



Technical Guidelines of the Technology Contests for the Purposes of the National Technology Initiative “Mutual Learning in Russian¹” and “Mutual Learning in English” (Up Great)²

¹ Technology contests “Mutual Learning in Russian” and “Mutual Learning in English” are also referred to as the Technology contest READ//ABLE.

² This document is the translation of the Russian version of «Technical Guidelines of the Technology Contests for the Purposes of the National Technology Initiative “Mutual Learning in Russian” and “Mutual Learning in English” (Up Great)», and may contain the references to documents that may be available only in Russian. In case of any queries on these documents, compliance with the requirements or other questions, please contact the organizing committee by email: ai@upgreat.one

Content

1. General information	3
2. Terms and definitions	3
3. Participants and Terms of Participation.....	6
4. Cycles and Stages.....	6
5. Contest Task and Assessment of the Accuracy of the Participants' Solutions	7
6. Qualification Stage	8
7. Final Test stage	9
8. Judging Panel and Technical Supervision Commission	10
9. Rights of the Organizing Committee.....	11
10. Protests and Appeals	11
11. Participants are not Allowed the Following.....	11
12. Safety Control and Additional Terms	12
13. Appendices.....	12
Appendix 1. Marking accuracy metrics.....	13
Appendix 2. Program Modules.....	16
1. Software modules for parsing, comparing, and evaluating markups.....	16
2. Software module for data exchange between the PSP and RVS.....	16
3. Algorithm for calculating the criteria and the final essay grade	17
Appendix 3. Error Classifier and Markup Language	20
1. Data markup language for the contest	20
2. Description of the data markup in JSON format	25
3. Error Classifier	27
Literature	53



1. General information

- 1.1. These Technical Guidelines define the requirements for holding the technology contest READ//ABLE for the purposes of the National Technology Initiative (hereinafter referred to as the Contest), as well as other technical parameters thereof.
- 1.2. The Technical Guidelines are published on the official Contest Website (hereinafter referred to as the Website): <http://ai.upgreat.one/>
- 1.3. These Technical Guidelines are a document detailing the Terms and Conditions of the Contest as published on the Contest Website according to the test regulations and the procedure for determining the result and the winner of the Contest.

This document does not contradict the Terms and Conditions, but solves the problem of detailing the Terms and Conditions at a technical level. The technical details contained herein are a necessary and sufficient description of the Contest's purpose. The Technical Guidelines determine the requirements for holding the Contest Tests, the Participants' Software Packages, the control procedure, and the algorithm for determining the level of teams' developments.

- 1.4. The main purpose of this document is to provide a unified set of documentation and regulations for participation in the Contest that ensures transparency and objectivity for both Participants and external observers.

2. Terms and definitions

- 2.1. These Technical Guidelines, in addition to the terms listed below, contain terms with meanings that are defined in clause 1.10. of the Terms and Conditions of the Technology Contests for the Purposes of the National Technology Initiative “Mutual Learning in Russian” and “Mutual Learning in English”.
- 2.2. **Markup Algorithm** – an artificial intelligence-based algorithm used in the Participant's Software Package to convert a file with an unmarked essay text into a file containing a marked text of the same essay.
- 2.3. **Algorithmic Markup** – a markup of the essay text generated by the Markup Algorithm in the Participant's Software Package.
- 2.4. **DB, Essay Database** – a collection of all essay texts in Russian, both marked and unmarked, which are presented in the form of files organized according to certain rules.
- 2.5. **Data Set (DS)** – a set of text essay files, marked or unmarked.
- 2.6. **Set for Markup Algorithm Training, Training Set, Marked Reference Open DS** – a set of marked essay text files that is provided to the Participants on equal terms and used by Participants to train their Markup Algorithms.
- 2.7. **Set for Markup Algorithm Testing, Test Set, Marked Reference Closed DS** – a set of unmarked text essay files provided to Participants on equal terms for independent assessment of the accuracy of the Markup Algorithms at the Qualification Stage. The Results Verification System (RVS) uses a set of marked files of the same essays, available only to the Contest Operator and not available to the Participants.
- 2.8. **Set for the Final Test, Final Set, Final DS** – a set of unmarked text essay files provided to the Participants on equal terms within the framework of the Final Tests stage. Markup of the Final Set by

experts is carried out after algorithmic markups of all the files from the final set have been received from all Participants.

- 2.9. **Tests, Final Tests** – a limited period of time beginning on the day appointed by the Organizing Committee, when Participants use their developed Software Packages for stream markup of the texts contained in the files of the Final Test Set.
- 2.10. **Contest** – technology contests “Mutual Learning in Russian” and “Mutual Learning in English” (variants of the name “ИПО//ЧТЕНИЕ” and “READ//ABLE” are also used), a technology contest, which aims to find the best solution for the automatic detection of semantic errors in text documents in Russian and English. The software systems are developed by the Contest Participants.
- 2.11. **TC, Terms and Conditions** – the main document that defines the aims, objects, and procedures for the Contest. The document is approved by the Technology Contests’ Commission for the purposes of the National Technology Initiative.
- 2.12. **EC, Error Classifier** – a list of error types and semantic blocks that can be used in expert and algorithmic markups of essay texts. The Error Classifier and markup language are described in Appendix 3 to this Technical Guidelines.
- 2.13. **Leaderboard** – a list of Participants, ranked in descending order of the criterion of relative accuracy of algorithmic markup (RAAM).
- 2.14. **PA, Personal Account** – a section of the user web interface of the Platform, in which the user can access data, analytical information, and the Platform’s services in accordance with their role in the Contest and access rights.
- 2.15. **Essay Score** – a score that is calculated based on the markup of the essay text in accordance with the rules for checking an essay of the relevant type (for example, a score for a mini-essay in social studies within the USE (unified state exam)). The score can be calculated both by expert and algorithmic markups.
- 2.16. **Error** – a fragment of an essay text that has the mandatory parameters of the beginning, end, and type of error according to EC, as well as optional parameters of the error subtype according to EC, comment (to specify the error subtype without reference to the essay text), explanation (to explain the error with reference to the essay text), correction, and tag.
- 2.17. **Contest Parameters** – numerical parameters that affect the values of automatically calculated criteria for the relative accuracy of algorithmic markup (RAAM). The Contest Parameters are announced before the beginning of each cycle and before the Final Tests, and can also be changed by the decision of the Organizing Committee, as agreed with the Judging Panel and the Technical Supervision Commission of the Contest.
- 2.18. **Platform** – a website on the internet, providing access to data, analytical information, and services for all categories of users who are involved in participating and holding the Contest. In particular, the Participants can test the Markup Algorithms, the Experts can mark up the essay texts in the Markup Program, etc. The Platform’s interface is available in both Russian and English.
- 2.19. **Markup Program** – software with a web interface based on the Platform, which allows text files to be marked up, and errors and semantic blocks to be described according to the EC. Access to the full functionality of the Marker is given to the categories of users defined by the Organizing Committee. Participants may view the markup, but may not independently make a markup in the Marker interface.

- 2.20. **PSP, Participant's Software Package** – a package of software and hardware developed by the Participant's Team, which automatically accepts an unmarked essay file as its input and returns a marked-up essay file generated by the Markup Algorithm as its output.
- 2.21. **SCP, Solution Comparison Program** – software with a web interface based on the Platform, which allows the user to visually compare two markup files of the same essay (for example, an algorithmic markup from an expert or two expert markups from different Experts).
- 2.22. **Marked-up File, File Markup** – an essay text file that has been marked up in accordance with the rules described in Appendix 3 to these Technical Guidelines, with the extension *.txt and UTF-8 encoding without specifications. The markup can be expert or algorithmic.
- 2.23. **RVS, Results Verification System** – an automatic verification system that accepts a set of files with an algorithmic markup, compares them with the corresponding expert markup and provides a comparison result in the form of an analytical report. The result of the verification can be visually checked for any file from the set in the Solution Comparison Program.
- 2.24. **Semantic Block** – a fragment of the essay text with the required parameters of the beginning, end, and type of the semantic block, according to the EC. Semantic Blocks are not errors, but their presence can affect the grade for the essay.
- 2.25. **Technological Barrier** – the task of creating a stable, functional Software Package for detecting semantic errors in academic essays, the average accuracy of which is not worse than the average accuracy of a reviewing expert operating within a limited period of time. Overcoming the Technological Barrier is a prerequisite for determining the winner of the Contest.
- 2.26. **Participant, Contest Participant** – a Russian or foreign legal entity or individual, or a collaborative group of such entities, whose application for participation in the Contest has been approved by the Organizing Committee. All the requirements of the Terms and Conditions apply to the Participant as well as to the Team.
- 2.27. **File, File with the Essay Text** – a marked-up or unmarked text file with the extension *.txt and UTF-8 encoding without specifications.
- 2.28. **Cycle** – a single logical block of the Contest, which consists of qualification, the final test, and the summing up stages.
- 2.29. **Expert, Specialist** – a specialist from a general education institution who is involved in the Contest, and who marks the essay text and compiles a marked-up file on the basis thereof.
- 2.30. **Expert Markup** – a markup of an essay text that has been compiled by an expert on the basis of a manual check of the unmarked essay file. The Platform features a special service that automates the expert markup process.
- 2.31. **Essay, Text, Essay Text** – a graded piece of written work by a student, in the form of a detailed and reasoned text regarding a specified problem or topic (for example, an essay written by a student of the 11th grade of a general education school as part of a unified state exam in Russian language, literature, social science, or history).

3. Participants and Terms of Participation.

- 3.1. To participate in the Contest, the Participant submits an application by filling out an electronic form on the Contest Website, in accordance with clause 3.3 and Appendix 1 of TC. The procedure for forming the Participant's Team is described in clause 4 of TC.
- 3.2. The conditions under which the Participant can be disqualified are described in clause 3.4 of TC.
- 3.3. Additional terms for participation are described in clause 3.5 of TC.

4. Cycles and Stages.

- 4.1. The Contest is held as repeated Cycles according to clauses 3.2 and 5.1 of TC. Each Cycle includes the Preliminary Stage (clause 5.2 of TC), the Qualification Stage (clause 5.3 of TC), the Final Tests Stage (clause 5.4 of TC), and the Summing up Stage (clause 6 of TC).
- 4.2. After successfully passing the Qualification Stage, the Participant gets access to the initial data and source codes of the common software modules described in Appendix 2 via their Personal Account on the Platform, which should be used by all Participants:
 - 4.2.1. The initial data is a set of the marked-up essay texts for training.
 - 4.2.2. A software module for parsing, comparing, and evaluating markups.
 - 4.2.3. A software module for exchanging data with the Platform.
 - 4.2.4. A pilot application that implements the minimum set of PSP functions.
- 4.3. The use of common software modules is mandatory, as it ensures equal conditions for all Participants. Participants can't make changes to the common software modules themselves. If any errors are detected therein, the Participant can submit an application "On Making Changes to the Common Software Modules" to the Organizing Committee by e-mail without requiring a specific form. In the application, the errors found and/or any suggestions for improving the operation of the modules should be indicated.
- 4.4. During the Qualification Stage, the Participant can repeatedly pass two types of qualifications – algorithmic and technical. For this purpose, the PSP opens data exchange sessions with the Platform. The goal of algorithmic qualification is to analyze the accuracy of the Markup Algorithms on an independent Test Set. The purpose of the technical qualification is to test the PSP in real time, as close as possible to the Final Tests.
- 4.5. At the Final Tests stage, the Participant can open and complete only one data exchange session with the Platform.
- 4.6. The functions of qualification and participation in the Final Tests should be implemented in the PSP using a common software module for data exchange with the Platform. To facilitate this technological task, the Participants are provided with a pilot software package with a minimum set of PSP functions.
- 4.7. The Organizing Committee reserves the right to update the Training and Test Sets, as well as the source code of the common software modules both between the Cycles and during the Qualifying Stage, but no later than 10 days before the Final Tests. In the event of an update, the Organizing Committee shall notify all Participants via a publication on the Platform's website, informing them of the purpose and nature of the modifications made (for example, an error in the code has been eliminated, a useful function has been added at the request of the Participants, the volume of marked-up data has been increased, etc.).

- 4.8. The condition for the end of the Contest is one of the Participants overcoming the Technological Barrier according to the results of the Final Tests (clause 6 of TC). If no Team overcomes the Technological Barrier in the current Cycle, then the next Contest Cycle will be started within the established period (clause 6.1.6 of TC). If no Team overcame the Technological Barrier in the last Cycle of the Contest timeline, the Technological Barrier is considered not to have been overcome (clause 6.1.7 of TC).

5. Contest Task and Assessment of the Accuracy of the Participants' Solutions

- 5.1. The contest task is to overcome the Technological Barrier by building a Markup Algorithm based on the Training Set of the Marked-up Essay Files.
- 5.2. A Markup Algorithm will overcome the Technological Barrier if the average accuracy of its algorithmic marking (AAAM) on the Final Set (during the Final Tests) is not worse than the average accuracy of an expert markup (AAEM), as calculated on the basis of expert markups obtained within a limited period of time.
- 5.3. The average accuracy of algorithmic markup (AAAM) is estimated on the basis of a set that has at least two expert markups for each essay file. The AAAM is defined as the weighted average of the paired accuracy of the algorithmic file markup relative to the expert markup of the same essay.
- 5.4. The average accuracy of expert markup (AAEM) is estimated on the basis of a set that has at least two expert markups for each essay file. The AAEM is defined as the weighted average of the paired accuracy of the expert file markup relative to the expert markup of the same essay.
- 5.5. The paired markup accuracy, relative to another markup of the same essay, is calculated according to the algorithm described in Appendix 1 to these Technical Guidelines. The algorithm is based on the calculation of seven metrics, the values of which are averaged with the weights of the metrics $w_1 \dots w_7$.
- 5.6. The weight coefficient used in the weighted averaging of the paired accuracies in AAAM and AAEM is calculated as follows. The maximum paired accuracy of the algorithmic markup for a given essay is taken into account with the weight $(1 - H)$, while the rest of the markups of this essay are taken into account with a weight of H . The minimum paired accuracy of the expert markup for a given essay is taken into account with the weight $(1 - H)$, while the rest of the markups of this essay are taken into account with a weight of H . The *hardness parameter* H takes a value between 0 and 1 and allows the hardness of the criterion to be controlled. The greater H is, the lower the AAAM and the higher the AAEM, and the more difficult it is to overcome the Technological Barrier. For $H = 1$, AAAM and AAEM are calculated as the arithmetic mean of the paired accuracies. Recommended starting value $H = 0$.
- 5.7. The relative accuracy of algorithmic markup (RAAM) is determined from a given set of essays as the ratio $RAAM = AAAM / AAEM * 100\%$. A value for the relative accuracy that is greater than or equal to 100% indicates that the Technological Barrier has been overcome by this Algorithm. Relative accuracy is used when compiling the ratings of the Participants and Algorithms at all Contest stages and when summing up the results.
- 5.8. Metric weights $w_1 \dots w_7$ and stiffness parameter H are the *Contest parameters*, which can also be changed by the decision of the Organizing Committee as agreed with the Judging Panel and the Technical Supervision Commission of the Contest. The Organizing Committee includes the values of these parameters in their announcements at the beginning of each cycle and of the Final Tests. In case of an extraordinary change in the Contest parameters, the Organizing Committee shall notify the Participants at least one week before the change comes into force.

6. Qualification Stage

- 6.1. This section supplements clause 5.3 of TC.
- 6.2. All registered Participants who have passed the Preliminary Stage, received a notification of registration from the Organizing Committee (clause 5.2.4 of TC), developed a Participant's Software Package (PSP), and integrated their own Markup Algorithms therein, can participate in the Qualification Stage.
- 6.3. During the Qualification Stage, the Participant can repeatedly pass two types of qualification – algorithmic and technical by opening data exchange sessions with the Platform.
- 6.4. Algorithmic qualification is carried out in order to check the quality of the markup algorithms of the Training Set or of the independent Test Set, as well as to analyze and compare them. The volume of the Test Set is at least 300 files. To conduct the algorithmic qualification, the Participant can open a data exchange session with the Platform at any time. When opening a session, he/she needs to select the type of set (Training or Test) and a limit on the number of files. The Platform Server sends unmarked files to the PSP and receives marked-up files from the PSP without intentional time delays. The number of sessions is limited to twenty per day.
- 6.5. At the end of the qualifying session, the Participant can view an automatically generated report regarding the algorithmic qualifications in his/her Personal Account. The report shows:
 - 6.5.1. estimates of the average and relative accuracy of the algorithm (AAAM, RAAM):
 - 6.5.1.1. across the entire set,
 - 6.5.1.2. in the context of essay types,
 - 6.5.1.3. in the context of accuracy metrics (see Appendix 1);
 - 6.5.2. a list of all processed essays, ranked according to the user-selectable accuracy metric;
 - 6.5.3. a comparison of an algorithmic markup with an expert markup for any essay that was selected by the user from the list of processed essays via SCP (only if the algorithmic qualification was carried out on the Training Set);
 - 6.5.4. a conclusion regarding the degree of readiness of the PSP for the Final Test.
- 6.6. The purpose of the technical qualification is to test a PSP in real time, as close as possible to the Final Tests. Technical qualification mode is launched by the Platform server automatically according to a schedule, every four hours. The frequency and number of files to be marked up can vary from run to run for testing purposes. Files are randomly selected from the Training Set. The volume of the set is at least thirty files.
- 6.7. At the end of the qualifying session, Participants can view an automatically generated report regarding technical qualification in their Personal Account. The report shows:
 - 6.7.1. the average and maximum file processing time for the Participant's Software Package;
 - 6.7.2. the average and maximum delay time between the moment access to the file was granted on the Platform server and the moment the file was returned at the request of the PSP;
 - 6.7.3. the proportion of files that were canceled due to non-compliance with the speed limits for receiving and processing files;
 - 6.7.4. a conclusion regarding the degree of readiness of the PSP for the Final Test.

- 6.8. Technical information regarding the procedure for organizing a data exchange session with the Platform during algorithmic and technical qualification is given in Appendix 2 to these Technical Guidelines.
- 6.9. The Qualification Stage will be considered to have been successfully completed if the conclusions regarding the degree of readiness of the PSP for the Final Test are positive on at least one occasion for algorithmic and technical qualification.
- 6.10. Based on the results of algorithmic qualification for the Test Set, a qualification leaderboard is compiled. This is a list of Participants, ranked in descending order by RAAM. To compile the ranking, the best RAAM value of all algorithmic qualifications is used, in which the Participant's PSP has processed at least 95% of the Test Set essay files.

7. Final Test stage

- 7.1. This section supplements clause 5.4 of TC.
- 7.2. The Final Test stage is open to all Participants who have successfully passed the Preliminary and Qualification stages.
- 7.3. The Final Tests are conducted on the Final Set of Unmarked Essays (Final DS). The expert markup of these essays is performed within 15 working days after the algorithmic markup has been completed by Participants. Thus, during the Final Tests, the expert markup of the Final Set does not exist.
- 7.4. To participate in the Final Tests, the Participant must open a data exchange session using the Participant's Software Package before the start of the Final Tests.
- 7.5. During the Final Tests, unmarked files from the Final Set become available on a schedule, beginning from the start of the tests and, subsequently, at regular one-minute intervals, in order to ensure equal conditions for all Participants and exclude the possibility of organizing manual data markup. The schedule of the Final Tests (the date and time of the start and the frequency of file uploads) shall be announced at least ten days in advance thereof.
- 7.6. During the data exchange session, the Participant's Software Package (PSP) must promptly download and process unmarked files from the Platform server, and then send files with an algorithmic markup to the Results Verification System (RVS). If the threshold delay time of 1 minute is exceeded (after the file has been shared), the file will be canceled for the respective Participant and won't be taken into account when calculating the average AAAM and RAAM criteria.
- 7.7. If, for a given Participant, the share of canceled files during the Final Tests exceeds 5% of the total volume of the Final Set, then this Participant's solution (the entire set of submitted files) will be canceled in its entirety and won't be part of the summary of the Contest results.
- 7.8. At the end of the session, the Participant can view an automatically generated report, which is similar to the report regarding technical qualification, in his/her Personal Account.
- 7.9. Technical information regarding the procedure for organizing a data exchange session with the Platform during the Final Tests is given in Appendix 2 to these Technical Guidelines.
- 7.10. Upon completion of the processing of files by all Participants of the Final Test, all data exchange sessions are closed, and all files with an algorithmic markup that have been received from Participants are saved in the Essay Database (generating a closed DS with the final results of the Participants). It is at this point that the expert markup process for the Final Set begins. Each essay file is reviewed by two independent experts. In case of a significant discrepancy between their assessments, a third expert is appointed, who

creates a third markup. The rules for the appointment of a third expert depend on the type of essay and are given in Appendix 3. All experts carry out the checking procedure within a limited time (the limit on the time for checking essays may vary depending on the type of essay). The third expert is provided with the data from the two previous checks and an opportunity to compare them using the SCP. The expert markups of the two or three experts are saved in the Essay Database (forming the Closed Final DS) and are used to calculate the AAAM and RAAM criteria for the Final Set.

- 7.11. Upon completion of the expert markup of the Final Set, the results of the Final Stage are summed up within no more than five working days. All Participants are provided with access to the expert markups of the Final Set. Participants can view an automatically generated report, which is similar to the algorithmic qualification report, in their Personal Account.
- 7.12. Based on the results of the Final Stage, a final leaderboard is compiled – a list of Participants ranked in descending order of RAAM, calculated according to the Final Set. Canceled solutions are not included in the final leaderboard. Solutions with a RAAM value of at least 100% are considered to have overcome the Technological Barrier.
- 7.13. The summing-up procedure is described in clause 6 of TC.
- 7.14. According to clause 8.2 of TC, the essays that have been marked up by the PSP during the Tests (but not the software itself, which is the intellectual property of the teams) are transferred to the Organizers under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) open license for subsequent publication on the Contest Website in order to ensure transparency in determining the winners and runners-up.
- 7.15. The procedure for determining the prizes for the winner and runners-up is described in clause 7 of TC.
- 7.16. Procedure for restarting Tests:
 - 7.16.1. Tests can be restarted in case of a massive technical failure. This can occur by the Organizing Committee's decision on the basis of the corresponding protocols of the Judging Panel and the Technical Supervision Commission.
 - 7.16.2. The restart can be scheduled for the next day, which shall be reported separately on the official website.
 - 7.16.3. New files as used in case of a restart. Those that have already been submitted are not counted.
 - 7.16.4. The results for previously submitted files are not counted.

8. Judging Panel and Technical Supervision Commission

- 8.1. This section clarifies clauses 5.3.10.1, 5.4.4, 5.4.14 of TC.
- 8.2. At the Qualification Stage:
 - 8.2.1. The Judging Panel verifies the automatically generated conclusions in terms of their degree of readiness of the PSP for the Final Test based on the results of algorithmic and technical qualification.
 - 8.2.2. The Technical Supervision Commission verifies the test results, which are then submitted to the Organizing Committee for approval.
 - 8.2.3. The Organizing Committee makes the final decisions regarding admission of Participants to the Final Test.

8.3. At the Final Tests stage

- 8.3.1. The Technical Supervision Commission verifies the correctness of the files received from Participants and identifies the conditions and reasons for cancellations, if any.
- 8.3.2. The Technical Supervision Commission controls the process of assigning tasks for the expert markup of the Final Set.
- 8.3.3. At the end of the expert markup of the Final Set, the Judging Panel verifies the automatically generated reports regarding Participants' solutions and the conditions for overcoming the Technological Barrier.
- 8.3.4. The Technical Supervision Commission verifies the test results, which are then submitted to the Organizing Committee for approval.
- 8.3.5. The Organizing Committee makes its decisions based on the results of the Final Test.

9. Rights of the Organizing Committee

- 9.1. This section clarifies clauses 3.1.8, 3.4.1, 5.1.5, 5.4.8, 8.1.4 of TC.
- 9.2. The Organizing Committee reserves the right to update the Training and Test Sets, as well as the source codes of common software modules both between Cycles and during the Qualifying Stage, but no later than 10 days before the Final Tests. In the event of an update, the Organizing Committee notifies all Participants via a publication on the Platform's website, informing them of the purpose and nature of the modifications that have been made.
- 9.3. The Organizing Committee may verify the team's solution in person.

10. Protests and Appeals

- 10.1. This section clarifies clause 6.2 of TC.
- 10.2. If, during the Final Tests, more than 5% of a Participant's files are canceled, then the Participant's solution cannot be included in the results of the Final Tests, without the possibility of filing a protest or appeal. In this case, the Participant is not disqualified and can participate in the subsequent Contest Stages.

11. Participants are not Allowed the Following

- 11.1. This section supplements clause 3.4 of TC.
- 11.2. Participants are not allowed to perform an expert markup of the Test and Final Set manually, using crowdsourcing, or in any other way involving of human labor.
- 11.3. Involving human labor in order to extract any additional data from essay texts in unmarked sets (Test and Final) is prohibited. The analysis and markup of these texts must be carried out by the Participant only by means of an algorithm.
- 11.4. Independently modifying common software modules is prohibited.
- 11.5. Sending a file with an algorithmic markup of the same essay more than ten times at the Final Test stage is prohibited. All subsequent submissions will be ignored and the last submission will be considered final.



- 11.6. It is forbidden at the stage of the Final Tests to send any other information instead of the algorithmic markup of the essay file, including the algorithmic markup of another essay. Such submissions will be ignored.
- 11.7. In case of a violation of the listed prohibitions, the Organizing Committee, at the request of the Judging Panel or the Technical Supervision Commission, may suspend the Team from participating in all subsequent Contest Cycles.

12.Safety Control and Additional Terms

- 12.1. Safety and environmental requirements, restrictions regarding the disclosure and distribution of information, intellectual property rights, and the procedure for changing the conditions of the Contest are described in clause 8 of TC.
- 12.2. The Participants' activities during the course of the Contest should comply with the environmental standards in force in the Russian Federation and with applicable safety requirements.
- 12.3. In case of on-site events, the Organizing Committee of the Contest shall provide the Participants with safety and environmental rules that all members of the Team must familiarize themselves with, sign, and adhere to.
- 12.4. All technical instructions for the operation of the Platform, as well as the parameters and methods of operation of the Platform are the subject of separate documents that are published on the official website of the Contest, or on the Platform or links to them lead from the Site or the Platform. These documents do not affect the essence of the TC and the Technical Guidelines, but describe the technical parameters of the Platform's functionality and are equally available to all Teams, including ensuring equal conditions for Teams' participation.

13.Appendices

- 13.1. Marking accuracy metrics
- 13.2. Program modules
- 13.3. Error Classifier and markup language

Appendix 1. Marking accuracy metrics

Paired accuracy $M(X, Y)$ of X markup relative to Y markup is estimated by metrics $M_1(X, Y) \dots M_7(X, Y)$, which are then averaged with the weights $w_1 \dots w_7$.

Metrics $M_1 \dots M_7$ are measured as a percentage between 0% to 100%. The higher the percentage, the more accurate the match. The final metric of the *paired accuracy* of the X markup relative to the Y markup is calculated as a weighted average of seven metrics:

$$M(X, Y) = \frac{\sum_{i=1}^7 w_i M_i(X, Y)}{\sum_{i=1}^7 w_i}$$

The weight coefficients w_i allow the degree of importance of each metric to be set. An average can only be taken for some of the metrics, then $w_i = 0$ for the rest of the metrics. For example, in the absence of expert estimates for the metric M_7 , or insufficient reliability thereof, it can be excluded from the formula, and reincluded when a sufficient amount of data regarding the explanations appears in the expert markups.

Set of weights $w_1 \dots w_7$ is a parameter of the method for measuring the accuracy of a markup.

To calculate metrics $M_2 \dots M_6$, the fragments are compared in two markups. The fragment matching algorithm is described in section 3 of this Appendix.

Optimistic relative paired accuracy of the algorithmic markup of one essay, when compared with the whole set $\{E\}$ of expert markups of this essay:

$$M_{opt}(A, \{E\}) = \frac{\max_E M(A, E)}{\min_{E, E'} M(E, E')} 100\%,$$

where the maximum in the numerator is taken for all expert markups, the minimum in the denominator – for all pairs of expert markups for this essay.

Average relative paired accuracy of the algorithmic markup of one essay, when compared with the whole set $\{E\}$ of expert markups of this essay:

$$M_{avr}(A, \{E\}) = \frac{\text{avr}_E M(A, E)}{\text{avr}_{E, E'} M(E, E')} 100\%,$$

where the average in the numerator is taken for all expert markups, the average in the denominator – for all pairs of expert markups for this essay.

Relative accuracy of algorithmic markup of one essay, taking into account the hardness parameter H :

$$RAAM = \frac{H \text{avr}_E M(A, E) + (1 - H) \max_E M(A, E)}{H \text{avr}_{E, E'} M(E, E') + (1 - H) \min_{E, E'} M(E, E')} 100\%.$$

The hardness parameter of criterion H can have values from 0 to 1; the smaller H is, the higher and more optimistic the RAAM estimate is. When $H = 0$ $RAAM = M_{opt}$, when $H = 1$ $RAAM = M_{avr}$.

By taking an average of the numerators over a set of essays, we get the *Average accuracy of algorithmic markup*, taking into account the hardness parameter H :

$$AAAM = \text{avr}_{\text{essay}} \left(H \text{avr}_E M(A, E) + (1 - H) \max_E M(A, E) \right)$$

By taking an average of the denominators over a set of essays, we get the *Average accuracy of expert markup*, taking into account the hardness parameter H :

$$AAEM = \text{avr}_{\text{essay}} \left(H \text{avr}_{E, E'} M(E, E') + (1 - H) \min_{E, E'} M(E, E') \right)$$

Relative accuracy of algorithmic markup for a set of essays, taking into account the hardness parameter H :

$$RAAM = \frac{AAAM}{AAEM} 100\%,$$

Thus, RAAM can be calculated both for a single essay and for any set of essays, including the training, test, and final sets.

M1. Essay Grade Prediction Accuracy.

If $K(X)$ and $K(Y)$ – two grades for the essay, in points, calculated according to markups X and Y , respectively. Then *essay grade prediction accuracy*

$$M_1(X, Y) = \left(1 - \frac{|K(X) - K(Y)|}{\max K} \right) \cdot 100\%$$

where $\max K$ is the maximum possible mark for an essay of this type.

The method of calculating the grades $K(X)$ and the maximum grade K depends on the type of essay. Appendix 3 lists formulae for calculating the essay grade in points for different types of essays.

M2. Accuracy and Completeness of the Search for Fragments.

Search accuracy is defined as the proportion of fragments in markup X that have a matched fragment in markup Y .

Search completeness is defined as the proportion of fragments in markup Y that have a matched fragment in markup X .

The aggregated metric of *accuracy and completeness* is defined as their harmonic average value (F_1 -rate).

M3. Prediction Accuracy of the Codes.

Multiplied by 2 proportion of fragments of markup X that have a matched fragment in markup Y with the same fragment code (type of error or semantic block).

M4. Prediction Accuracy for Error Subtypes and Comments.

Multiplied by 2 proportion of fragments of markup X that have a matched fragment in markup Y with the same error subtype or with a comment that is a paraphrase.

M5. Accuracy of Fragment Localization.

Multiplied by 2 the average accuracy of matches of fragments of markup X with their associated fragments in markup Y . The accuracy of coincidence of two fragments is calculated as the Jaccard index – a ratio of the number of words in the intersection in relation to the number of words in the consolidation of the two fragments.

M6. Accuracy of Bug Fixes.

Multiplied by 2 proportion of fragments in markup X that have a fix and a matched fragment in markup Y with the same fix.



M7. Accuracy of Explanations.

Average expert assessment of the explanation accuracy for all fragments of markup X with explanations. This is the only metric that is not based on an automatic comparison with markup Y , but on the expert assessment.

Experts rate each explanation in the verifiable algorithmic markup with a score from 0 to 5 points. The score is multiplied by 20% to obtain an accuracy that is expressed as percentage. The total score consists of the answers to the following questions (yes = 1, no = 0) with regard to this explanation:

- it will probably be clear to the essay author
- correctly explains the nature of the error
- leaves no opportunity for appeal
- refers to the text of the work and specifically to the selected fragment
- solves a pedagogical problem and helps to avoid similar errors in the future

The explanation in the expert markup automatically gets 100%.



Appendix 2. Program Modules

1. Software modules for parsing, comparing, and evaluating markups

The following algorithms are provided to the Participants to ensure equal status. They should be designed and implemented before the start of the Contest.

1. Parsing (sentence analysis) of the markup algorithm:

Input:

marked-up text;

Output:

list of metadescription field values and a list of fragments;

for each fragment:

start and end positions in the original unmarked text,
fragment text,
code(s),
comment,
explanation,
fix,
tag.

2. Algorithm for finding the optimal match between fragments:

Input:

two markups of the same text in the form of two lists of fragments;

Output:

a list of pairs of numbers of the corresponding fragments from the first and second lists.

3. Algorithm for calculating the criteria and the final essay grade:

Input:

list of metadescription field values;

markup in the form of a list of fragments;

Output:

list of criteria values in points;

final grade according to the rules of assessment for this type of essay.

The formulae, according to which the criteria and the final grade for the essay in points are calculated, depend on the type of essay. Appendix 3 to the Technical Guidelines contains formulae for grades, which are given as points, for five types of essays – USE essays in Russian language, literature, social studies, history, and English language.

4. Algorithm for calculating the metric of the paired accuracy of the markup:

Input:

two markups of the same text in the form of two lists of fragments;

Output:

value of the markup paired accuracy;

values of metrics M1–M7.

2. Software module for data exchange between the PSP and RVS

The software module for exchanging data between the Participant's Software Package (PSP) and the Results Verification System (RVS) is provided to all Participants who have passed the Preliminary Stage.

The main purpose of the module is to provide synchronous distribution of unmarked files to the Participants at the Final Test stage in order to exclude the possibility of manual data markup.

The module is also used at the Qualification Stage in two modes: *algorithmic qualification* (when the quality of algorithms is mainly worked out) and *technical qualification* (when the issues of reliability of the PSP and the speed of the algorithms are mainly worked out).

The module allows several scenarios for communication between the PSP and the RVS to be implemented. During the data exchange session, the PSP acts as a client, with the RVS as a server.

A scenario for a data exchange session at the algorithmic qualification stage

- PSP: Request to open a session.
- RVS: Verification of the Participant and, if successful, transfer of permission to open a session.
- PSP: Request for the next file of unmarked essay.
- RVS: Transfer of an unmarked essay file or a set completion message.
- PSP: Submission of the markup file of the same essay.
- RVS: Measurement of the response time, saving the markup, evaluation of the markup quality.
- RVS: Upon completion of the set, an analytical report is generated and saved.

Scenario for a data exchange session at the Final Tests stage

- PSP: Request to open a session (must be received no later than two minutes after the start of Final Tests).
- RVS: Verification of the Participant and, if successful, transfer of permission to open a session.
- PSP: Request for the next unmarked essay file (must be received no later than 10 seconds after access to the file was granted).
- RVS: Transfer of an unmarked essay file or a set completion message.
- PSP: Submission of the markup file of the same essay (must be received no later than 50 seconds after the unmarked file was sent).
- RVS: Measurement of the response time, saving the markup, evaluation of the markup quality.
- RVS: Upon completion of the set, an analytical report is generated and saved.

If the RVS does not obtain a request from the PSP to receive an unmarked essay file within the given time, or if the RVS does not receive a marked file from the PSP within the given time, then the file will be canceled for this Participant and will not be taken into account when calculating the average AAAM and RAAM criteria.

Scenario for a data exchange session at the technical qualification stage

The scenario of the data exchange session during technical qualification at the Qualification Stage may differ from the scenario of the Final Tests only in terms of time delays.

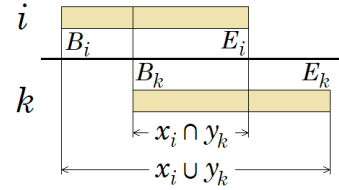
3. Algorithm for calculating the criteria and the final essay grade

The *markup* will be a sequence of n fragments $X = \{x_i = (B_i, E_i, C_i)\}_{i=1}^n$, where B_i is the position of the beginning of the fragment, E_i is the position of the end of the fragment, and C_i is the fragment type.

Matching two markups $X = \{x_i = (B_i, E_i, C_i)\}_{i=1}^n$ and $Y = \{y_k = (B_k, E_k, C_k)\}_{k=1}^m$ is a plurality of M pairs of fragments (i, k) , when each x_i from X matches no more than one y_k and each y_k from Y matches no more than one x_i . If there is no match for fragment x_i , " $x_i \rightarrow \emptyset$ " should be written.

For the discretionary pair of fragments (x_i, y_k) , we define the *Jaccard distance*:

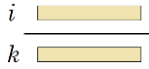
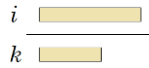
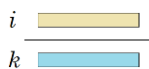
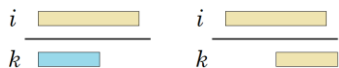
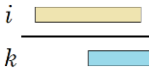

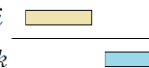
$$J_{ik} = 1 - \frac{|x_i \cap y_k|}{|x_i \cup y_k|}$$



The Jaccard distance J_{ik} can have a value between 0 and 1. If fragments x_i and y_k match, then $J_{ik} = 0$. If the fragments do not intersect, then $J_{ik} = 1$.

Define a *loss matrix* $L[i, k]$ of $n \times m$ size:

$$L[i, k] = J_{ik} + [J_{ik} = 1] + [B_i \neq B_k] + [C_i \neq C_k].$$

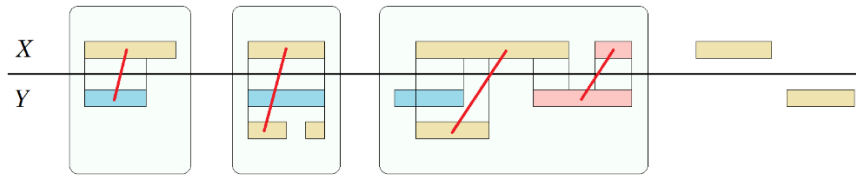
$L[i, k]$	Examples of situations of matching a pair of fragments (x_i, y_k)	
0	fragments match and have the same type	
0..1	fragments intersect, have a common origin and the same type	
1	fragments match, but have different types	
1..2	fragments overlap and have either a different origin or different types	
2	<i>fragments don't match</i>	
2..3	fragments overlap, have a different origin and different types	
3	fragments don't intersect and have the same type	
4	fragments don't intersect and have different types	

We need to find a match between markups that minimizes the amount of losses:

$$Q(M) = \sum_{(i,k) \in M} L[i, k] + \sum_i [x_i \rightarrow \emptyset] + \sum_k [y_k \rightarrow \emptyset] \rightarrow \min_M.$$

The task of finding the optimal match between markups is a generalized assignment problem, which can be solved using one of the variants of the Hungarian algorithm. However, in our case, the loss function is designed so specifically that there is a fast search algorithm for finding the optimal match.

It is easier to describe the algorithm in terms of the graph theory. A *bipartite graph* G is plotted, parts of which are multiple fragments X and Y , and intersecting fragments $(i, k) : J_{ik} < 1$ are connected with edges. If a pair of fragments doesn't intersect, then it can't be included in the optimal solution, since it's better not to connect this pair at all. Thus, the problem of minimizing total losses equates to the problem of finding optimal *matching* – a subgraph M of paired adjacent edges in a bipartite graph G .



The figure shows an example of the correlation of two markups: the top markup X consists of 5 fragments, while the bottom one Y consists of 8 fragments. An optimal match consists of 4 edges: 1 fragment from the top markup and 4 from the bottom markup remain unconnected.

The task of finding the optimal match is solved separately for each connected component of the graph, which greatly reduces the search. A complete search of matches is performed within each component. When an edge is transferred from the original graph to the match, the component can split into even smaller connected components, which, in turn, can be processed separately, reducing the search even more. This idea is implemented by a recursive algorithm for a complete search of all matches in connected components of a given bipartite graph. To speed up the search, the edges can be sorted in descending order of the number of edges adjacent to them.

Algorithm:

1. bipartite graph G is set; match M is empty; $Q(M) = n + m$;
2. transfer from G in M all edges (i, k) , for which $L[i, k] = 0$;
3. $M_{\min} := M$; $Q_{\min} := Q(M)$;
4. $\text{sort_matches}(G, M)$;
5. return the best match M_{\min} ;

Function $\text{sort_matches}(G, M)$:

1. split graph G into connected components;
2. for each connected component G' from G :
3. sort edges (i, k) , non-adjacent to M , in descending order of the number of adjacent edges;
4. transfer edge (i, k) from G' into M ;
5. calculate $Q(M)$;
6. if $Q(M) < Q_{\min}$, then remember $M_{\min} := M$; $Q_{\min} := Q(M)$;
7. $\text{sort_matches}(G', M)$;
8. return edge (i, k) from G' into M ;



Appendix 3. Error Classifier and Markup Language

This document describes a method for marking up the texts of academic essays in order to create training and test data samples for holding the NTI Technology Contest “Mutual Learning in Russian” (Up Great) (<https://ai.upgreat.one>).

The methodology is based on the formalization of criteria for checking essays in Russian language, English language, literature, social studies, and history within the Unified State Examination.

1. Data markup language for the contest

The main objective of the Contest is to develop models and algorithms for natural language understanding (NLU) that can overcome *the Technological Barrier, enabling us to search for, explain, and correct semantic errors (including grammatical, speech, logical, factual, and ethical errors) in academic essays at the level of a specialist (checking expert) within a limited time*. In order to create such algorithms, a large set of texts is required, in which errors have been marked up by experts.

The purpose of the formation of a body of marked-up academic essays is to provide initial data for both the Contest itself and subsequent research in the field of natural language understanding.

This document describes the *rules for marking up essays* and the *types of errors* in essays on the Russian language, Russian literature, social studies, and history, which are checked within the Unified State Examination.

The methodology and error classifier are based on the 2020 USE materials.

1.1.General markup principles

The following general considerations were taken into account when developing the markup rules.

1. The main aim of the markup is to indicate the *localization* and *type* of each error in the essay text, which is convenient for both humans and machines.
2. Each type of essay has its own *error type classifier*. Separate classifiers are used for grammatical, speech, logical, factual, and ethical errors.
3. The fragment that localizes the error should be continuous, sufficient to explain the type and nature of the error, and nonredundant. That is to say, it should not capture redundant parts of text that are not related to the specific error.
4. When evaluating some types of essays, it is not only errors that are taken into account, but also the presence of certain *semantic blocks*. In particular, the essays in Russian language, social studies, and history should contain examples. Moreover, the number of examples that are correctly provided affects the assessment.
5. The text markup should be sufficient for automatic scoring according to the formal criteria for checking an essay of this type (except for spelling and punctuation errors, if they don't lead to distortion of the meaning of a statement). If an expert has taken a point off according to one criterion or other, then the markup should contain full information about the reason for taking this point off. This may be either an error or the absence of the semantic blocks that are required for this type of essay.
6. The markup may contain corrections of *non-core errors* that are not taken into account when evaluating this type of essay, but which may make the automatic processing of the text more difficult. For example, in essays on history and social science, typos, spelling, punctuation, grammatical, and speech errors are non-core.
7. Experts should not make any corrections or comments in the texts themselves, except as provided for by the markup rules.

8. The syntax of the markup should not allow any opportunity for the parsing program to confuse elements of the markup with the original essay text. This is ensured through the selection of special characters for highlighting fragments, error types, and expert comments.

1.2. Marking up fragments

The main syntactic construction of the *markup language* is the selection of a text fragment using special brackets “(\ ... \)” or “(* ... *)”, between which special characters can be used “\”, “:”, “>”, “#” in order to highlight the type of error and expert comments.

An example of the “violation of management norms” error, along with a correction, from an essay on the Russian language:

before markup:

Все удивлялись его силой.

after markup:

Все удивлялись его (* Gr. \ силой >> силе*).

Notes on the syntax of the markup language. Special characters “\”, “*” и “(” are selected because inputting them doesn’t require switching case or language on the keyboard. The special character “\” or “*” is placed next to the parentheses so that the markup can’t be confused with the author’s parentheses in the source text. The backslash “\” is used, rather than “/”, since the straight line is sometimes found in texts, for example: “and/or”. Two types of brackets “(\ ... \)” и “(* ... *)” are required so that, in the case of embedded fragments, it would be more convenient for the markup to distinguish external and internal fragments. Opening and closing parentheses should be of the same type, that is to say, you can’t start a fragment with a “(\” parenthesis and close with “*)” a parenthesis.

After the opening parenthesis, “(\” or “(*”, one or two keywords (codes) are recorded, denoting the error type or semantic block and an optional error subtype.

An example of highlighting a “concept” semantic block in a mini-essay on social studies:

before markup:

Cognition is the process of human comprehension of the surrounding reality.

after markup:

(*CONCEPT\ Cognition is the process of human comprehension of the surrounding reality.*)

Errors are highlighted in the same way. An example of the “**E.theorconn**” error type:

(\ **E.theorconn**\ Action is a process of purposeful activity of people. \)

The error types are rigidly fixed in the classifier and are associated with formal assessment criteria. The error can be specified using the error subtype code, which is written after the error code, separated by a space, or using *comment*, which is written after the selected fragment, separated by a backslash “\”:

(\ **E.theorconn idea**\ Action is a process of purposeful activity of people. \)

(\ **E.theorconn idea** \ Action is a process of purposeful activity of people. \The concept is not associated with the main idea. \)

The error subtype or comment briefly explains the error to the student. Each error type in the classifier may contain several subtypes. Each subtype has its own standard comment (that is, *error subtype* is, in fact, an abbreviation for a standard comment). If the expert believes that none of the standard comments are suitable for a particular case, then he/she can write down his/her own comment. The comment should be laconic and enounce a typical error that is encountered in many works. The comment should not refer to the text of the

particular work in question. If both the error subtype and the comment are specified in the markup, then the comment will take precedence and the subtype should be ignored.

If the expert believes that the comment is not enough to explain the nature of the error to the author of the essay, he/she can add a detailed explanation separated by two colons “::”. He/she can also add his/her own correction by separating it with an arrow “>>” composed of two “more” symbols. As distinct from the comment, the explanation should refer to the text of this particular work and specifically to the highlighted fragment.

Example:

(\ **E.theorconn idea** \ Action is a process of purposeful activity of people. \)
 :: Here, it would be necessary to consider the features of the cognitive process. The more general
 concept of activity distracts from the main idea of the statement. \)

To mark an error that is not specific to this particular essay type, a predefined code is used “**CORR**”. A correction separated by an arrow “>>” is mandatory. The comment, explanation, and tag are usually omitted.

Example:

Все удивлялись его (**CORR** \ силой >> силе \).

Selected fragments can be embedded. The most typical case is when a long semantic block is highlighted, within which errors and other semantic blocks can be highlighted. Example:

(* **CONCEPT** \ (\ **E.theorconn idea** \ Action is a process of purposeful activity of people. \The
 concept is not associated with the main idea. \) *)

In such cases, one should carefully monitor the balance of the parentheses: the number of opening parentheses “(” should be equal to the number of closing ones “)”. Fragments enclosed by parentheses can be embedded, but they can’t overlap. Different brackets can be used to visually highlight embedded fragments: “(* ... (\ ... \) ... *)”, however, this recommendation is not mandatory.

Some error types don’t refer to a specific fragment, but to the entire text. In this case, the error can be indicated at the end of the entire text without highlighting the fragment. Examples:

(* **S.topic** *)
 (* **S.topic** :: The topic was not fully covered. To fully cover the topic, it was necessary to ... *)

1.3.Fragment markup: formal definition

A fragment with an error or a semantic block is highlighted with brackets “(\ ... \)” or “(* ... *)”. Inside the brackets and before the fragment, the code is indicated, or indeed several codes separated by spaces. The fragment can be followed by a comment, explanation, correction, and a tag, each of which begins with its own starting special character:

(* Codes \ FragmentText \ Comment :: Explanation >> Correction #Tag *)

Legend:

- **Codes** are one or more keywords (abbreviations) that denote the *semantic block type*, the *error type*, or the *error subtype* according to the classifier; error codes take the format “x.zzz”, where “x” is a letter indicating which type of essay this error belongs to, “zzz” is an abbreviation of the error; when writing the code, the case of characters is not important: you can write **F.fact**, **f.Fact**, **F.FACT**, **f.fact** – all these options are correct;
- **FragmentText** is an unchanged fragment of the original essay text. The boundaries of the fragment should clearly indicate the localization of the error or semantic block. If there is no **FragmentText**, then the error

refers to the text as a whole (it is preferable to avoid this).

- **Comment** is an error subtype that explains the nature of the error to the student in more detail. For typical errors, if possible, standard comments or their abbreviations (error subtypes) are used according to the classifier. There may not be a **Comment**.
- **Explanation** is a detailed comment referring directly to the essay text and specifically to the highlighted fragment. There may not be an **Explanation**.
- **Correction** is the expert's suggested spelling, replacing **FragmentText** as a whole. There may not be a **Correction**.
- **Tag** is a string of letters or numbers, which serves to link several fragments related to the same error or semantic block. There may not be a **Tag**, if the error is localized in one fragment only.

There may be no **\Comment**, **::Explanation**, **>>Correction** or **#Tag** markup elements. If there is no markup element, then its initial special character will be also omitted. Spaces around initial special characters **"\"**, **"::"**, **">>"**, **"#"** are ignored.

The **"#Tag"** structure is only used in (rather rare) cases when an error or semantic block can't be localized in one fragment, and we need to select several fragments and link them to one another. The tag that links them should be unique, that is, different from other tags in the same text.

1.4.Data view format

Each piece of work is saved in a separate text file (with the .txt extension) in UTF8 encoding. The use of other formats, including MS Word (with the .doc or .docx extension), is not permitted.

If spelling, punctuation, grammar, and speech errors are not supposed to be checked for this essay type, then, when transferring the text of the student's work from the handwritten version to a .txt file, these errors, including obvious typos, should be corrected. In these cases, you can use a speller.

The file name should consist of two parts: a unique identifier for the piece of work and a unique identifier of the reviewing expert. Thus, simply using the file name, you can select all the reviews of a particular piece of work and all the pieces of works checked by a particular expert.

1.5.File meta description

At the beginning of the text file, there should be lines with meta description fields. The meta description is separated from the main text by a blank line. Each line of the meta description takes the following form:

Field : Value

where **Value** is text, to the end of the line, and **Field** is one of the predefined keywords:

Field	Field value
Topic	essay topic or the text on which the essay is written
Original text	full text of the assignment according to which this essay was written
Subject	type of essay or name of the subject matter, which defines the possible types of errors: Russian language, literature, social studies, history, English language, English-fluent, Russian-fluent.
Line	area related to the topic of the mini-essay; for example, for social studies, this is one of six lines: philosophy, economy, sociology, social psychology, politology, legal studies
Class	String values, e.g. «11», «5», «2 course»

Year	a four-digit number (e. g. 2018, 2019, 2020). The year is important, since the rules for reviewing this type of essay may be different for different years
Test	the type of exam, which determines the criteria for evaluating a piece of work: GSE, USE, training etc.
Expert	the unique identifier of the expert who reviewed the given work
CN	assessment of work by criterion number <i>N</i>

If **Value** contains several lines (for example, the text about which the essay has been written in Russian), then it will be enclosed in brackets “(\ ... \)” or “(* ... *)”:

Field: (* Value *)

The value of any field can be empty.

Example of a meta description:

Topic: (* What are the strengths and weaknesses of Bazarov’s nihilism? *)
 Class: 11
 Year: 2020
 Subject: literature
 Test: USE training
 Expert: SeverusSnape2020

1.6.Syntax errors

A syntactic parser is a program that accepts a marked up .txt file of an essay as its input and returns a list of metadescription fields and a list of fragments as its output. The parser can detect the following errors in a markup (the parser operation is indicated in parentheses):

1. Unknown metadescription field (this field is ignored).
2. Unknown code, or separator forgotten “\” (this code is taken as the beginning of the fragment).
3. Fragment code is not specified (a fragment with an empty code is created).
4. Opening parenthesis without a corresponding closing one (it is considered that the closing parenthesis is at the end of the text).
5. Closing parenthesis without a corresponding opening one (the closing parenthesis is ignored).
6. The closing parenthesis doesn’t match the opening bracket (it is considered as matching)
7. The “**CORR**” fragment doesn’t contain the structure “>>Correction” (the correction is ignored).
8. The original essay text is corrupt (warning, ignored).

1.7.Operation algorithm of the expert marker

Expert actions with the essay .txt file that was received for markup.

1. Enter all metadescription fields, for which reliable data is available, in a text file.
2. If there is a semantic block in the text, then it should be highlighted regardless of how correct its content is.
3. If there is an error in the text, based on which the expert is going to reduce the score according to one of the applicable criteria, then he/she should:
 - precisely localize the beginning and end of the faulty text fragment;
 - choose the code of the appropriate error type from the classifier;
 - choose a code of the appropriate error subtype from the classifier; if none of them fits, then write a comment;

- if the nature of the error may be misunderstood by the essay author, then add an explanation and/or correction;
 - if a fragment requires linking with other fragments, then think up a unique tag and put it in all related fragments.
4. Put grades on all formal criteria (C1, C2, etc.) and verify that the grades, as a points score, are uniquely calculated from the markup according to the rules for checking this essay type.

1.8. Visual markup interface

When marking up an essay file in the visual web interface, the following rules apply.

1. The semantic block is marked with a light green background color.
2. The error is marked with a light red background color.
3. The error combined with a semantic block is marked with a light yellow background color.
4. Embedded blocks of the same color are marked with an increased intensity of the background color.
5. “**CORR**” corrections are marked with a light blue background color.
6. The linked fragments are marked at the end of the fragment with a tag similar to a footnote.
7. The window for entering the fragment attributes (type and subtype codes, comment, explanation, correction, and tag) are displayed in the fields by clicking on this fragment.
8. If a fragment combines several semantic blocks and errors, then when clicking on this fragment, the attribute windows for all error types will appear.

2. Description of the data markup in JSON format

JSON format is used to represent the marked-up essay texts in a format that is more convenient for computer processing.

Description of JSON format fields

Field name	Data type	Empty value	Compliance with the syntactic structures of the markup language
meta	object	not permitted	Metadescription
meta.id	string	not permitted	Unique ID of the text, which is used to contact the support team. Can be empty for non-public texts.
meta.uuid	string	not permitted	Unique technical text ID
meta.theme	string	not permitted	Theme
meta.third	boolean	not permitted If there is no field in the file, the algorithms consider it equal to <i>false</i>	Indication that the expert has marked the data in the viewing mode of the previous two markings
meta.class	string	empty string	Class
meta.year	number	not permitted	Year
meta.subject	string	not permitted	Subject permitted values: rus, eng, lit, social, hist, rus-free, eng-free
meta.taskText	string	empty string	Original text - the text of the task for which the essay was written
meta.category	string	empty string	Line
meta.expert	string	can be omitted or be an empty string	Expert is a required field for files with an expert markup
meta.test	string	not permitted	Test

criteria	array	permitted	CN
selections	object	missing from unmarked essay files	sequence of marked-up fragments
selections[N].id	number	not permitted	fragment number
selections[N].startSelection	number	not permitted	position of the beginning of the fragment, character by character, starting from zero, relative to the text. Line breaks are counted as a character.
selections[N].endSelection	number	not permitted	position of the end of the fragment, character by character
selections[N].comment	string	empty string	Comment
selections[N].explanation	string	empty string	Explanation
selections[N].correction	string	empty string	Correction
selections[N].tag	string	empty string	Tag
selections[N].group	string	not permitted	fragment type: error — mistake meaning — semantic block
selections[N].type	string	not permitted	code of the error type or semantic block according to the classifier
selections[N].subtype	string	empty string	error subtype code according to the classifier
Text	string	not permitted	Essay text written by the student. Has no leading or trailing whitespaces. Paragraphs are marked with a line break character in unix format (\n)

Example

```
{
  "meta": {
    "subject": "literature",
    "test": "USE training",
    "category": null,
    "year": 2017,
    "class": "11",
    "theme": "The problem of overcoming fear in ordinary, peaceful life",
    "taskText": "Excerpt from a text by F. A. Vigdorova"
  },
  "selections": [
    {
      "id": 424,
      "tag": "",
      "type": "R.sim",
      "group": "error",
      "comment": "",
      "subtype": "",
      "correction": "",
      "explanation": "",
      "endSelection": 211,
      "startSelection": 53
    },
    {
      "id": 529,
```

```

    "tag": "",
    "type": "CORR",
    "group": "error",
    "comment": "",
    "subtype": "",
    "correction": "Text correction",
    "explanation": "",
    "endSelection": 447,
    "startSelection": 330
  },
  {
    "id": 530,
    "tag": "",
    "type": "ARGUMENT",
    "group": "meaning",
    "comment": "",
    "subtype": "",
    "correction": "",
    "explanation": "Test explanation",
    "endSelection": 627,
    "startSelection": 523
  }
],
"text": "Is it true that one of the hardest trials in human life is the test of fear? In the text proposed for analysis the author raises a problem..."
}

```

3. Error Classifier

The error classifier that is corrected and taken into account when assessing student work is based on the methodological recommendations of the Federal Institute of Pedagogical Measurements (FIPM) [1] and published demos of control measurement materials from the Federal Institute of Pedagogical Measurements (FIPM) [2, 3, 4, 5].

The classifier describes both types of errors and types of semantic blocks.

3.1. Grammatical errors

Grammatical errors (Γ) – these are errors in the structure of a linguistic unit: words, phrases or sentences, that is, a violation of any grammatical norm - derivational, morphological, and syntactic.

The errors described in this section refer to essay types **Russian, Russian-free**.

Code	Standard comments	Examples
Г1 Г.слов	<p>словообр Ошибочное словообразование.</p> <p>сущ Ошибочное образование формы существительного.</p> <p>прил Ошибочное образование формы прилагательного.</p> <p>числ Ошибочное образование формы числительного.</p> <p>мест Ошибочное образование формы местоимения.</p> <p>глагол Ошибочное образование формы глагола.</p> <p>прич Ошибочное образование формы причастия.</p> <p>деепр Ошибочное образование формы деепричастия.</p>	<p>Благородность, чуда техники, подчерк, надсмехаться; более интересное, красивше; с пятистами рублями; жонглировал обоими руками, ихнего пафоса, вокруг его ничего нет; сколько нравственных принципов мы лишились из-за утраты духовности;</p> <p>им (Г.слов глаг\ двигает >>движет\) чувство сострадания; ручейки воды, (Г.слов прич\ стекаемые >>стекающие\) вниз, поразили автора текста; (Г.слов деепр\ вышел >>выйдя\) на сцену, певцы поклонились.</p>
Г2 Г.согл	Нарушение норм согласования.	Я знаком с группой ребят, серьезно увлекающимися джазом.
Г3 Г.упр	<p>упр Нарушение норм управления.</p> <p>сущ Неверный выбор варианта падежного окончания имени существительного.</p> <p>пфс Неверный выбор полной или краткой формы прилагательных в функции сказуемого</p>	<p>Нужно сделать природу более красивую. Все удивлялись его (Г.упр\ силой>> силе\).</p> <p><i>В московских автобусах снова работают кондуктора.</i></p> <p><i>Статья интересная по форме и содержанию.</i></p>
Г4 Г.сказ	<p>подл Нарушение связи между подлежащим и сказуемым.</p> <p>спосвыр Нарушение способа выражения сказуемого.</p>	<p>Главное, чему теперь я хочу уделить внимание, это художественной стороне произведения.</p> <p>Он написал книгу, которая эпопея.</p> <p>Все были рады, счастливы и веселые.</p>
Г5 Г.однор	Ошибка в построении предложения с однородными членами.	Страна любила и гордилась поэтом. В сочинении я хотел сказать о значении спорта и почему я его люблю.
Г6 Г.деепр	Ошибка в построении предложения с деепричастным оборотом.	Читая текст, возникает такое чувство сопереживания.
Г7 Г.прич	Ошибка в построении предложения с причастным оборотом.	Узкая дорожка была покрыта проваливающимся снегом под ногами.
Г8 Г.сложн	Ошибка в построении сложного предложения.	Эта книга научила меня ценить и уважать друзей, которую я прочитал еще в детстве. Человеку показалось то, что это сон.
Г9 Г.смешен	Смещение прямой и косвенной речи.	Автор сказал, что я не согласен с мнением рецензента.
Г10 Г.границ	Нарушение границ предложения.	Его не приняли в баскетбольную команду. Потому что он был невысокого роста.
Г11 Г.видовр	Нарушение видовременной соотнесенности глагольных форм.	Замирает на мгновение сердце и вдруг застучит вновь.

G12 Г.эллипс	проп Пропуск члена предложения. элл Неверное употребление эллипсиса.	На собрании было принято (?) провести субботник.
G13 Г.частиц	част Ошибка в употреблении частицы. отрыв Отрыв частицы от того компонента предложения, к которому она относится.	Хорошо было бы, если бы на картине стояла бы подпись художника. В тексте всего раскрываются две проблемы.

3.2.Speech errors

Speech errors (P) – these are errors not in the construction of a sentence, not in the structure of a linguistic unit, but in its usage, most often in the use of a word, that is, a violation of lexical norms. These are pleonasm, tautology, speech clichés, inappropriate use of colloquial vocabulary, dialecticisms, jargon; expressive means, nondiscrimination of paronyms. Errors in the use of homonyms, antonyms, synonyms, ambiguity not eliminated

The errors described in this section refer to essay types **Russian, literature, Russian-free**.

Code	Standard comments	Examples
P1 P.знач	несвой Употребление слова в несвойственном ему значении. термин Неправильное употребление термина, терминологическая ошибка.	Мы были шокированы прекрасной игрой актеров. Благодаря пожару, лес сгорел.
P2 P.прост	Неоправданное употребление диалектных и просторечных слов.	Таким людям всегда удастся объегорить других. Обломов ничем не занимался и целыми днями валял дурака.
P3 P.мест	Неудачное употребление местоимений.	Текст написал В. Белов. Он относится к художественному стилю; У меня сразу же возникла картина в своем воображении.
P4 P.стиль	стиль Употребление слов иной стилистической окраски. эмоц Неуместное употребление экспрессивных, эмоционально окрашенных слов. эпохи Смешение лексики разных эпох. устар Неуместное употребление устаревшей лексики. неол Неуместное употребление неологизма. канц Неуместное употребление канцелярита. жарг Неуместное употребление жаргонизма. флог Неуместное употребление фразеологизма.	По задумке автора, герой побеждает; Молчалин работает секретарем Фамусова; В романе А.С. Пушкина имеют место лирические отступления; Автор то и дело прибегает к употреблению метафор и олицетворений. Если бы я был там, то за такое отношение к матери я бы этому кексу в грызло бы дал; Зощенко палец в рот не клади, а дай только посмешить читателя.

P5 Р.прист	Неразличение оттенков значения. прист Неразличение оттенков значения, вносимых в слово приставкой.	В таких случаях я взглядываю в словарь.
P5 Р.суфф	Неразличение оттенков значения. суфф Неразличение оттенков значения, вносимых в слово суффиксом.	
P6 Р.оним	син Неразличение синонимичных слов. пар Неразличение паронимов. ант Ошибка в употреблении антонимов при построении антитезы. фрлог Разрушение образного значения фразеологизма в неудачно организованном контексте.	Были приняты эффектные меры; Имя этого поэта знакомо во многих странах; В третьей части текста не веселый, но и не мажорный мотив заставляет нас задуматься; грампластинка не сказала еще своего последнего слова.
P7 Р.сочет	Нарушение лексической сочетаемости.	Автор использует художественные особенности.
P8, P12 Р.лишн	лишн Употребление лишних слов. плеон Плеоназм. избыт Лексическая избыточность. расщ Расщепление сказуемого. параз Слова-паразиты. сравн Объединение простой и сложной форм сравнения	Молодой юноша; очень прекрасный; более оптимальный; вести борьбу. Ученики приняли решение произвести уборку школьного двора. Тогда о том, чтобы вы могли улыбнуться, об этом позаботится книжный наш магазин. Обломов был избалованным, ну, в общем, ему в детстве все пытались угодить. Сейчас более худшее положение, чем ранее.
P9 Р.тавт	Употребление рядом или близко однокоренных слов (тавтология).	В этом рассказе рассказывается о реальных событиях.
P10 Р.повтор	Неоправданное повторение слова.	Герой рассказа не задумывается над своим поступком. Герой даже не понимает всей глубины содеянного им.
P11 Р.бедн	Бедность и однообразие синтаксических конструкций.	Когда писатель пришел в редакцию, его принял главный редактор. Когда они поговорили, писатель отправился в гостиницу.
Р.неполн	Лексическая неполнота высказывания. Пропуск необходимого в предложении слова.	Достоинство Куприна в том, что ничего лишнего.
Р.двусм	двусм Двусмысленность. омон Двусмысленность при употреблении омонимов или многозначных слов.	Экипаж находится в отличном состоянии. Аудитория не соответствовала требованиям лектора.
Р.шаблон	шаблон Неверная шаблонная фраза. употр Неверное употребление шаблонной фразы. неум Неуместное употребление шаблонной фразы.	В пример трусости и отсутствия любви к родине автор приводит двух молодых людей.

3.3.Logical errors

Logical errors (Л) –associated with a violation of the logical correctness of speech. They arise as a result of a violation of the laws of logic, admitted both within the limits of one sentence, judgment, and at the level of the whole text.

The errors described in this section refer to essay types **Russian, Russian-free**.

Code	Standard comments	Examples
Л1 Л.неоднор	Сопоставление (противопоставление) двух логически неоднородных (различных по объему и по содержанию) понятий.	На уроке присутствовали директор, библиотекарь, а также Анна Петровна Иванова и Зоя Ивановна Петрова; Он облокотился спиной на батарею; За хорошую учебу и воспитание детей родители обучающихся получили благодарственные письма от администрации школы.
Л2 Л.причслед	Нарушение причинно-следственных отношений.	В последние годы очень много сделано для модернизации образования, однако педагоги работают по-старому, так как вопросы модернизации образования решаются слабо.
Л3 Л.скачок	Пропуск звена в объяснении, «логический скачок».	Людской поток через наш двор перекрыть вряд ли возможно. [?] А как хочется, чтобы двор был украшением и школы, и поселка.
Л4 Л.перест	Перестановка частей текста.	Пора вернуть этому слову его истинный смысл! Честь... Но как это сделать?
Л5 Л.подмена	Неоправданная подмена лица, от которого ведется повествование.	Автор пишет о природе, описывает природу севера, вижу снега и просторы снежных равнин.
Л6 Л.несопост	Сопоставление логически несопоставимых понятий.	Синтаксис энциклопедических статей отличен от других научных статей.
Л.повтор	необосн Необоснованный повтор в логических рассуждениях. неразв Мысль повторяется и не развивается.	В предложенном для анализа тексте В.М. Песков поднимает проблему роли детских впечатлений в формировании личности. (Л.повтор\ Именно над ней он и размышляет.\)
Compositional and text errors		
Л7 Л.зачин	Неудачный зачин.	Текст начинается предложением, содержащим указание на предыдущий контекст, который в самом тексте отсутствует, наличием указательных словоформ в первом предложении, например: В этом тексте автор...
Л8 Л.основн	Ошибки в основной части.	а). Сближение относительно далеких мыслей в одном предложении. б). Отсутствие последовательности в изложении; бессвязность и нарушение порядка предложений. в). Использование разнотипных по структуре предложений, ведущее к затруднению понимания смысла.

Л9 Л.конц	Неудачная концовка.	Дублирование вывода, неоправданное повторение высказанной ранее мысли.
Л.абзац	Нарушение абзацного членения, текст неверно разделен на микротемы, абзацы. слито перед данным предложением необходимо начать новый абзац разд данное предложение должно продолжать предыдущий абзац	(*ПОЗИЦИЯ \Абрамов считает, что сохранению и защите непреходящих духовных ценностей стоит уделять столько же внимания, сколько сейчас уделяют сохранению природной среды или памятников материальной культуры. И именно книги позволяют сохранить все эти ценности.*) (\Л.абзац слито\Я согласна с мнением писателя\). (*ОТНОШЕНИЕ\Литература, буквально хранящая в себе людей прошедших эпох, испытания и подвиги наших предков, справляется с ролью нравственного ориентира для молодых поколений лучше всего.*)

3.4.Errors in facts

Errors in facts (Φ) – a kind of non-linguistic errors, which consists in the fact that the author gives facts that contradict reality, gives incorrect information about the factual circumstances, both related and not related to the analyzed text (background knowledge).

The errors described in this section refer to essay types **Russian, Russian-free**.

Code	Standard comments	Examples
Ф1 Ф.искаж	содерж Искажение содержания литературного произведения. толк Неправильное толкование. прим Неудачный выбор примеров.	Базаров был нигилист и поэтому убил старуху топором; Ленский вернулся в свое имение из Англии; Счастьем для Обломова было одиночество и равнодушие.
Ф2 Ф.цит	цит Неточность в цитате. нетавт Отсутствие указания на автора цитаты. автор Неверно названный автор цитаты.	Книга очень много для меня значит, ведь еще Ленин сказал: «Век живи – век учись!»
Ф3 Ф.факт	факт Незнание фактов. время Временное смещение.	Великая Отечественная война 1812 года; Столица США - Нью-Йорк.
Ф4 Ф.назван	герой Неточность в имени литературного героя. произв Искажение в названии литературного произведения. автор Ошибка в указании автора.	Тургенев; «Тарас и Бульба»; в повести Тургенева «Преступление и наказание».

3.5.Ethical mistakes

Ethical mistakes (Э) –violation of the system of values and rules of ethics: statements that humiliate human dignity, expressing an arrogant and cynical attitude towards the human person, malignity, manifestations of speech aggression, slang words and phrases.

The errors described in this section refer to essay types **Russian, Russian-free**.

Code	Standard comments	Examples
Э.агресс	ненапр Ненаправленная речевая агрессия. некорр Речевая некорректность. оскор Грубое, оскорбительное высказывание без явного указания адресата. негат Выражение негативных эмоций, чувств или намерений в неприемлемой в данной речевой ситуации форме. жарг Употребление бранных слов, вульгаризмов, жаргонизмов, аргю.	Этот текст меня бесит; Нужно быть полностью сумасшедшим, чтобы читать книги сегодня; Почему школьная программа принуждает к прочтению всего старья, что именуется классикой? Это настоящий пиар! Нечего морочить людям мозги устаревшими истинами.
Э.направл	напр Направленная речевая агрессия. угроз Угроза, грубое требование, обвинение, насмешка с явным указанием адресата. униж Унижение человеческого достоинства. цин Высокомерное или циничное отношение к человеческой личности.	Мне хотелось бы сделать автору замечание за его неумение передавать свои мысли. Михалков в своём репертуаре! Пишет детские книги, поэтому и требует, чтобы читали именно в детстве.

3.6.Errors in Russian language essays

In an essay based on the text read (task No. 27 in the USE 2020), the student is required to formulate the problem set by the author of the text; include in the comments two examples with illustrations from the text read, explain them; formulate the position of the author and express your attitude to the position of the author [2].

The errors described in this section refer to essay types **Russian**.

Code	Standard comments	Examples
Types of semantic blocks		
ПРОБЛЕМА	Формулировка одной из проблем исходного текста.	(*ПРОБЛЕМА\ Какова роль литературы в жизни человека? Над этим вопросом рассуждает в своем тексте российский писатель и публицист Ф.А. Абрамов. *)
ПРИМЕР	Пример-иллюстрация из исходного текста, важный для понимания проблемы.	(*ПРИМЕР\ Писатель раскрывает эту проблему на примере женщины, которая была приветлива и гостеприимна, но одинока, потому что её родных забрала война: «Пятеро не вернулись у меня с войны: сам, трое сыновей и деверь...». *)
ПОЯСНЕНИЕ	Пояснение к примеру, значение примера для понимания проблемы.	(*ПОЯСНЕНИЕ\ Для этой женщины "изба пуста" стала, потому что невозможно вернуть родных и любимых, погибших на войне. Но «люди, несмотря ни на какие невзгоды, сохраняют и несут по жизни распахнутую, неунывающую душу».*)

СВЯЗЬ	Смысловая связь между примерами.	(*СВЯЗЬ \А также *) писатель обращает внимание на то, что в Вологодчине много недостроенных и обветшалых изб — это означало, что некому закончить начатое...
ПОЗИЦИЯ	Позиция автора по проблеме.	(*ПОЗИЦИЯ\ Авторская позиция ясна: В.П. Астафьев считает, что несмотря на все невзгоды, которые пережил русский народ, он продолжает жить дальше, не унывать. И это было характерно для всего народа в целом, потому что общая \беда сплотила его. *)
ОТНОШЕНИЕ	Отношение к позиции автора по проблеме.	(*ОТНОШЕНИЕ\ Таким образом, можно сделать вывод, что главенствующей ролью литературы является духовное и нравственное воспитание человека, а также сохранение накопленного за века опыта предыдущих поколений. *)
Formulation of problems of the original text (criterion K1)		
П.проблема	Проблема сформулирована неверно.	
П.факт	Фактическая ошибка, связанная с пониманием и формулировкой проблемы.	
Comment on the formulated problem of the original text (criterion K2)		
П.опора	Проблема прокомментирована без опоры на исходный текст.	
П.пересказ	Вместо комментария дан простой пересказ текста.	
П.факткомм	Фактическая ошибка в комментарии, связанная с пониманием исходного текста.	
П.другая	Прокомментирована другая, не сформулированная экзаменуемым проблема.	
П.копир	Вместо комментария цитируется большой фрагмент исходного текста.	Одной из главнейших задач современной литературы — «предостеречь молодежь от опасности душевного очерствения, помочь ей усвоить и обогатить духовный багаж, накопленный предшествующими поколениями».
Reflection of the position of the author of the original text (criterion K3)		
П.позиция	искаж Позиция автора сформулирована неверно или искажена. факт Фактическая ошибка, связанная с пониманием позиции автора.	
Attitude towards the author's position on the problem of the source text (criterion K4)		
П.отнош	Высказанное отношение к позиции автора не соответствуют сформулированной проблеме.	

П.обоснов	необосн Отношение к позиции автора не обосновано. форм Отношение к позиции автора заявлено лишь формально.	Я согласен / не согласен с автором
Semantic integrity, speech coherence and sequence of presentation (criterion K5) - logical errors, see section 2.3.		
Accuracy and expressivity of speech (criterion K6)		
П.однообр	Однообразие грамматического строя речи.	Прочитав текст Белова, я могу понять автора. Я понимаю, что Белов очень рад приходу весны. Леса, поля, реки оживают. Автор любит весну, землю, солнце, березы, ветер. Автор наделяет их душой, сердцем. Автор призывает нас оглядеться вокруг, любоваться природой, радоваться жизни. Я согласна с Беловым. Весна - пора любви, счастья, веселья. Я очень люблю весну. Весна - это самое прекрасное время.
П.точность	Нарушение точности выражения мысли.	
Compliance with spelling and punctuation standards (criteria K7, K8) - are not checked within the framework of the Contest and are not included in the markup.		
Compliance with grammatical, speech, ethical standards (criteria K9, K10, K11) - see sections 2.1, 2.2, 2.5.		
Compliance with factual accuracy in background material (criterion K12) - see sections 2.4.		

Evaluation formula in points

The formula for evaluating an essay in the Russian language formalizes the rules given in [2]. Notation [x] means one if condition x is true and zero if condition x is false.

если (объём < 70 слов) то $K=K1=...=K12=0$ и далее оценивание не производится;
 если (**ПРОБЛЕМА**>0) и (**П.проблема**=0) и (**П.факт**=0) то $K1=1$;
 иначе $K1=K2=K3=K4=0$ и переход к вычислению $K5-K12$;
 если (**П.опора**+**П.пересказ**+**П.факткомм**+**П.другая**+**П.копир**>0) то $K2=0$;
 иначе если (**ПРИМЕР**>=2) и (**ПОЯСНЕНИЕ**>=2) и (**СВЯЗЬ**>=1) то $K2=5$;
 иначе если (**ПРИМЕР**>=2) и (**ПОЯСНЕНИЕ**+**СВЯЗЬ**>=2) то $K2=4$;
 иначе если (**ПРИМЕР**+**ПОЯСНЕНИЕ**+**СВЯЗЬ**>=3) то $K2=3$;
 иначе если (**ПРИМЕР**=2) то $K2=2$;
 иначе если (**ПРИМЕР**=1) то $K2=1$;
 иначе $K2=0$;
 если (**ПОЗИЦИЯ**>0) и (**П.позиция**=0) то $K3=1$ иначе $K3=0$;
 если (**ОТНОШЕНИЕ**>0) и (**П.отнош**+**П.обоснов**=0) то $K4=1$ иначе $K4=0$;
 $ошГрам$ = число грамматических ошибок согласно разделу 2.1;
 $ошРеч$ = число речевых ошибок согласно разделу 2.2;
 $ошЛог$ = число логических ошибок согласно разделу 2.3;
 $ошФакт$ = число фактических ошибок согласно разделу 2.4;

$ошЭтич$ = число этических ошибок согласно разделу 2.5;
 $ошОрф$ = число орфографических ошибок (равно нулю в рамках Конкурса);
 $ошПункт$ = число пунктуационных ошибок (равно нулю в рамках Конкурса);
 $КороткийТекст$ = [объём < 150 слов]
 $K5 = \max(2 - ошЛог, 0)$;
 $K7 = \max(0, \text{округление_вниз}(3 - 0.5 * ошОрф - КороткийТекст))$;
 $K8 = \max(0, \text{округление_вниз}(3.5 - 0.5 * ошПункт - КороткийТекст))$;
 $K9 = \max(0, \text{округление_вниз}(2 - 0.5 * ошГрамм - КороткийТекст))$;
 $K10 = \max(0, \text{округление_вниз}(2.5 - 0.5 * ошРеч - КороткийТекст))$;
 $K11 = [ошЭтич = 0]$;
 $K12 = [ошФакт = 0]$;
 если ($П.однообр + П.точность = 0$) и ($K10 \geq 2$) то $K6 = 2$;
 иначе если ($П.однообр + П.точность = 1$) или ($K10 < 2$) то $K6 = 1$;
 иначе $K6 = 0$;
 $K = K1 + \dots + K12$. Максимальное значение $K = 24$.

Condition for assigning a third check

Third check is carried out under the following conditions of significant discrepancy between the assessments of two independent experts; at the same time, the third expert is provided with the data of two previous checks [2]:

- the discrepancy between the final marks K of two experts is 8 or more points,
- either the difference in $K7$ scores is 2 or more points,
- or the difference in $K8$ scores is 2 or more.

3.7.Errors in essays on literature

In an essay on a literary topic (task No. 17.1-4 in the USE 2020), the student is required to reveal the topic deeply and from different perspectives, using the text for argumentation, based on theoretical and literary concepts. The length of the essay is at least 200 words [3].

The errors described in this section refer to essay types: **literature**.

Code	Standard comments	Examples
Types of semantic blocks		
АРГУМЕНТ	Аргументация с привлечением текста произведения.	Он строго следует моде – «как dandy лондонский одет», чтобы быть принятым в свете. Евгений по три часа проводит за своим туалетом, а потом старается везде успеть: на веселый ужин – «Вошел: и пробка в потолок», на бал, в театр.
ПОНЯТИЕ	Теоретико-литературное понятие используется для анализа текста произведения в целях раскрытия темы сочинения.	Архетип Прием зеркальной композиции Художественный образ Фабула
Compliance of the essay with the topic and its disclosure (criterion 1)		
С.одностор	Тема сочинения раскрыта глубоко, но односторонне.	
С.поверхн	Тема сочинения раскрыта поверхностно.	
С.тема	Тема сочинения не раскрыта	

Using the text of the work for argumentation (criterion 2)		
С.опора	Суждения не аргументируются текстом произведения(-ий).	Основная часть произведения показывает читателю другую сторону образа Архипа Савельича – он упрям, ворчлив, готов отстаивать своё мнение до последнего и любит читать наставления. Он очень бережлив, задумывается о благосостоянии своего «воспитанника», поэтому старается оградить его от необдуманных расходов. (*С.опора\ Главный герой – типичный представитель своего времени. Он наделён чертами человека своей эпохи и своего социального круга. Для него характерна (\С.послед прот\ холодность, мятежность, страстность :: Как можно, не комментируя, соединять в одном человеке холодность, мятежность и страстность? \) натуры и противопоставление себя обществу. *)
С.упрощен	общ Для аргументации текст привлекается на уровне общих рассуждений о его содержании. анализ Нет анализа важных для раскрытия темы сочинения фрагментов, образов, микротем, деталей и т.п.	Главный герой, Родион Раскольников, - молодой человек, имеющий довольно философский склад ума и анализирующий свои поступки и мысли. Казалось бы, он не способен пойти на преступление, отнять жизни у невинных людей ради достижений цели, но, увы, это произошло.
С.пересказ	Аргументация подменяется пересказом текста.	Чтобы подтвердить мысль современного литературоведа, обратимся к тексту произведения. Перед нами два героя, они оба красивы и молоды. Вся жизнь предстает перед ними в необычайных красках, они влюблены друг в друга. Героям ничего не мешает, они оба достаточно обеспечены, московская праздная жизнь начинает поглощать героев. Бунин неслучайно воссоздает насыщенную картину интеллектуальной и культурной жизни России 20 века. Для этого рассказа очень существенна привязанность событий к определенному времени.
С.позиция	Позиция автора сформулирована неверно или искажена.	
С.факт	Фактическая ошибка.	На протяжении всего романа Печорин не меняется, он всегда остаётся эгоистом, который не задумывается о чувствах других людей. Во время нападения (\С.факт\ крестьян\ на Белогорскую крепость Савельич совершает благороднейший поступок, достойный внимания и уважения читателей.
Reliance on theoretical and literary concepts (criterion 3)		
С.понятие	Ошибка при использовании теоретико-литературного понятия.	Таким образом, у Григория Александровича присутствует много черт романтического героя, однако всё же его следует относить к (\С.понятие\ архетипу\ «лишний» человек.\)

С.неиспол	Теоретико-литературное понятие введено, но не использовано для аргументации.	В пьесе “Гроза” Островского описываются душевные метания Катерины. Существует три рода литературы: эпос, лирика и драма. Пьеса относится к драматическому роду литературы. Жизнь Катерины не проста...
Compositional integrity and consistency (criterion 4)		
С.послед	прот Рассуждения непоследовательные или противоречивые. затруд Грубое нарушение последовательности частей высказывания, существенно затрудняющее понимание смысла сочинения. повт Необоснованный повтор.	Таким образом, у Григория Александровича присутствует много черт романтического героя, (*С.послед прот \ однако всё же его следует относить к архетипу «лишний» человек. *)
С.неразв	Мысль повторяется и не развивается.	
С.связь	Нарушение композиционной связи между смысловыми частями	
С.компози	В сочинении не прослеживается композиционный замысел	
Compliance with speech norms (criterion 5)		
- speech errors, see section 2.2.		

Note. Errors С.одностор, С.поверхн, С.тема, С.компози refer to the entire text of the work and do not require the selection of a fragment.

Evaluation formula in points

The formula for evaluating an essay on literature formalizes the rules given in [3].

если (объём < 150 слов) то $K=K1=...=K5=0$ и далее оценивание не производится;
 если (**С.тема**>0) то $K1=0$;
 иначе если (**С.поверхн**>0) то $K1=1$;
 иначе если (**С.одностор**>0) то $K1=2$;
 иначе $K1=3$;
 если (**АРГУМЕНТ**=0) или (**С.опора**>0) или (**С.позиция**>0) или (**С.факт**>=4) то $K2=0$;
 иначе если (**С.упрощен**>0) или (**С.пересказ**>0) или (**С.факт**>=3) то $K2=1$;
 иначе если (**С.факт**>=2) то $K2=2$;
 иначе если (**С.факт**=0) то $K2=3$;
 если (**ПОНЯТИЕ**=0) или (**С.понятие**>=2) то $K3=0$;
 иначе если (**С.неиспол**>0) или (**С.понятие**>=1) то $K3=1$;
 иначе $K3=2$;
 если (**С.компози**>0) то $K4=0$;
 иначе если (**С.неразв**>0) или (**С.связь**>0) то $K4=1$;
 иначе если (**С.послед**>0) то $K4=2$;
 иначе $K4=3$;
 $ошРеч$ = число речевых ошибок согласно разделу 2.2;
 $K5 = \max(0, \text{округление_вниз}(3.5 - 0.5 * ошРеч))$;

$K = K1 + \dots + K5$. Максимальное значение $K = 14$.

Condition for assigning a third check

Third check is carried out under the following conditions of significant discrepancy between the assessments of two independent experts; at the same time, the third expert is provided with the data of two previous checks [3]:

- the discrepancy between the final marks of K of two experts is 7 or more points,
- either the discrepancy in any of the criteria $K1..K5$ is 2 or more points,
- either a discrepancy according to the estimate $K1 = 0$.

3.8.Errors in social studies essays

In a *mini-essay on social studies* (task number 29 in the USE 2020), the student is required to reveal the topic of a given statement based on social science knowledge. Disclosure of the meaning of the statement, the correctness of explanations of key concepts and theoretical provisions, the quality of the examples given are evaluated [4].

The errors described in this section refer to essay types: **social studies**.

Code	Standard comments	Examples
Types of semantic blocks		
ИДЕЯ	Основная идея высказывания или тезис, требующий обоснования.	Данным высказыванием Карлейль хотел сказать, что безработица — негативное явление, из-за которого люди не могут обеспечивать себя необходимыми вещами и удовлетворять свои потребности.
ПОНЯТИЕ	Объяснение ключевого понятия.	Безработица — социально-экономическое явление, сущность которого заключается в том, что часть экономически активного населения, желающего работать, не может найти работу.
ТЕОРИЯ	Формулировка теоретического положения.	У безработицы, как ни странно, есть положительные черты: она формирует мобильный резерв рабочей силы, снижает уровень инфляции, а также повышает мотивацию работников, ведь никто не хочет потерять работу. Но всё-таки безработица — это негативное явление, поэтому и отрицательных черт у нее больше: снижается уровень жизни населения; у людей теряются профессиональные знания и навыки, что существенно затрудняет возможность трудоустройства.
ЛОГИКА	Последовательность рассуждений или причинно-следственных связей, включая выводы.	Таким образом, хоть и при наличии положительных черт, безработица все равно носит негативный характер и делает несчастными не только людей, но и государства.
ПРИМЕР.ОБЩ	Факт или пример из общественной жизни, в том числе по материалам СМИ.	Из новостного портала "infox" мне стало известно о такой печальной статистике: 45 тысяч самоубийств происходит из-за потери работы. Люди просто не могут пережить такое потрясение, и поэтому идут на такой отчаянный шаг.

ПРИМЕР.ЛИЧ	Факт или пример из личного социального опыта.	Так, изучая экономическую теорию, я провела своего рода «эксперимент» - посчитала стоимость проезда на метро в случае если бы на рынке были две компании и стоимость билета оказалась выше, нежели при присутствии на рынке только одной фирмы.
ПРИМЕР.ИСТ	Факт или пример из истории литературы, искусства, науки, техники.	Так, в 30х годах двадцатого века известный экономист Дж. Кейнс предложил государствам стимулировать потребительский спрос, который подталкивает производителей к большому инвестициям в производство. А если инвестиции растут, растёт и выпуск товара и тем самым растёт ВВП, а это значит экономика развивается. Последовав рекомендациям Кейнса, правительство Рузвельта уменьшило уровень безработицы в стране и достигла высокого уровня ВВП.
Errors in revealing the meaning of the statement (criterion K1)		
о.смысл	неверн Неверно выделена основная идея высказывания. неотр Сформулированный тезис не отражает смысла высказывания.	Тема: «Вникните в причины всякой распущенности и вы увидите, что она проистекает из безнаказанности» (Ш. Монтескьё) (*о.смысл неотр\ В выбранном мною высказывание автор говорит о том, что именно безнаказанность способствует распущенности обществу и всему тому, что творится внутри него. Ошибочно считать, что один безнаказанный случай не приведёт к тому, что такие преступления не будут повторяться вновь. ∴ сформулирован тезис: безнаказанность - причина преступности, авторский тезис иной/шире*)
о.подмена	общ Произведена подмена смысла высказывания рассуждениями общего характера, не отражающими специфики предложенного высказывания дз домашняя заготовка	
о.пересказ	Раскрытие смысла высказывания подменяется его прямым пересказом или перефразированием.	Тема: «Как только человек имеет всю свободу на свете, он становится животным» (А. С. Кончаловский) (*о.пересказ \Кончаловский в своём высказывании говорит о том, что человек не может обладать абсолютной свободой, так как начинает превращаться в животного, а именно поведение человека начинает быть схожим с инстинктами животными.*)
Errors in concepts and theoretical provisions (criterion K2)		
о.понятие	Есть неточность в объяснении ключевого понятия, не искажающая его научного смысла.	Общественный прогресс — это процесс развития общества от низшего к высшему, от примитивного, дикого состояния к высшему, цивилизованному.

о.теория	Есть неточность в объяснении теоретического положения, не искажающая его научного смысла.	(* о.теория \Каждый предприниматель сталкивается с таким явлением, как конкуренция – это соперничество между участниками рыночных отношений за лучшие условия купли и продажи товаров и услуг. Чаще всего в рыночной экономике представляется свободная конкуренция, основным признаком которой является свобода входа/выхода из бизнеса. ∴ "каждый", "основной признак" *)
о.теорсвязь	идея Понятие или теоретическое положение не связано с основной идеей или тезисом. пон Понятие не связано с другими понятиями. нерас Теоретическое положение не раскрывает смысла высказывания.	Но всё-таки безработица — это негативное явление, поэтому и отрицательных черт у нее больше: снижается уровень жизни населения; у людей теряются профессиональные знания и навыки, что существенно затрудняет возможность трудоустройства. Выделяют следующие виды безработицы: фрикционная, структурная, циклическая и сезонная. Таким образом, хоть и при наличии положительных черт, безработица все равно носит негативный характер и делает несчастными не только людей, но и государства.
о.нехватает	Не хватает ключевых понятий, необходимых для раскрытия смысла высказывания.	(*о.нехватает \Конкурируя, производители стремятся угодить потребителю, вводя нововведения в свой продукт. Большое количество потребителей является показателем успешности бизнеса, так как чем больше клиентов, тем большие прибыли получает предприниматель, что ведет к процветанию бизнеса. ∴ отсутствуют понятия: издержки, спрос *)
о.упрощ	Упрощённые рассуждения на уровне обыденных представлений без опоры на обществоведческие знания.	Свобода — это возможность поступать так, как хочется.
Errors in the logic of reasoning (criterion K3)		
о.рассужд	Рассуждения непоследовательные или противоречивые.	
о.вывод	Сформулированный вывод не обоснован или не достоверен с точки зрения научного обществознания	
Errors in facts or examples (criterion K4)		
о.примсвязь	Приведённый факт или пример не связан с иллюстрируемой идеей или тезисом.	
о.подтв	неподтв Приведённый факт или пример не подтверждает иллюстрируемую идею или тезис. общ Вместо примера приведены общие рассуждения.	Не так давно мной было прочитано произведение Н. Главный герой произведения Т всю свою жизнь только и мечтал о будущем, о путешествиях, о семье, но так ничего и не добился. Но почему? Герой всему находил отговорки, и вместо полноценной жизни, просто существовал, откладывая все дела на потом, либо не делал ничего вообще.

о.дубл	Пример дублирует предыдущий по содержанию.	Мария Петровна решила открыть салон красоты в маленьком городе, где уже имелось большое количество похожих предприятий. Не обращая внимания на данный факт, она открыла салон. У неё были клиенты, но их было мало. Из-за недостатка прибыли гражданка М была вынуждена закрыть салон. Моя бабушка имеет небольшой магазин обуви в маленьком городе. Она, как предприниматель, всегда должна оценивать ситуации на рынке, а именно, что нужно потребителям. Так как если она произведет необходимый товар, то она получит большую прибыль.
о.факт	знач Факт приведён некорректно. незнач Незначимое искажение факта, не влияющее на корректность выводов	(*ПРИМЕР.ИСТ о.факт незнач \ Примером из истории может являться предприниматель Генри Форд. Он первый в истории поставил изготовление автомобилей через конвейер, что позволяло создавать множество экземпляров значительно быстрее, чем при ручной сборке. Данный ход был высоко оценен потребителями, машины стали покупаться, что приносило Г. Форду огромные прибыли, что позволило открывать заводы по всему миру. :: главным фактором для покупателей была цена, а не скорость производства *)
о.подробн	Факт сформулирован недостаточно развёрнуто.	(*ПРИМЕР.ОБЩ о.подробн \ Недавно по телевизору шла телепередача, в которой говорилось об аварии на АЭС Фукусима-1. Произошла крупная радиационная авария, в результате сильнейшего землетрясения, а после и цунами. Данная авария привела к заражению различных видов животных, людей также настигли различные заболевания, вследствие которых многие из них заканчивались летальным исходом. Данный пример иллюстрирует техногенную катастрофу, которая настигла не только жителей, но и другие страны. Это стало общей, глобальной проблемой. :: пример не полностью раскрывает понятие "противоречивость общественного прогресса", не говорится о роли АЭС в жизни общества *)

Evaluation formula in points

The formula for evaluating a mini-essay on social studies formalizes the rules given in [4]. Notation [x] means one if condition x is true and zero if condition x is false.

если (ИДЕЯ>0) и (о.смысл=0) и (о.подмена=0) и (о.пересказ=0) то K1=1;
иначе K=K1=K2=K3=K4=0 и далее оценивание не производится;
если (о.упрощ>0) то K2=0; иначе
K2 = max(0, [ПОНЯТИЕ>0] + [ТЕОРИЯ>0] – [о.понятие+о.теория>0] – о.теорсвязь – о.нехватает);
K3 = max(0, [ЛОГИКА>0] – о.рассужд – о.вывод);

$K4 = \max(0, \min(2, [\text{ПРИМЕР.ОБЩ}] > 1] + [\text{ПРИМЕР.ЛИЧ}] > 1] + [\text{ПРИМЕР.ИСТ}] > 1]) - \text{о.примсвязь} - \text{о.подтв} - \text{о.дубл} - \text{о.факт} - \text{о.подробн});$
 $K = K1 + \dots + K4.$ Максимальное значение $K = 6.$

Condition for assigning a third check

The third check is carried out under the following conditions of significant discrepancy between the assessments of two independent experts; at the same time, the third expert is provided with the data of the two previous checks [4]:

- the discrepancy between the final marks K of two experts is 3 or more points,
- either the discrepancy on the $K2$ or $K4$ criterion is 2 points,
- either a discrepancy according to the estimate $K1 = 0.$

An example of marking up a mini-essay on social studies

Тема: (* Подлинные достижения человека откладываются не только вне его, в тех или иных порожденных им объектах, но и в нем самом.
 (С.Л.Рубинштейн) *)

Линия: социология

Класс: 11

Год: 2020

Предмет: обществознание

Тест: егэ тренировка

Эксперт: СеверусСнейп

$K1: 1$

$K2: 2$

$K3: 1$

$K4: 1$

(*Идея\Смысл данного афоризма заключается в том, что результаты деятельности индивида видны не только в тех объектах, на которые эта деятельность была направлена, но и в самом человеке. Автор утверждает, что главными результатами и достижениями человека являются внутренние изменения, которые с ним происходят в процессе социализации. Рубинштейн освещает актуальную тему самовоспитания, становления личности и её социализации. Я полностью согласна с мнением автора. Настоящие достижения закладываются в процессе деятельности внутри самого человека.*)

Обратимся к теории для подтверждения. (*Логика\От рождения человек является (\Понятие\индивидом – одиночным представителем человеческого вида homo sapiens\). В процессе (\Понятие\социализации - усвоения норм и ценностей, принятых в обществе\)) индивид становится (\Понятие\личностью, которую можно определять как человека, как обладателя социальных черт общества, необходимых ему для жизнедеятельности в обществе\). Процесс социализации длится всю жизнь. Каждое новое достижение помогает человеку укрепить свой статус личности, даёт ему толчок для дальнейших достижений и открытий. Кроме того, у человека постепенно формируется «Я-концепция», помогающая ему найти своё предназначение и занять своё место в обществе. (\Понятие\Деятельность – это процесс целенаправленной активности людей.\)) К видам деятельности относят труд, игру, общение, и др. Любая деятельность сопровождается (\Понятие\мотивом – тем, почему человек занимается деятельностью\), и (\Понятие\целью – для чего он ей занимается\). Достигнув цели, человек приобретает уверенность в себе и желание трудиться ещё больше для поддержания положительных результатов. Победы становятся стимулом к движению вперед.*) Для поддержания этого мнения приведу несколько аргументов.

(\Пример.общ\ Как известно из многих СМИ, Стив Джобс был основоположником компании (*о.факт незнач\ по производству телефонов и планшетов «Apple» >> «Apple» по производству персональных компьютеров *). Его идея телефонов нового поколения стала настоящим открытием и принесла создателю много прибыли и успеха. (*О.подробн\ Однако такой успех не расслабил Джобса, а наоборот, стал стимулом :: здесь не раскрыто, как именно достижения Джобса повлияли на него самого *), воодушевившим его на создание новых электронных устройств, пользующихся невероятной популярностью и по сей день.\)

(*Пример.ист\ По сюжету серии книг о Гарри Поттере, о которой нам рассказывал учитель литературы в школе, Гермиона Грейнджер с первого курса школы чародейства и волшебства являлась лучшей ученицей среди сверстников. Её достижения в большинстве учебных предметов сделали из Гермионы уверенную в себе девушку, которая поставила себе цель быть отличницей по всем предметам в течение всего семилетнего обучения. Результаты экзаменов всегда были высоки, что помогло Грейнджер сформироваться, как личности. Результаты в учёбе влияли на поведение Гермионы, она становилась веселой и уверенной в своих силах, что видели и одобряли как учителя, так и ученики. *)

Таким образом, можно сказать, что Рубинштейн в своём высказывании абсолютно прав.

(*Логика\ Достижения человека в какой-либо сфере деятельности сказываются на его внутренней уверенности в себе, вере в свои силы и желании двигаться в том же направлении для достижения все больших успехов.*)

3.9.Errors in essays on history

In a historical essay about a given period of history (task number 25 in the USE 2020), the student is required to describe significant events (at least two), the reasons for these events (at least two), historical figures (at least two), their actions and roles. Knowledge of historical facts and / or opinions of historians, the ability to assess the impact of events on further history, the correctness of the use of historical terms and concepts, and the coherence of presentation are assessed [5].

The errors described in this section refer to essay types: **history**.

Code	Standard comments	Examples
Types of semantic blocks		
СОБЫТИЕ СЯП	Историческое событие, явление или процесс.	Однако наиболее значимым событием данного периода является (*СЯП\ развенчание культа Сталина на XX съезде КПСС в 1956 году.*)
РОЛЬ	Роль исторической личности с указанием её конкретных действий, в значительной степени повлиявших на ход и (или) результат события.	Так, в 1961 году Юрий Гагарин первый совершил полёт в космос.
ПРИЧИНА СЛЕДСТВИЕ	Причина и следствие события, явления или процесса (выделяются два фрагмента, и им назначается один и тот же тег).	(*ПРИЧИНА\ Этот доклад Хрущёва #1*) по праву изменил мировоззрение всех советских граждан. Кроме того, (*СЛЕДСТВИЕ\ прямым следствием этого съезда стала реабилитация политзаключённых, арестованных в годы управления страной Сталиным. На свободу были выпущены сотни тысяч репрессированных, а расследования политических дел, начатые ещё при Сталине, приостановлены (например, «Дело врачей»). #1*)

ОЦЕНКА	Оценка влияния события, явления или процесса на дальнейшую историю с опорой на исторические факты и/или мнения историков.	Из-за победы группировки Сталина в во внутрипартийной борьбе большинство в партии выступило за форсированную модернизацию. В дальнейшем это приведёт к осуществлению политики «большого скачка» — индустриализации промышленности и коллективизации сельского хозяйства.
Errors when specifying a historical event (criterion K1) An error related to an event, phenomenon or process is associated with the corresponding СЯП block using a tag. If the tag is not specified, then the error refers to the nearest preceding СЯП block.		
И.событие И.сяп	Событие, явление или процесс указано неверно.	Хронологические рамки периода с марта 1953 года по октябрь 1964 года определяются смертью И.В. Сталина и назначением Н. С. Хрущёва первым секретарем ЦК КПСС.
И.период	Событие не относится к данному периоду истории страны.	
Errors in specifying a historical person and his/her role (criterion K2) An error related to a historical person is associated with the corresponding РОЛЬ block using a tag. If no tag is specified, then the error refers to the closest preceding РОЛЬ block.		
И.личность	Историческая личность названа неверно.	Владимир Владимирович Ленин.
И.лсвязь	Деятельность исторической личности не связана с указанными событиями.	
И.лпериод	Деятельность исторической личности не относится к данному периоду истории страны.	
И.лроль	Роль личности охарактеризована неверно.	Ньютон внес большой вклад в развитие химии 20 века.
И.лдейств	неверн Действие исторической личности указано неверно. неконк Указанное действие исторической личности не является конкретным и одномоментным.	Также значимой личностью, принявшей участие в Чехословацком кризисе, был А. Дубчек. Он, будучи недовольным большим влиянием СССР на Чехословакию, поднял антисоветское восстание. Дубчек сверг старое правительство и создал новое с собой во главе. В этом периоде Алексей Косыгин занимал должность председателя Совмина и сыграл важную роль в реализации Косыгинской экономической реформы.
Errors in causal relationships (criterion K3)		
И.причин	неверн Причина события, явления или процесса указана неверно. неук Причины события, явления или процесса не указаны.	(*ПРИЧИНА \В результате невероятного подъема силы духа русского солдата и чувства национального самосознания #1*) Россия изгнала Наполеона со своей территории и (*СЛЕДСТВИЕ\ тем самым сделала первый шаг к освобождению стран Европы от наполеоновского владычества.#1*) (*ПРИЧИНА \ (И.причин неверн \Также во внешней политике Россия преследовала цель захвата Финляндии и устранения угрозы северным рубежам страны.) #1*)

И.следств	<p>неверн Следствие события, явления или процесса указано неверно.</p> <p>неук Следствия события, явления или процесса не указаны.</p> <p>период Следствие события, явления или процесса указано, но выходит за рамки рассматриваемого периода.</p>	По инициативе Александра были проведены преобразования в области образования (*И.следств \ следствием которых стало систематизирование школ, открытие новых учебных заведений. ∴ перечисленные события не являются следствием преобразований, так как произошли в ходе преобразований, а не после них*)
Errors in assessing the impact of an event on further history (criterion K4)		
И.влиян	Некорректная оценка влияния события, явления или процесса на дальнейшую историю.	Историческая оценка данного периода многогранна. После смерти Брежнева новым Генеральным секретарём станет Ю. Андропов, который умрёт в 1984-ом. За время своего правления он проведёт частичную чистку партийного и государственного аппаратов, сместив с должностей 18 министров СССР и переизбрав 37 первых секретарей обкомов. В 1985-ом к власти придёт М. Горбачёв, вследствие чего начнётся "перестройка".
И.упрощ	Упрощённые рассуждения на уровне обыденных представлений без привлечения исторических фактов и (или) мнений историков.	Таким образом этот период имеет большое значение для развития России. Смута невероятно ослабила Россию, но вместе с тем, она же и показала всю истинную силу русского люда
Errors in the use of historical terminology (criterion K5)		
И.понятие	Некорректное использование исторического термина или понятия.	
И.неиспол	Не использованы необходимые исторические термины или понятия	
Actual errors (criterion K6)		
И.факт	<p>предст Неверно представлен исторический факт.</p> <p>ош Фактическая ошибка.</p>	И вскоре в 1606 году вспыхнуло восстание в Москве под предводительством Василия Шуйского. Косыгин приказал перейти к отраслевому управлению, отменил Совнархозы и вернул на их место министерства и ведомства. ∴ Это решение было принято членами Пленума ЦК КПСС в сентябре 1965 г.
Errors in the presentation (criterion K7)		
И.излож	<p>отрыв Ответ представлен в виде отдельных отрывочных положений.</p> <p>несвяз Несвязное изложение.</p>	

Evaluation formula in points

The Historical Essay Evaluation Formula formalizes the rules given in [5]. Notation $[x]$ means one if condition x is true and zero if condition x is false.

$$K1 = \max(0, \min(2, \text{СЯП} - \text{И.сяп} - \text{И.период}));$$

если ($K1=0$) то $K=K1=\dots=K7=0$ и далее оценивание не производится;
 выделяются две совокупности фрагментов, описывающих две исторические личности:
 $Личн_1 = [РОЛЬ_1 > 0] - [И.личность_1 + И.лсвязь_1 + И.лпериод_1 + И.лроль_1 + И.лдейств_1 > 0]$;
 $Личн_2 = [РОЛЬ_2 > 0] - [И.личность_2 + И.лсвязь_2 + И.лпериод_2 + И.лроль_2 + И.лдейств_2 > 0]$;
 $K2 = Личн_1 + Личн_2$;
 выделяются две пары фрагментов, описывающих две причинно-следственные связи:
 $ПСС_1 = [ПРИЧИНА_1 > 0 \text{ и } СЛЕДСТВИЕ_1 > 0] - [И.причин_1 + И.следств_1 > 0]$;
 $ПСС_2 = [ПРИЧИНА_2 > 0 \text{ и } СЛЕДСТВИЕ_2 > 0] - [И.причин_2 + И.следств_2 > 0]$;
 $K3 = ПСС_1 + ПСС_2$;
 $K4 = [ОЦЕНКА > 0] - [И.влиян + И.упрощ > 0]$;
 $K5 = 1 - [И.понятие + И.неиспол > 0]$;
 $K6 = \max(0, 3 - И.факт)$;
 если ($K1+K2+K3+K4 < 5$) то $K7 = 0$;
 иначе $K7 = [И.излож=0 \text{ и } K1+K2+K3+K4 \geq 5]$;
 $K = K1 + \dots + K7$. Максимальное значение $K = 12$.

Condition for appointing a third expert

The third check is carried out under the following conditions of significant discrepancy between the assessments of two independent experts; at the same time, the third expert is provided with the data of the two previous checks [5]:

- the discrepancy between the final marks K of two experts is 5 or more points,
- either the discrepancy on the criterion $K1$ or $K2$ or $K3$ or $K6$ is 2 or more points,
- or discrepancy in any four or more of the seven criteria.

An example of a markup of a historical essay

Тема: март 1881 г. – октябрь 1894 г.

Класс: 11

Год: 2019

Предмет: история

Тест: егэ тренировка

Эксперт: Белик Александра Алексеевна

$K1: 2$

$K2: 0$

$K3: 1$

$K4: 0$

$K5: 1$

$K6: 0$

$K7: 0$

Рамками данного исторического периода России являются (\СЯП\ вступление на престол и конец правления Александра III \).

В данном периоде можно выделить Победоносцева, занимавшегося воспитанием Александра III в юности. Также можно выделить (\РОЛЬ и.лдейств неконк\ самого Александра III, проводившего консервативную политику внутри государства \).

(\ПРИЧИНА\ Александр III считал преобразования своего отца ошибочными #1 \), видел в них причину убийства Александра II. (*СЛЕДСТВИЕ\ Из-за этого и из-за воспитания Победоносцева Александром III был взят (\и.понятие\ рекреативный курс\). #1 *) (\СЯП\ Цензура стала более жёсткой, была отменена автономия университетов и был увеличен контроль полиции в стране\). (*ОЦЕНКА\ Всё это нашло

отражение и в политике (\и.факт\ Николая II, выпустившего манифест о незыблемости самодержавия\)) и консервативных реформ, приведших к свержению самодержавия в России.*)

(\и.факт\ Александр III умер в результате покушения.\)) Была заложена бомба под поезд, в котором он ехал с семьёй. Это произошло по причине недовольства части населения, требующей либеральных реформ.

Данный период можно назвать одним из самых острых в истории России, так как именно в это время появилось основное недовольство консервативностью власти императоров и (\и.факт\ возникли группы по борьбе с самодержавием, и двум революциям в России\)), в результате которых самодержавие всё-таки пало.

3.10. Errors in a free essay in Russian *

* - As part of the 2020 trials, this type of essay will not be included in the training and test samples.

In a free essay in Russian (essay type Russian-free) grammatical, speech, factual, logical and ethical errors are assessed. Meaning blocks are not highlighted.

- Compliance with grammatical norms (criterion K1) - grammatical errors, section 2.1.
- Compliance with speech norms (criterion K2) - speech errors, section 2.2.
- Consistency and semantic integrity of presentation (criterion K3) - logical errors, section 2.3.
- Compliance with factual accuracy (criterion K4) - factual errors, sections 2.4.
- Compliance with ethical standards (criterion K5) - in accordance with section 2.5.
- Compliance with spelling and punctuation standards (criteria K6, K7) - are not checked within the framework of the Competition and are not included in the markup.

Evaluation formula in points

The grading formula for a free essay in Russian is based on some rules for checking an essay in Russian [2].

$ошГрам$ = число грамматических ошибок согласно разделу 2.1;
 $ошРеч$ = число речевых ошибок согласно разделу 2.2;
 $ошЛог$ = число логических ошибок согласно разделу 2.3;
 $ошФакт$ = число фактических ошибок согласно разделу 2.4;
 $ошЭтич$ = число этических ошибок согласно разделу 2.5;
 $ошОрф$ = число орфографических ошибок (равно нулю в рамках Конкурса);
 $ошПункт$ = число пунктуационных ошибок (равно нулю в рамках Конкурса);
 $K1 = \max(4 - ошГрам, 0);$
 $K2 = \max(4 - ошРеч, 0);$
 $K3 = \max(4 - ошЛог, 0);$
 $K4 = \max(3 - ошФакт, 0);$
 $K5 = \max(3 - ошЭтич, 0);$
 $K6 = \max(0, \text{округление_вниз}(3 - 0.5 * ошОрф));$
 $K7 = \max(0, \text{округление_вниз}(3.5 - 0.5 * ошПункт));$
 $K = K1 + \dots + K7.$ Максимальное значение $K = 24.$

Condition for assigning a third check

The third check is carried out under the following conditions of significant discrepancy between the assessments of two independent experts; at the same time, the third expert is provided with the data of the two previous checks [2]:

- the discrepancy between the final marks K of the two experts is 8 or more points.

3.11. Errors in essays in English language

In an essay in English language (task No. 40 in the USE 2020), the student is required to express his/her opinion according to the following plan:

- (1) paraphrased problem statement,
- (2) personal opinion with 2–3 arguments;
- (3) opposite opinion with 1–2 arguments;
- (4) rationale of disagreement with the opposite opinion;
- (5) conclusion that confirms the author's position.

These five points are called *aspects*, and the sixth aspect is styling in a neutral style. The solution to the communicative problem, text organization, vocabulary, grammar, spelling, and punctuation are all evaluated [6,7].

The errors described in this section apply only to the **English** essay type.

Code	Standard comments	Examples Topic: (*Sport unites people*)
Types of semantic blocks		
PROBLEM	Introduction, problem statement.	(*PROBLEM\ Doing and watching sport brings many strong feelings in people's lives. *)
POPINION	Personal opinion.	(*POPINION\ I firmly believe that sport unites people. *)
OPOPINION	Opposite opinion.	(*OPOPINION\ However, some people think that sport does not unite people because sports supporters of different teams behave aggressively and fight with each other. *)
ARGUMENT	Argument.	(*ARGUMENT\ Firstly, athletes become close friends when they train and compete together in the same team. *) (*ARGUMENT A.logic arg\ Secondly, athletes should train much to achieve high results and to win a competition. :: this argument doesn't answer the question "why do you think so?" *)
RATIONALE	Rationale of disagreement with the opposite opinion.	(*RATIONALE\ I do not agree with that because extreme sports fanatics are small groups of people who are condemned by society*)
CONCLUSION	Conclusion.	(*OUTPUT\ I think that sport unites people. The first reason is that one-team athletes always build close friendship when they train and compete. The second reason is that strangers become strongly linked emotionally when they support the same national team. *)

Errors in solving a communicative task (criterion C1, from 0 to 3)		
A.style	Stylistic error. rhet Rhetorical question. coll Colloquial expression. reduc Reduced vocabulary. contr Contractions.	(\A.style conv\ Let's >> Let us \) (\A.style reduc\ folks >> people \) (\A.style contr\ I'm >> I am \) (\A.style contr\ aren't >> are not \)
A.periph	The statement in the introduction isn't paraphrased, but is quoted word for word; it's necessary to replace at least two words or a grammatical structure ("more or less" according to the additional evaluation scheme).	
A.probl	The introduction doesn't reflect the problematic nature of the topic ("more or less" according to the additional evaluation scheme).	
A.aspect	undiscl An aspect was not covered.	(\A.aspect undiscl\ Many people like doing sport in my country. I am very sporty too :: the author writes about their love for sports, while the essay topic is devoted to the relationships between people who are fond of sports, and their feelings towards one another \)
A.aspm	An error regarding coverage of an aspect, giving "more or less" according to the additional evaluation scheme. incompl An aspect wasn't fully covered. inacc An aspect was not covered accurately.	
A.fact	Actual error ("more or less" according to the additional evaluation scheme).	
A.volume	The answer doesn't correspond to the required length (200–250 words).	
A.unprod	This part of the answer is unproductive (textually identical to the published source).	
Errors in organization of the text (criterion C2, from 0 to 3)		
A.logic	Logical error. arg The argument doesn't support the opinion.	
A.comm	An error in using the means of logical connection.	
A.para	The text is not divided into paragraphs.	

Lexical errors (criterion K3, 0 to 3)		
A.lex	Lexical error. cont Incorrect use of the word in context compat Violation of lexical compatibility skip Skipping a word when it doesn't affect the grammatical structure of the sentence wform Error in word formation that doesn't change the part of speech postpos Postposition in a composite verb word Spelling error in a word that changes its meaning	(\A.lex wform\ unregular >> irregular) (\A.lex word\ think >> thing)
A.stock	The word-stock used doesn't correspond to the high level of task complexity. repeat Repeating the same words syn Lack of synonyms ant Lack of antonyms incorr Incorrect use of the word compat Violation of compatibility	(\A.wstock repeat\ Scientists think think think that \) (\A.wstock syn\ Many people think that sport unites people, but some people do not agree. ∴ people – three times \)
Grammatical errors (criterion K4, 0 to 3)		
A.gramm	Grammatical error. typetime Tense and aspectual form of the verb imp Impersonal verb form mod Modal verb plur Plural form poss Possessive noun comp Comparative form of an adjective or adverb art Article prep Preposition pron Pronoun ord Order of words in a sentence skip Skipping a word (subject or predicate) in such a way that affects the sentence's grammatical structure wform Word formation error if a part of speech changes	...you will be able to make a good choice of a career only when you (\A.gramm typetime\ will be older >> are older) ...it will be helpful for you to find (\A.gramm art\ right job >> the right job) (\A.gramm wform\ actively >> activity)
A.lev	The grammatical forms used don't correspond to the high level of the task's complexity. repeat Repeating the same construction simp Using only simple, short sentences comm Absence of commonly used sentences incorr Incorrect use of grammatical forms	
Punctuation and orthographical errors (criterion C5, from 0 to 2)		
A.spell	Spelling error.	
A.punct	Punctuation error.	

Note 1. In expert markup, the error A.volume may be absent, since the volume is calculated automatically.

Note 2. *Repeated errors*, i.e. mistakes related to the same rule of grammar, vocabulary, spelling, or punctuation are counted as one error. Tags are used to highlight repeated errors. All fragments containing the same repeated error should have the same tag. Different repeated errors should have different tags. It is preferable to take the abbreviation of the error type with the error identification number as a tag (for example, **#plur1**).

Scoring formula (points)

The scoring rules from [6,7] don't allow for an unambiguous calculation of criteria based on simply counting the number of semantic blocks and the number of errors, since they contain interval threshold values and ambiguous "AND/OR" conditions. In the proposed formulae, ambiguities are eliminated based on the *non-appealability principle* of the evaluation (which leads, on average, to a shift in evaluation in the student's favor). The number of aspects that were not covered is calculated separately (variable *rlAspects*), as well as the number of deviations from the proposed plan (variable *rlPlan*). Record $[x]$ means one, if condition x is true, and zero, if condition x is false.

$$rlAspects = [\text{PROBLEM}=0 \text{ or } \text{A.periph}>0 \text{ or } \text{A.probl}>0] \\ + [\text{POPINION}=0] + [\text{OPOPINION}=0] + \max(3 - \text{ARGUMENT}, 0) + [\text{RATIONALE}=0];$$

or ($\text{A.volume}<180$ words) or ($\text{A.unprod}>30\%$ words), then $C1 = 0$;
 or if ($rlAspects=0$) и ($\text{A.aspect}=0$) or ($\text{A.style}\leq 1$), then $C1 = 3$;
 or if ($rlAspects=0$) и ($\text{A.aspect}=2$) or ($\text{A.style}\leq 3$), then $C1 = 2$;
 or if ($rlAspects*2+\text{A.aspect}\leq 4$) or ($\text{A.style}\leq 4$), then $C1 = 1$;
 or $C1 = 0$;
 if ($C1 = 0$), then $C = 0$ and other criteria are not calculated;
 $rlPlan$ = the number of blocks missing or out of order in the sequence {**PROBLEM**, **POPINION**, **OPOPINION**, **RATIONALE**, **CONCLUSION**};
 if ($rlPlan=0$) or ($\text{A.logic}=0$) or ($\text{A.comm}=0$) and ($\text{A.para}=0$), then $C2 = 3$;
 or if ($rlPlan + \text{A.logic} + \text{A.comm} + \text{A.para}\leq 4$), then $C2 = 2$;
 or if ($rlPlan + \text{A.logic} + \text{A.comm} + \text{A.para}\leq 8$), then $C2 = 1$;
 or $C2 = 0$;
 if ($\text{A.lex}\leq 1$) and ($\text{A.stock}=0$), then $C3 = 3$;
 or if ($\text{A.lex}+3*\text{A.stock}\leq 3$), then $C3 = 2$;
 or if ($\text{A.lex}\leq 4$) and ($\text{A.stock}\leq 1$), then $C3 = 1$;
 or $C3 = 0$;
 if ($\text{A.gramm}\leq 2$) and ($\text{A.lev}=0$), then $C4 = 3$;
 or if ($\text{A.gramm}\leq 4$) and ($\text{A.lev}=0$), then $C4 = 2$;
 or if ($\text{A.gramm}\leq 7$) and ($\text{A.lev}\leq 1$), then $C4 = 1$;
 or $C4 = 0$;
 if ($\text{A.spell}\leq 1$) and ($\text{A.punct}\leq 1$), then $C5 = 2$;
 or if ($\text{A.spell}+\text{A.punct}\leq 4$), then $C5 = 1$;
 or $C5 = 0$;
 final grade $C = C1 + C2 + C3 + C4 + C5$ (maximum value $C = 14$).

Condition for appointment of a third expert

The discrepancy between the final grades of C of two experts is 4 or more points.

Example of a markup of an essay in English language

Topic: (*Sport unites people*)

Class: 11

Year: 2017

Subject: English language

Test: USE training

Expert: Severus Snape

(*PROBLEM\\ Some people (\A.punct\ think, >>think \) that sport unites people, while the others (\A.punct\ think, >>think \) that it is definitely not right. So, it is a great topic to (\A.spell\ disquiss >>discuss \). *)

(*POPINION\\ As for me, I believe that sport (\A.spell\ can not >>cannot \) work in the bad side, only in the good side of our life. (*ARGUMENT\\ Because it is a great way to spent time together, side by side. *) Also, I think, that (*ARGUMENT\\ with the help of sport you can meet new friends. *) Besides, I am convinced that if you and your friend have the same sport activity, you become (\A.gramm\ more closer \). *)

(*OPOPINION\\ Nevertheless, some people believe that (*ARGUMENT\\ sport (\A.gramm plur\ make >>makes \) people very nervous and (\A.spell\ exousted >>exhausted \). *). More than that they think that (*ARGUMENT\\ sport is just a waste of time*). *)

(*RATIONALE\\ I do not fully agree with that, because sport (\A.gramm\ help >>helps \) people to relax physically and mentally. Also, sport is the best thing for time spending for (\A.gramm\ your >>you \) and your family, because there are a lot of different kinds of sport which (\A.gramm\ specially were made >> were specially made \) for family *)

(*CONCLUSION\\ In conclusion, I would like to say that sport is great thing, that allows us to spend our time not only just for fun and joy, but also for our good health and our happy future. *)

Literature

1. I. A. Sherstobitova, S. P. Belokurova, L. G. Gvozdinskaya The classification of errors that are corrected and taken into account when evaluating a student's work. FIPM, 2012.
2. Demonstration version of control grading materials for the 2020 Unified State Exam in Russian Language. FIPM, 2019.
3. Demonstration version of control grading materials for the 2020 Unified State Exam in Literature. FIPM, 2019.
4. Demonstration version of control grading materials for the 2020 Unified State Exam in Social Studies. FIPM, 2019.
5. Demonstration version of control grading materials for the 2020 Unified State Exam in History. FIPM, 2019.
6. Demonstration version of control grading materials for the 2020 Unified State Exam in English Language. FIPM, 2019.
7. M. V. Verbitskaya, K. S. Makhmuryan, Yu. B. Kurasovskaya Guidance papers for the chairmen and members of the thematic commissions of the constituent entities of the Russian Federation regarding checking the fulfillment of tasks with a detailed answer on the 2020 USE exam papers. English language. "Letter" section. FIPM, 2020.