MS Windows operating system, but the standard client part of the BOINC platform is used [13]. Number of users about 4000.

It is worth noting significant differences in the weights of characteristics in the project Gerasim@home and for all projects (Fig.11). A significant failure in submitting publications for the volunteer community and feedback problems are clearly visible in (Fig.12) And may be the project's primary improvements for project organizers.

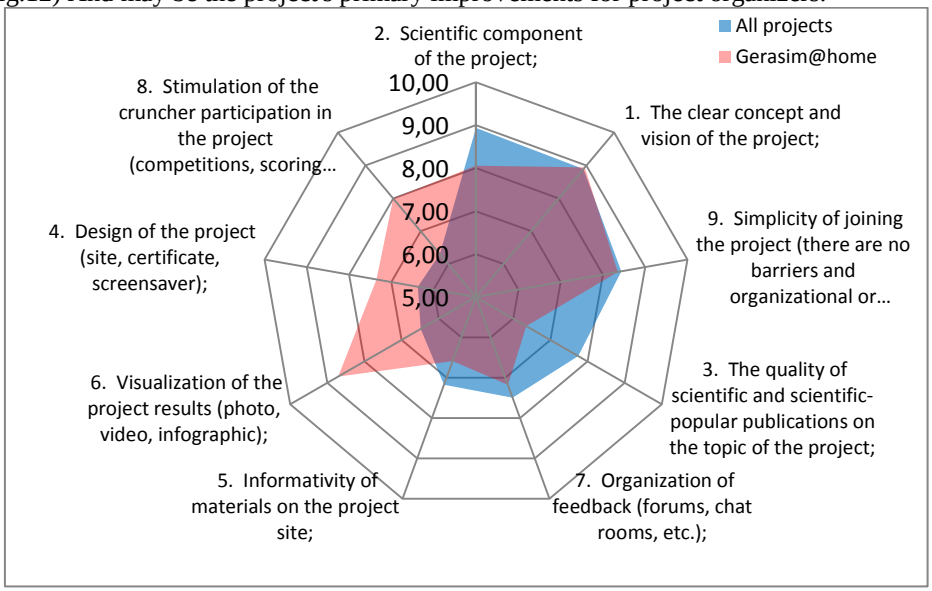

Fig. 11. Weights of characteristics for the project Gerasim@home

14

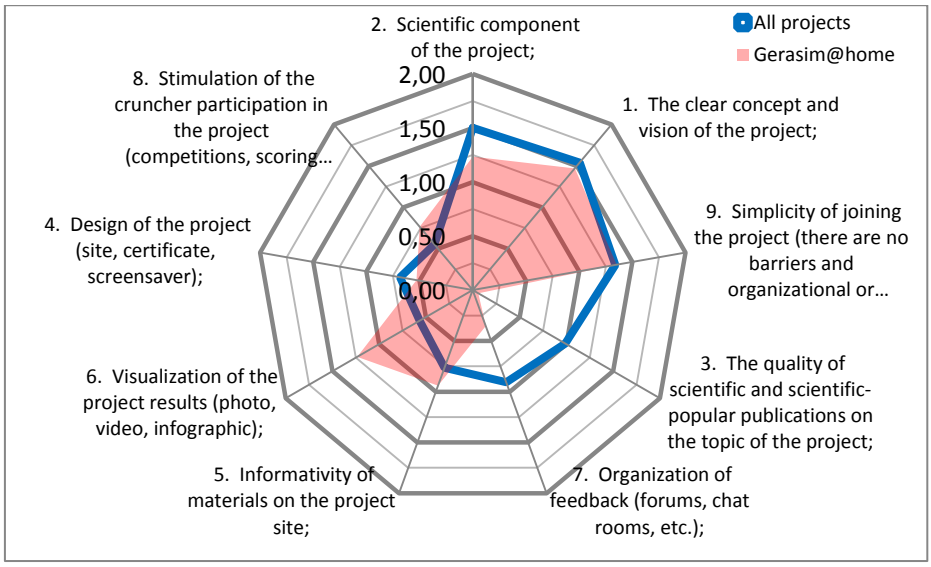

Fig. 12. Evaluation of the characteristics of the project Gerasim@home

4 Performance of Calculations of the YaK-index for Selected VDC Projects

Using the calculated average (for the considered subset of projects) and individual values of the weights of the characteristics of the VDC projects for each project separately, as well as expert estimates of the characteristics, the values of the YaK-index were calculated (Table 1).

Table 1. YaK-index.

№	Project Title	YaK-index (with average weights of characteristics)	YaK-index (with individual weights of characteristics)
1	POGS@home	0.66	0.69
2	Einstein@home	0.65	0.69
3	Folding@home	0.65	0.68
4	LHC@home	0.64	0.64
5	Asteroids@home	0.62	0.62
6	Rosetta@home	0.61	0.63
7	SETI@home	0.60	0.61
8	MilkyWay@home	0.60	0.61
9	Gerasim@home	0.59	0.61
10	SAT@home	0.58	0.57

Judging by the magnitude of the YaK-index values, all 10 VDC projects have a certain capabilities for development. So, in the case of using individual characteristics scales (right column), the highest values of the index (0.69) have both the Einstein@home projects, and the POGS@home project, and the lowest values has the SAT@home project. The teams of the first two projects, referring to the values of the characteristics weights, can determine which characteristics of their projects they should pay attention to in order to increase the values of the YaK-index.

For the project Einstein@home, judging by Fig. 7 and Fig. 8, the growth of the values of the YaK-index is possible when a visualization of the project results and a stimulation of the participation of the crunchers in the project are improved.

The lowest values of the YaK-index in both cases (right and second from the right columns of the table) were obtained by the Gerasim@home and SAT@home projects. This is just a bit over half of the maximum possible values of the index. It is clear that this project is "younger" than the rest ones. Nevertheless, by comparing the radar diagrams with weights and estimates of characteristics, we can recommend to the project teams to pay attention to the scientific component and description of the project's concept, to visualize the results and design the project site.

5 Conclusions

The proposed approach to assessing the state of VDC projects through a specially organized survey of project participants and calculation of the values of the developed YaK-index allows teams to obtain in visualized form comparable information on 9 significant characteristics. These materials can be an important tool for project teams to improve work and project management.

The estimated values for a number of well-known VDC projects show the adequacy of the proposed approach, clearly illustrate the advantages of well-organized projects with a solid operating experience, and create certain conditions for monitoring the process of formation and development of newly created and "young" VDC projects.

Acknowledgements

This work was funded by Russian Science Foundation (№16-11-10352).

References

1.M.J. Litzkow, M. Livny, M.W. Mutka “Condor-a hunter of idle workstations”, Distributed Computing Systems, IEEE,1988.
2.Grimshaw, A. S., & Wulf, W. A. (1997). The Legion vision of a worldwide virtual computer. Communications of the ACM, 40(1), 39-45.
3.D.P. Anderson “BOINC: a system for public-resource computing and storage”, Grid Computing, IEEE, 2004.

4. The server of statistics of volunteer distributed computing projects on the BOINC platform. http://boincstats.com.
5. Homepage of the Folding@home project https://foldingathome.org/, (date of access: 30.04.2018).
6. Homepage of the Einstein@home project https://einsteinathome.org/ru/, (date of access: 30.04.2018).
7. Homepage of the LHC@home project http://lhcathome.web.cern.ch/, (date of access: 30.04.2018).
8. Vatutin E.I., Titov V.S. Volunteer distributed computing for solving discrete combinatorial optimization problems using Gerasim@home project // Distributed computing and grid-technologies in science and education: book of abstracts of the 6th international conference. Dubna: JINR, 2014.
9. Posypkin, M., Semenov, A., & Zaikin, O. (2012). Using BOINC desktop grid to solve large scale SAT problems. Computer Science, 13(1), 25. (date of access: 30.01.2018).
10. Lombraña González, D., Harutyunyan, A., Segal, B., Zacharov, I., McIntosh, E., Jones, P. L., ... & Grey, F. (2012). LHC@Home: A volunteer computing system for massive numerical simulations of beam dynamics and high energy physics events. In Conf. Proc. (Vol. 1205201, No. IPAC-2012-MOPPD061, pp. 505-507).
11. Zaikin, O.S., Posypkin, M.A., Semenov, A.A., Khrapov, N.P.: Experience in organizing volunteer computing: a case study of the OPTIMA@home and SAT@home projects. Vestnik of Lobachevsky State University of Nizhni Novgorod, No. 5–2, pp. 340–347 (2012).
12. V. I. Tishchenko, A. L. Prochko. Russian participants in BOINC-based volunteer computing projects. The activity statistics. Computer Research and Modeling, 2015, vol. 7, no. 3, pp. 727–734
13. Clary EG, Snyder M, Ridge RD, Copeland J, Stukas AA, Haugen J, Miene P. Understanding and assessing the motivations of volunteers: a functional approach. Journal of personality and social psychology. 1998 Jun;74(6):1516.
14. Webpage of World Community Grid project. 2013 Member Study: Findings and Next Steps https://www.worldcommunitygrid.org/about_us/viewNewsArticle.do?articleId=323&utm_source=email&utm_medium=email&utm_campaign=user_study_20130815 (date of access: 30.01.2018).
15. Yakimets V.N., Kurochkin I.I.. Voluntary distributed computing in Russia: a sociological analysis. Collection of scientific articles of the XVIII Joint Conference "Internet and Contemporary Society" (IMS-2015), St. Petersburg, June 23, 2015, St. Petersburg: ITMO University, 2015. Sc.345-352. ISBN 978-5-7577-0502-6.
16. Kurochkin I.I., Yakimets V.N. Evaluation of the voluntary distributed computing project SAT@home. National Supercomputer Forum (NSCF-2016), Pereslavl-Zalessky, Russia, November 29 - December 2, 2016.