

# Технический регламент технологического конкурса в целях реализации Национальной технологической инициативы Up Great «Взаимное обучение на русском языке»

## 1. Общая информация

- 1.1. Настоящий Технический регламент определяет требования к процедуре проведения технологического конкурса ПРО//ЧТЕНИЕ в целях реализации Национальной технологической инициативы (далее Конкурс), обеспечению и другим техническим параметрам Конкурса.
- 1.2. Технический регламент публикуется на официальном Сайте Конкурса (далее - Сайте) <http://ai.upgreat.one/>
- 1.3. Настоящий Технический регламент является документом, детализирующим Конкурсное задание Конкурса, опубликованное на Сайте Конкурса, по регламентам проведения испытаний и процедуре определения результата и победителя Конкурса. Настоящий документ не противоречит Конкурсному заданию, но решает задачу детализации Конкурсного задания до технического уровня; технические детали, указанные в настоящем документе являются необходимым и достаточным описанием задачи Конкурса. Технический регламент определяет требования к процедуре проведения Испытаний Конкурса, Программным комплексам Участников, процедуре контроля и алгоритму определения уровня разработок команд.
- 1.4. Основная цель настоящего документа: обеспечение единого набора документации и регламента участия в Конкурсе, обеспечивающих прозрачность и объективность как для Участников, так и для сторонних наблюдателей.

## 2. Термины и определения

- 2.1. В настоящем Техническом регламенте помимо терминов, перечисленных ниже, используются в том числе, термины в значении, определенном в п. 1.10. Конкурсного задания технологического конкурса в целях реализации Национальной технологической инициативы «Взаимное обучение на русском языке».
- 2.2. **Алгоритм разметки** – алгоритм на основе искусственного интеллекта, используемый в Программном комплексе Участника для преобразования файла с неразмеченным текстом эссе в файл с размеченным текстом того же эссе.
- 2.3. **Алгоритмическая разметка** – разметка текста эссе, сгенерированная алгоритмом разметки в Программном комплексе Участника.

- 2.4. **БД, База данных эссе** – совокупность всех текстов эссе на русском языке, как размеченных, так и не размеченных, представленных в виде файлов, организованная по определённым правилам.
- 2.5. **Выборка данных, дата-сет (ДС)** – набор текстовых файлов эссе, размеченных или неразмеченных.
- 2.6. **Выборка для обучения алгоритмов разметки, Обучающая выборка, Размеченный эталонный открытый ДС** – набор размеченных текстовых файлов эссе, предоставляемый Участникам Конкурса на равных условиях, используемый Участниками для обучения своих алгоритмов разметки.
- 2.7. **Выборка для тестирования алгоритма разметки, Тестовая выборка, Размеченный эталонный закрытый ДС** – набор неразмеченных текстовых файлов эссе, предоставляемый Участникам Конкурса на равных условиях для независимого оценивания точности алгоритмов разметки в рамках Квалификационного этапа. В Системе проверки решений используется выборка размеченных файлов тех же эссе, доступная только Оператору Конкурса и недоступная Участникам.
- 2.8. **Выборка для финального испытания, Финальная выборка, Финальный ДС** – набор неразмеченных текстовых файлов эссе, предоставляемый Участникам Конкурса на равных условиях в рамках этапа Финальных испытаний. Разметка финальной выборки экспертами осуществляется после того, как от всех Участников будут получены алгоритмические разметки всех файлов финальной выборки.
- 2.9. **Испытания, Финальные испытания** – ограниченный по времени период, начинающийся в назначенный Оргкомитетом день, когда Участники проводят с помощью разработанных ими Программных комплексов потоковую разметку текстов, содержащихся в файлах Выборки финального испытания.
- 2.10. **Конкурс** – конкурс «Взаимное обучение на русском языке» (также используется варианты названия «ПРО//ЧТЕНИЕ» и «READ//ABLE») – технологический конкурс, предметом которого является выбор наилучшего решения для автоматического выявления смысловых ошибок в текстовых документах на русском и английском языке. Программные комплексы разрабатываются участниками конкурса.
- 2.11. **КЗ, Конкурсное задание** – основной документ, определяющий цели, задачи и порядок проведения Конкурса. Конкурсное задание утверждается конкурсной комиссией технологических конкурсов в целях реализации Национальной технологической инициативы.
- 2.12. **КО, Классификатор ошибок** – перечень типов ошибок и смысловых блоков, которые могут использоваться в экспертных и алгоритмических разметках текстов эссе. Классификатор ошибок и язык разметки описаны в Приложении 3 к данному Техническому регламенту.
- 2.13. **Лидерборд** — список Участников, ранжированный по убыванию критерия относительной точности алгоритмической разметки (ОТАР).

- 2.14. **ЛК, Личный кабинет** – раздел пользовательского веб-интерфейса Платформы, в котором пользователь может получить доступ к данным, аналитической информации и сервисам Платформы в соответствии с его ролью в Конкурсе и правами доступа.
- 2.15. **Оценка за эссе** – оценка в баллах, вычисляемая на основании разметки текста эссе согласно правилам проверки эссе данного типа (например, оценка за мини-сочинение по обществознанию в рамках ЕГЭ). Оценка может быть вычислена как по экспертной разметке, так и по алгоритмической.
- 2.16. **Ошибка** – фрагмент текста эссе, обладающий обязательными параметрами начала, конца и типа ошибки согласно КО, а также необязательными параметрами подтипа ошибки согласно КО, комментария (для детализации подтипа ошибки без привязки к тексту эссе), пояснения (для объяснения ошибки с привязкой к тексту эссе), исправления и тега.
- 2.17. **Параметры Конкурса** — числовые параметры, влияющие на значения автоматически вычисляемых критериев относительной точности алгоритмической разметки (ОТАР). Параметры Конкурса объявляются перед началом каждого цикла и перед Финальными испытаниями, а также могут быть изменены решением Оргкомитета при согласовании с Судейской коллегией и Технической комиссией Конкурса.
- 2.18. **Платформа** – сайт в компьютерной сети Интернет, обеспечивающий доступ к данным, аналитической информации и сервисам для всех категорий пользователей, вовлечённых в участие в Конкурсе и проведение Конкурса. В частности, Участникам предоставляется возможность тестирования Алгоритмов разметки, Экспертам предоставляется возможность размечать тексты эссе в Программе Разметчике, и т.д. Платформа имеет интерфейс на русском и английском языках.
- 2.19. **Программа Разметчик** – программное обеспечение с веб-интерфейсом на базе Платформы, которое позволяет проводить разметку текстовых файлов, выделяя и описывая ошибки и смысловые блоки согласно КО. Доступ к полному функционалу Разметчика имеют определённые Оргкомитетом категории пользователей. Участники имеют права на просмотр разметки, но не могут самостоятельно делать разметку в интерфейсе Разметчика.
- 2.20. **ПКУ, Программный комплекс Участника** – разработанный Командой Участника Конкурса комплекс программного и аппаратного обеспечения, который в автоматическом режиме принимает на входе неразмеченный файл эссе и выдаёт на выходе размеченный файл эссе, сгенерированный Алгоритмом разметки.
- 2.21. **ПСР, Программа сравнения решений** – программное обеспечение с веб-интерфейсом на базе Платформы, позволяющее пользователю визуально сравнить два размеченных файла одного и того же эссе (например, алгоритмическую разметку с экспертной или две экспертных разметки от разных Экспертов).
- 2.22. **Размеченный файл, разметка файла** – текстовый файл эссе, размеченный в соответствии с правилами, описанными в Приложении 3 к данному Техническому

регламенту, имеющий расширение \*.txt и кодировку UTF-8 без спецификаций. Разметка может быть экспертной или алгоритмической.

- 2.23. **СПР, Система проверки решений** – автоматическая система проверки, которая принимает выборку файлов с алгоритмической разметкой, сравнивает их с соответствующей экспертной разметкой и выдаёт результат сравнения в форме аналитического отчёта. Результат проверки может быть визуально проконтролирован для любого файла из выборки в Программе сравнения решений.
- 2.24. **Смысловый блок** – фрагмент текста эссе, обладающий обязательными параметрами начала, конца и типа смыслового блока согласно КО. Смысловые блоки не являются ошибками, но их наличие может влиять на оценку за эссе.
- 2.25. **Технологический барьер** – задача создания стабильно работающего Программного комплекса для выявления смысловых ошибок в академических эссе, средняя точность которого не хуже средней точности проверяющего эксперта, действующего в условиях ограниченного времени. Преодоление Технологического барьера является условием для определения победителя Конкурса.
- 2.26. **Участник, Участник Конкурса** — российское или иностранное юридическое или физическое лицо, или объединение таких лиц, чья заявка на участие в Конкурсе прошла одобрение Оргкомитетом. Все требования положений Конкурсного задания распространяются на Участника также, как и на Команду.
- 2.27. **Файл, файл с текстом эссе** – размеченный или неразмеченный текстовый файл, имеющий расширение \*.txt и кодировку UTF-8 без спецификаций.
- 2.28. **Цикл** – единый логический блок Конкурса, который состоит из этапов квалификации, финального испытания и подведения итогов.
- 2.29. **Эксперт, Специалист** – привлечённый профильный специалист общеобразовательного учреждения, который проверяет текст эссе и формирует в результате проверки размеченный файл.
- 2.30. **Экспертная разметка** – разметка текста эссе, сформированная экспертом в результате проверки вручную неразмеченного файла данного эссе. В Платформе реализован специальный сервис для автоматизации процесса экспертной разметки.
- 2.31. **Эссе, текст, текст эссе** – оцениваемая письменная работа учащегося в виде развернутого и аргументированного текста, написанная по заданной проблеме или теме (например, сочинение, написанное учеником 11 класса общеобразовательной школы в рамках единого государственного экзамена по русскому языку, литературе, обществознанию или истории).

### 3. Участники конкурса и условия участия

- 3.1. Для участия в конкурсе Участник подаёт заявку путем заполнения электронной формы на Сайте конкурса, согласно п.3.3 и Приложению 1 КЗ. Порядок формирования Команды Участника описан в п.4 КЗ.

- 3.2. Условия, при которых участник может быть дисквалифицирован, описаны в п.3.4 КЗ.
- 3.3. Дополнительные условия участия описаны в п.3.5 КЗ.

#### **4. Циклы и этапы конкурса**

- 4.1. Конкурс проводится путём повторения Циклов, согласно п.3.2 и п.5.1 КЗ. Каждый Цикл включает в себя Отборочный этап (п.5.2 КЗ), Квалификационный этап (п.5.3 КЗ), этап Финальных испытаний (п.5.4 КЗ) и Подведение итогов (п.6 КЗ).
- 4.2. После успешного прохождения Отборочного этапа участник получает через Личный кабинет Платформы доступ к исходным данным и исходным кодам общих программных модулей, описанных в Приложении 2, которые должны использоваться всеми Участниками Конкурса:
  - 4.2.1. Исходные данные – обучающая выборка размеченных текстов эссе.
  - 4.2.2. Программный модуль для парсинга, сравнения и оценивания разметок.
  - 4.2.3. Программный модуль для обмена данными с Платформой.
  - 4.2.4. Пилотное приложение, реализующее минимальный набор функций ПКУ.
- 4.3. Использование общих программных модулей является обязательным, поскольку обеспечивает равные условия для всех Участников Конкурса. Участники Конкурса не имеют право самостоятельно вносить изменения в общие программные модули. В случаях обнаружения ошибок в них Участник может подать в Оргкомитет по электронной почте в свободной форме заявку «О внесении изменений в общие программные модули», описав в ней обнаруженные ошибки и/или свои предложения по улучшению работы модулей.
- 4.4. На протяжении Квалификационного этапа Участник может многократно проходить два вида квалификации – алгоритмическую и техническую. Для этого ПКУ открывает сессии обмена данными с Платформой. Целью алгоритмической квалификации является анализ точности алгоритмов разметки на независимой Тестовой выборке. Целью технической квалификации является тестирование ПКУ в режиме реального времени, максимально приближенном к Финальным испытаниям.
- 4.5. На этапе Финальных испытаний Участник может открыть и провести до конца только одну сессию обмена данными с Платформой.
- 4.6. Функции квалификации и участия в Финальных испытаниях должны быть реализованы в ПКУ с использованием общего программного модуля обмена данными с Платформой. Для облегчения данной технологической задачи Участникам предоставляется пилотный программный пакет с минимальным набором функций ПКУ.
- 4.7. Оргкомитет оставляет за собой право обновлять Обучающую и Тестовую выборки, а также исходные коды общих программных модулей как между Циклами, так и во время Квалификационного этапа, но не позднее, чем за 10 дней

до Финальных испытаний. В случае обновления Оргкомитет оповещает всех Участников путём публикации на сайте Платформы, сообщая цель и суть произведённых модификаций (например, устранена ошибка в коде, добавлена полезная функция по запросам участников, увеличен объём размеченных данных, и т.д.).

- 4.8. Условием окончания Конкурса является преодоление Технологического барьера одним из Участников по итогам Финальных испытаний (п.6 КЗ). Если ни одна из Команд не преодолела Технологический барьер в текущем Цикле, то в установленные сроки запускается следующий Цикл Конкурса (п.6.1.6 КЗ). Если ни одна из Команд не преодолела Технологический барьер в последнем Цикле в рамках сроков проведения Конкурса, то Технологический барьер считается не преодоленным (п.6.1.7 КЗ).

## **5. Конкурсная задача и оценивание точности решений Участников**

- 5.1. Конкурсная задача заключается в преодолении Технологического барьера путём построения Алгоритма разметки по Обучающей выборке размеченных файлов эссе.
- 5.2. Алгоритм разметки преодолевает Технологический барьер, если его средняя точность алгоритмической разметки (СТАР) на Финальной выборке (во время Финальных испытаний) не хуже средней точности экспертной разметки (СТЭР), вычисленной по экспертным разметкам, полученным в условиях ограниченного времени.
- 5.3. Средняя точность алгоритмической разметки (СТАР) оценивается по выборке, в которой для каждого файла эссе имеется одна или несколько экспертных разметок. СТАР определяется как взвешенное среднее парной точности алгоритмической разметки файла относительно экспертной разметки того же эссе.
- 5.4. Средняя точность экспертной разметки (СТЭР) оценивается по выборке, в которой для каждого файла эссе имеется, как минимум, две экспертные разметки. СТЭР определяется как взвешенное среднее парной точности экспертной разметки файла относительно экспертной разметки того же эссе другим экспертом.
- 5.5. Парная точность разметки относительно другой разметки того же эссе вычисляется согласно алгоритму, описанному в Приложении 1 к данному Техническому Регламенту. Алгоритм основан на вычислении семи метрик, значения которых усредняются с весами метрик  $w_1 \dots w_7$ .
- 5.6. Весовой коэффициент, используемый при взвешенном усреднении парных точностей в СТАР и СТЭР, вычисляется следующим образом. Максимальная парная точность алгоритмической разметки по заданному эссе учитывается с весом  $(1 - H)$ , остальные разметки данного эссе учитываются с весом  $H$ . Минимальная парная точность экспертной разметки по заданному эссе учитывается с весом  $(1 - H)$ , остальные разметки данного эссе учитываются с весом  $H$ . *Параметр жёсткости  $H$  (hardness parameter) принимает значения от 0 до 1 и позволяет управлять жёсткостью критерия. Чем больше  $H$ , тем меньше*

СТАР и больше СТЭР, и тем труднее преодолеть Технологический барьер. При  $H = 1$  СТАР и СТЭР вычисляются как средние арифметические парных точностей. Рекомендуемое начальное значение  $H = 0$ .

- 5.7. Относительная точность алгоритмической разметки (ОТАР) определяется по заданной выборке эссе как отношение  $ОТАР = СТАР / СТЭР * 100\%$ . Значение относительной точности, большее или равное 100%, соответствует преодолению Технологического барьера данным Алгоритмом. Относительная точность используется при построении рейтингов Участников и Алгоритмов на всех этапах Конкурса и при подведении итогов.
- 5.8. Веса метрик  $w_1 \dots w_7$  и параметр жёсткости  $H$  являются *параметрами Конкурса*, которые могут быть изменены решением Оргкомитета при согласовании с Судейской коллегией и Технической комиссией Конкурса. Оргкомитет включает значения этих параметров в объявления о начале каждого цикла и о Финальных испытаниях. В случае внеочередного изменения параметров Конкурса Оргкомитет уведомляет Участников не менее чем за неделю до вступления изменения в силу.

## **6. Квалификационный этап**

- 6.1. Данный раздел дополняет п.5.3 КЗ.
- 6.2. В Квалификационном этапе могут участвовать все зарегистрированные Участники, прошедшие Отборочный этап, получившие уведомление о регистрации от Оргкомитета (п.5.2.4 КЗ), разработавшие Программный комплекс Участника (ПКУ) и интегрировавшие в него собственные Алгоритмы разметки.
- 6.3. На протяжении Квалификационного этапа Участник может многократно проходить два вида квалификации – алгоритмическую и техническую, открывая сессии обмена данными с Платформой.
- 6.4. Алгоритмическая квалификация проводится с целью проверки качества, анализа и сравнения собственных алгоритмов разметки на Обучающей выборке или на независимой Тестовой выборке. Объём Тестовой выборки составляет не менее 300 файлов. Для проведения алгоритмической квалификации Участник может открыть сессию обмена данными с Платформой в любой момент. При открытии сессии выбирается тип выборки (Обучающая или Тестовая) и ограничение на число файлов. Сервер Платформы отправляет ПКУ неразмеченные файлы и принимает размеченные файлы от ПКУ без намеренных задержек по времени. Количество сессий ограничено 20 сессиями в сутки.
- 6.5. По окончании квалификационной сессии Участник может просмотреть автоматически сформированный отчёт об алгоритмической квалификации в Личном кабинете. Отчёт показывает:
  - 6.5.1. оценки средней и относительной точности алгоритма (СТАР, ОТАР):
    - 6.5.1.1. по всей выборке,
    - 6.5.1.2. в разрезе типов эссе,

- 6.5.1.3. в разрезе метрик точности (см. Приложение 1);
- 6.5.2. список всех обработанных эссе, ранжированный по выбираемой пользователем метрике точности;
- 6.5.3. сравнение алгоритмической разметки с экспертной для любого эссе, выбираемого пользователем из списка обработанных эссе через ПСР (только в случае, если алгоритмическая квалификация проводилась на Обучающей выборке);
- 6.5.4. заключение о степени готовности ПКУ к проведению Финального испытания.
- 6.6. Техническая квалификация проводится с целью отработки функционирования ПКУ в режиме реального времени, максимально приближенном к Финальным испытаниям. Режим технической квалификации запускается сервером Платформы автоматически по расписанию, каждые четыре часа. Частота отдачи файлов на разметку и их количество может варьироваться от запуска к запуску в целях тестирования. Файлы выбираются случайным образом из Обучающей выборки. Объём выборки составляет не менее 30 файлов.
- 6.7. По окончании квалификационной сессии Участник может просмотреть автоматически сформированный отчёт о технической квалификации в Личном кабинете. Отчёт показывает:
  - 6.7.1. среднее и максимальное время обработки файла Программным комплексом Участника;
  - 6.7.2. среднее и максимальное время задержки между моментом открытия доступа к файлу на сервере Платформы и моментом отдачи файла по запросу от ПКУ;
  - 6.7.3. доля файлов, аннулированных по причинам несоблюдения ограничений по скорости приёма и обработки файлов;
  - 6.7.4. заключение о степени готовности ПКУ к проведению Финального испытания.
- 6.8. Техническая информация о порядке организации сессии обмена данными с Платформой во время алгоритмической и технической квалификации приведена в Приложении 2 к настоящему Техническому регламенту.
- 6.9. Квалификационный этап считается успешно пройденным, если Участник получил положительные заключения о степени готовности ПКУ к проведению Финального испытания, хотя бы по одному разу для алгоритмической и для технической квалификации.
- 6.10. По результатам алгоритмической квалификации Тестовой выборки строится квалификационный лидерборд — список Участников, ранжированный по убыванию критерия ОТАР. Для ранжирования используется лучшее значение ОТАР из всех алгоритмических квалификаций, в которых ПКУ данного Участника обработал не менее 95% файлов эссе Тестовой выборки.

## **7. Этап финальных испытаний**

- 7.1. Данный раздел дополняет п.5.4 КЗ.



- 7.2. В этапе Финальных испытаний могут участвовать все Участники, успешно прошедшие Отборочный и Квалификационный этапы.
- 7.3. Финальные испытания проводятся на Финальной выборке неразмеченных эссе (финальный ДС). Экспертная разметка этих эссе производится в течение 15 рабочих дней после завершения алгоритмической разметки Участниками Конкурса. Таким образом, во время проведения Финальных испытаний экспертная разметка Финальной выборки не существует.
- 7.4. Для участия в Финальных испытаниях Участник должен запустить сессию обмена данными средствами Программного комплекса Участника до момента старта Финальных испытаний.
- 7.5. Во время Финальных испытаний неразмеченные файлы из Финальной выборки становятся доступны по расписанию, начиная с момента старта, через равные промежутки времени в 1 минуту для обеспечения равных условий всем Участникам Конкурса и исключения любых возможностей для организации ручной разметки данных. Расписание Финальных испытаний (дата-время старта и периодичность отдачи файлов) объявляется не менее, чем за 10 дней до их начала.
- 7.6. В ходе сессии обмена данными Программном комплексе Участника (ПКУ) должен оперативно загружать и обрабатывать неразмеченный файлы с сервера Платформы, затем отдавать файлы с алгоритмической разметкой в Систему проверки решений (СПР). При превышении порогового времени задержки в 1 минуту (после открытия общего доступа к данному файлу) файл аннулируется для данного Участника и не учитывается при подсчёте усреднённых критериев СТАР и ОТАР.
- 7.7. Если доля аннулированных файлов Участника во время Финальных испытаний превышает 5% от объёма Финальной выборки, то решение (вся совокупность отданных файлов) данного Участника полностью аннулируется и не участвует в подведении итогов Конкурса.
- 7.8. По окончании сессии Участник может просмотреть в Личном кабинете автоматически сформированный отчёт, по структуре аналогичный отчёту о технической квалификации.
- 7.9. Техническая информация о порядке организации сессии обмена данными с Платформой во время Финальных испытаний приведена в Приложении 2 к настоящему Техническому регламенту.
- 7.10. По окончании обработки файлов всеми Участниками Финального испытания все сессии обмена данными закрываются, все файлы с алгоритмической разметкой, полученные от Участников, сохраняются в Базу данных эссе (формируя закрытый ДС с финальными решениями Участников). С этого момента стартует процесс экспертной разметки Финальной выборки. Каждый файл эссе проверяется двумя независимыми экспертами. В случае существенного расхождения их оценок назначается третий эксперт, который формирует третью разметку. Правила назначения третьего эксперта зависят от типа эссе и приводятся в Приложении 3. Все эксперты осуществляют проверку в условиях ограниченного времени (время проверки может ограничиваться сверху в зависимости от типа эссе). Третьему

эксперту предоставляются данные двух предыдущих проверок и возможность сравнить их с помощью ПСР. Экспертные разметки двух или трёх экспертов сохраняются в Базу Данных эссе (формируя Закрытый финальный ДС) и используются для вычисления критериев СТАР и ОТАР по Финальной выборке.

- 7.11. По окончании экспертной разметки Финальной выборки в течение не более чем 5 рабочих дней подводятся итоги Финального этапа. Всем Участникам предоставляется доступ к экспертным разметкам Финальной выборки. Участник может просмотреть в Личном кабинете автоматически сформированный отчёт, по структуре аналогичный отчёту об алгоритмической квалификации.
- 7.12. По итогам Финального этапа строится финальный лидерборд — список Участников, ранжированный по убыванию критерия ОТАР, рассчитанного по Финальной выборке. Аннулированные решения не участвуют в финальном лидерборде. Решения со значением ОТАР не менее 100% считаются преодолевшими Технологический барьер.
- 7.13. Порядок подведения итогов описан в п.6 КЗ.
- 7.14. В соответствии с п. 8.2. КЗ эссе размеченные ПКУ Участников во время Испытаний (но не само программное обеспечение, являющееся интеллектуальной собственностью команд) передаются Организаторам на условиях открытой лицензии Creative Commons Attribution 4.0 International (CC BY 4.0) для последующей публикации на Сайте конкурса с целью обеспечения прозрачности в определении победителей и призёров.
- 7.15. Порядок определения премий победителю и призёрам описан в п.7 КЗ.
- 7.16. Порядок перезапуска Испытаний:
  - 7.16.1. Испытания могут быть перезапущены в случае массового технического сбоя по решению Оргкомитета, на основании соответствующих протоколов Судейской Коллегии и Технической Комиссии.
  - 7.16.2. Перезапуск может быть назначен на следующий день, о чём сообщается отдельно на официальном сайте.
  - 7.16.3. Файлы на перезапуск новые. Уже отданные не учитываются.
  - 7.16.4. Результаты по отданным файлам не учитываются.

## **8. Судейская коллегия и Техническая комиссия**

- 8.1. Данный раздел уточняет пункты 5.3.10.1, 5.4.4, 5.4.14 КЗ.
- 8.2. На Квалификационном этапе:
  - 8.2.1. Судейская коллегия проверяет автоматически формируемые заключения о степени готовности ПКУ к проведению Финального испытания по результатам алгоритмической и технической квалификации.
  - 8.2.2. Результаты испытаний верифицируются Технической комиссией и передаются на утверждение Оргкомитету.

- 8.2.3. Оргкомитет выносит окончательные решения о допуске Участников к Финальному испытанию.
- 8.3. На этапе Финальных испытаний
  - 8.3.1. Техническая комиссия проверяет корректность полученных от Участников файлов, выявляет условия и причины аннулирований, если таковые имеются.
  - 8.3.2. Техническая комиссия контролирует процесс распределения заданий для экспертной разметки Финальной выборки.
  - 8.3.3. По окончании экспертной разметки Финальной выборки Судейская коллегия проверяет автоматически формируемые отчеты о решениях Участников и условия преодоления технологического барьера.
  - 8.3.4. Результаты испытаний верифицируются Технической комиссией и передаются на утверждение Оргкомитету.
  - 8.3.5. Оргкомитет выносит решения по итогам Финального испытания.

## **9. Права Оргкомитета**

- 9.1. Данный раздел уточняет пункты 3.1.8, 3.4.1, 5.1.5, 5.4.8, 8.1.4 КЗ.
- 9.2. Оргкомитет оставляет за собой право обновлять Обучающую и Тестовую выборки, а также исходные коды общих программных модулей, как между Циклами, так и во время Квалификационного этапа, но не позднее, чем за 10 дней до Финальных испытаний. В случае обновления Оргкомитет оповещает всех Участников путём публикации на сайте Платформы, сообщая цель и суть произведённых модификаций.
- 9.3. Оргкомитет имеет право проверить решение команды очно.

## **10. Протесты и апелляции**

- 10.1. Данный раздел уточняет пункт 6.2 КЗ.
- 10.2. Если в ходе Финальных испытаний аннулировано более 5% файлов Участника, то данное решение Участника не может участвовать в подведении итогов Финального испытания без возможности подачи протеста или апелляции. При этом Участник не дисквалифицируется и может участвовать в следующих Этапах Конкурса.

## **11. Участникам запрещается**

- 11.1. Данный раздел дополняет п.3.4 КЗ.
- 11.2. Участнику запрещается осуществлять экспертную разметку Тестовой и Финальной выборки вручную, с помощью краудсорсинга или любыми иными способами с привлечением труда людей.
- 11.3. Запрещается привлекать труд людей для извлечения каких-либо дополнительных данных из текстов эссе в неразмеченных выборках – Тестовой и Финальной.

Анализ и разметка этих текстов должны производиться Участником исключительно алгоритмически.

- 11.4. Запрещается самостоятельно модифицировать общие программные модули.
- 11.5. Запрещается на этапе Финальных испытаний отправлять файл с алгоритмической разметкой одного и того же эссе более 10 раз. Все последующие отправки будут проигнорированы, в качестве окончательного варианта будет рассматриваться последняя отправка.
- 11.6. Запрещается на этапе Финальных испытаний отправлять вместо алгоритмической разметки файла эссе какую-либо иную информацию, в том числе алгоритмическую разметку другого эссе. Такие отправки будут проигнорированы.
- 11.7. В случае нарушения перечисленных запретов Оргкомитет по запросу Судейской коллегии или Технической комиссии имеет право отстранить Команду от участия во всех последующих Циклах Конкурса.

## **12. Обеспечение безопасности и дополнительные условия**

- 12.1. Требования к безопасности и экологии, ограничения на раскрытие и распространение информации, права на интеллектуальную собственность и порядок изменения условий Конкурса описаны в п. 8 КЗ.
- 12.2. Деятельность Участников в рамках Конкурса должна соответствовать действующим на территории Российской Федерации природоохранным нормам, требованиям техники безопасности.
- 12.3. В случае проведения очных мероприятий Оргкомитет Конкурса предоставляет Участникам правила по технике безопасности и охране окружающей среды, с которыми все члены команды Участника должны ознакомиться под подпись и соблюдать.

## **13. Список приложений**

- 13.1. Метрики точности разметки
- 13.2. Программные модули
- 13.3. Классификатор ошибок и язык разметки данных

# Приложение 1. Метрики точности разметки

Парная точность  $M(X, Y)$  разметки  $X$  относительно разметки  $Y$  оценивается по метрикам  $M_1(X, Y) \dots M_7(X, Y)$ , которые затем усредняются с весами  $w_1 \dots w_7$ .

Метрики  $M_1 \dots M_7$  измеряются в процентах от 0% до 100%. Чем больше, тем точнее совпадение. Итоговая метрика *парной точности* разметки  $X$  относительно разметки  $Y$  вычисляется как взвешенное среднее семи метрик:

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

Весовые коэффициенты  $w_i$  позволяют задавать степень значимости каждой метрики. Возможно усреднение только по части метрик, тогда  $w_i = 0$  для остальной части метрик. Например, при отсутствии или недостаточной надёжности экспертных оценок для метрики  $M_7$  её можно исключить из формулы, и включить обратно, когда в экспертных разметках появится достаточное количество данных о пояснениях.

Набор весов  $w_1 \dots w_7$  является параметром методики измерения точности разметки.

Для вычисления метрик  $M_2 \dots M_6$  производится сопоставление фрагментов в двух разметках. Алгоритм сопоставления фрагментов описан в разделе 3 данного приложения.

*Оптимистичная относительная парная точность* алгоритмической разметки одного эссе, при сравнении со всем множеством  $\{\mathcal{E}\}$  экспертных разметок данного эссе:

$$M_{\text{опт}}(A, \{\mathcal{E}\}) = \frac{\max_{\mathcal{E}} M(A, \mathcal{E})}{\min_{\mathcal{E}, \mathcal{E}'} M(\mathcal{E}, \mathcal{E}')} 100\%,$$

где максимум в числителе берётся по всем экспертным разметкам, минимум в знаменателе — по всем парам экспертных разметок данного эссе.

*Средняя относительная парная точность* алгоритмической разметки одного эссе, при сравнении со всем множеством  $\{\mathcal{E}\}$  экспертных разметок данного эссе:

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

где среднее в числителе берётся по всем экспертным разметкам, среднее в знаменателе — по всем парам экспертных разметок данного эссе.

*Относительная Точность Алгоритмической Разметки* одного эссе с учётом параметра жёсткости  $H$ :

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

Параметр жёсткости критерия  $H$  принимает значения от 0 до 1; чем меньше  $H$ , тем выше и оптимистичнее оценка ОТАР. При  $H = 0$  ОТАР =  $M_{\text{опт}}$ , при  $H = 1$  ОТАР =  $M_{\text{ср}}$ .

Усредняя числители по выборке эссе, получим *Среднюю Точность Алгоритмической Разметки*, с учётом параметра жёсткости  $H$ :

$$\text{СТАР} = \text{avr}_{\text{эссе}} \left( H \text{avr}_{\mathcal{A}, \mathcal{E}} M(\mathcal{A}, \mathcal{E}) + (1 - H) \max_{\mathcal{A}, \mathcal{E}} M(\mathcal{A}, \mathcal{E}) \right)$$

Усредняя знаменатели по выборке эссе, получим *Среднюю Точность Экспертной Разметки*, с учётом параметра жёсткости  $H$ :

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

*Относительная Точность Алгоритмической Разметки* по выборке эссе, с учётом параметра жёсткости  $H$ :

$$\text{ОТАР} = \frac{\text{СТАР}}{\text{СТЭР}} 100\%$$

Таким образом, ОТАР может быть вычислена как по отдельному эссе, так и по любому множеству эссе, в том числе по обучающей, тестовой и финальной выборке.

### **М1. Точность предсказания оценки за эссе.**

Пусть  $K(X)$  и  $K(Y)$  – две оценки за эссе в баллах, вычисленные по разметкам  $X$  и  $Y$  соответственно. Тогда *точность предсказания оценки* за эссе

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

где  $\max K$  – максимальная возможная оценка за эссе данного типа.

Методика вычисления оценок  $K(X)$  и максимальная оценка  $K$  зависит от типа эссе. В Приложении 3 приведены формулы для вычисления оценки в баллах за эссе по разметке, для различных типов эссе.

### **М2. Точность и полнота поиска фрагментов.**

*Точность поиска* определяется как доля фрагментов разметки  $X$ , имеющих сопоставленный фрагмент в разметке  $Y$ .

*Полнота поиска* определяется как доля фрагментов разметки  $Y$ , имеющих сопоставленный фрагмент в разметке  $X$ .

Агрегированная метрика *точности и полноты* определяется как их гармоническое среднее ( $F_1$ -мера).

### **М3. Точность предсказания кодов.**

Доля фрагментов разметки  $X$ , имеющих сопоставленный фрагмент в разметке  $Y$  с равным кодом фрагмента (типом ошибки или смыслового блока).

### **М4. Точность предсказания подтипов ошибок и комментариев.**

Доля фрагментов разметки  $X$ , имеющих сопоставленный фрагмент в разметке  $Y$  с равным подтипом ошибки или с комментарием, который является парафразом.

### **М5. Точность локализации фрагментов.**

Средняя точность совпадения фрагментов разметки  $X$  с сопоставленными им фрагментами разметки  $Y$ . Точность совпадения двух фрагментов вычисляется как мерой Жаккара –

отношение числа слов в пересечении к числу слов в объединении двух фрагментов. Если фрагмент разметки  $X$  не имеет сопоставленного ему фрагмента разметки  $Y$ , то точность их совпадения принимается равной 0%.

#### **М6. Точность исправлений ошибок.**

Доля фрагментов разметки  $X$ , имеющих исправление и сопоставленный фрагмент в разметке  $Y$  с таким же исправлением.

#### **М7. Точность пояснений.**

Средняя экспертная оценка точности пояснения по всем фрагментам разметки  $X$ , имеющим пояснения. Это единственная метрика, основанная не на автоматическом сравнении с разметкой  $Y$ , а на оценках экспертов.

Эксперты оценивают каждое пояснение в проверяемой алгоритмической разметке от 0 до 5 баллов. Для получения точности в процентах число баллов умножается на 20%. Суммарный балл складывается из ответов на следующие вопросы (да=1, нет=0) относительно данного пояснения:

- скорее всего будет понятно автору эссе
- правильно объясняет суть ошибки
- не оставляет возможностей для апелляции
- обращается к тексту работы и конкретно к выделенному фрагменту
- решает педагогическую задачу, помогает избежать подобных ошибок впредь

Пояснение в экспертной разметке автоматически получает 100%.

## Приложение 2. Программные модули

### 1. Программные модули для парсинга, сравнения и оценивания разметок

Следующие алгоритмы предоставляются Участникам Конкурса для обеспечения равных условий. Они должны быть разработаны и реализованы до начала Конкурса.

#### 1. Алгоритм парсинга (синтаксического разбора) разметки:

**Вход:**

размеченный текст;

**Выход:**

список значений полей метаописания и список фрагментов;

для каждого фрагмента:

позиции начала и конца в исходном неразмеченном тексте,

текст фрагмента,

код(ы),

комментарий,

пояснение,

исправление,

тег.

#### 2. Алгоритм поиска оптимального соответствия между фрагментами:

**Вход:**

две разметки одного и того же текста в виде двух списков фрагментов;

**Выход:**

список пар номеров соответствующих фрагментов из первого и второго списка.

#### 3. Алгоритм вычисления критериев и итоговой оценки за эссе:

**Вход:**

список значений полей метаописания;

разметка в виде списка фрагментов;

**Выход:**

список значений критериев в баллах;

итоговая оценка согласно правилам оценивания данного типа эссе.

Формулы, по которым вычисляются критерии и итоговая оценка за эссе в баллах, зависят от типа эссе. В Приложении 3 к Техническому регламенту формулы оценки в баллах приведены для пяти типов эссе - сочинений ЕГЭ по русскому языку, литературе, обществознанию, истории и английскому языку.

#### 4. Алгоритм вычисления метрики парной точности разметки:

**Вход:**

две разметки одного и того же текста в виде двух списков фрагментов;

**Выход:**

значение парной точности разметки;

значения метрик M1–M7.

### 2. Программный модуль обмена данными между ПКУ и СПР

Программный модуль обмена данными между Программным комплексом Участника (ПКУ) и Системой проверки решений (СПР) предоставляется всем Участникам, прошедшим Отборочный этап.



Основным назначением модуля является обеспечение синхронной раздачи неразмеченных файлов Участникам Конкурса на этапе Финального испытания, чтобы исключить любую возможность ручной разметки данных.

Модуль используется также на Квалификационном этапе в двух режимах – *алгоритмической квалификации* (при которой в основном обрабатываются вопросы качества алгоритмов) и *технической квалификации* (при которой в основном обрабатываются вопросы надёжности ПКУ и быстродействия алгоритмов).

Модуль позволяет реализовать несколько сценариев коммуникации между ПКУ и СПР. В ходе сессии обмена данными ПКУ выступает в роли клиента, СПР – в роли сервера.

#### **Сценарий сессии обмена данными на этапе алгоритмической квалификации**

- ПКУ: Запрос на открытие сессии.
- СПР: Проверка Участника и, в случае успеха, передача разрешения на открытие сессии.
- ПКУ: Запрос на получение следующего файла неразмеченного эссе.
- СПР: Передача файла неразмеченного эссе или сообщения о завершении выборки.
- ПКУ: Отправка размеченного файла того же эссе.
- СПР: Замер времени отклика, сохранение разметки, оценивание качества разметки.
- СПР: По завершении выборки генерация и сохранение аналитического отчёта.

#### **Сценарий сессии обмена данными на этапе Финальных испытаний**

- ПКУ: Запрос на открытие сессии (должен поступить не позднее 2 минут после старта Финальных испытаний).
- СПР: Проверка Участника и, в случае успеха, передача разрешения на открытие сессии.
- ПКУ: Запрос на получение следующего файла неразмеченного эссе (должен поступить не позднее 10 секунд после открытия доступа к данному файлу).
- СПР: Передача файла неразмеченного эссе или сообщения о завершении выборки.
- ПКУ: Отправка размеченного файла того же эссе (должна произойти не позднее 50 секунд после отправки неразмеченного файла).
- СПР: Замер времени отклика, сохранение разметки, оценивание качества разметки.
- СПР: По завершении выборки генерация и сохранение аналитического отчёта.

Если СПР не получает от ПКУ запрос на получение файла неразмеченного эссе в течение отведённого времени или если СПР не получает от ПКУ размеченный файл в течение отведённого времени, то данный файл аннулируется для данного Участника и не будет учитываться при подсчёте усреднённых критериев СТАР и ОТАР.

#### **Сценарий сессии обмена данными на этапе технической квалификации**

Сценарий сессии обмена данными при технической квалификации на Квалификационном этапе может отличаться от сценария Финальных испытаний только параметрами временных задержек.

### 3. Алгоритм вычисления критериев и итоговой оценки за эссе

*Разметкой* будем называть последовательность  $n$  фрагментов  $X = \{x_i = (B_i, E_i, C_i)\}_{i=1}^n$ , где  $B_i$  – позиция начала фрагмента,  $E_i$  – позиция конца фрагмента,  $C_i$  – тип фрагмента.

*Соответствием* двух разметок  $X = \{x_i = (B_i, E_i, C_i)\}_{i=1}^n$  и  $Y = \{y_k = (B_k, E_k, C_k)\}_{k=1}^m$  будем называть множество  $M$  пар фрагментов  $(i, k)$ , такое, что каждому  $x_i$  из  $X$  соответствует не более одного  $y_k$  и каждому  $y_k$  из  $Y$  соответствует не более одного  $x_i$ . Если для фрагмента  $x_i$  нет соответствия, будем записывать « $x_i \rightarrow \emptyset$ ».

Для произвольной пары фрагментов  $(x_i, y_k)$  определим *расстояние Жаккара*:

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

Расстояние Жаккара  $J_{ik}$  принимает значения от 0 до 1. Если фрагменты  $x_i$  и  $y_k$  совпадают, то  $J_{ik} = 0$ . Если фрагменты не пересекаются, то  $J_{ik} = 1$ .

Определим *матрицу потерь*  $L[i, k]$  размера  $n \times m$ :

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

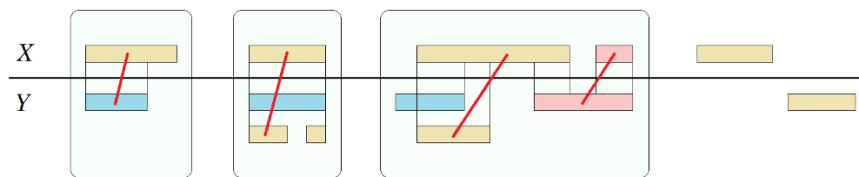
$L[i, k]$	Примеры ситуаций соответствия пары фрагментов $(x_i, y_k)$	
0	фрагменты совпадают и имеют одинаковый тип	
0..1	фрагменты пересекаются, имеют общее начало и одинаковый тип	
1	фрагменты совпадают, но имеют разные типы	
1..2	фрагменты пересекаются и имеют либо разное начало, либо разные типы	
2	<i>фрагменты не образуют соответствия</i>	
2..3	фрагменты пересекаются, имеют разное начало и разные типы	
3	фрагменты не пересекаются и имеют одинаковый тип	
4	фрагменты не пересекаются и имеют разные типы	

Требуется найти соответствие между разметками, минимизирующее сумму потерь:

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

Задача поиска оптимального соответствия между разметками является обобщённой задачей о назначениях, для решения которой может быть применён один из вариантов венгерского алгоритма. Однако в нашем случае функция потерь устроена настолько специфически, что существует быстрый переборный алгоритм поиска оптимального соответствия.

Алгоритм проще описывать в терминах теории графов. Задан *двудольный граф*  $G$ , долями которого являются множества фрагментов  $X$  и  $Y$ , рёбрами соединяются пересекающиеся фрагменты  $(i, k)$ :  $J_{ik} < 1$ . Если пара фрагментов не пересекается, то она не может входить в оптимальное решение, поскольку данную пару выгоднее вообще не связывать. Таким образом, задача минимизации суммарных потерь является задачей поиска оптимального *паросочетания* – подграфа  $M$  попарно несмежных рёбер в двудольном графе  $G$ .



На рисунке показан пример соответствия двух разметок: верхняя разметка  $X$  состоит из 5 фрагментов, нижняя  $Y$  состоит из 8 фрагментов. Оптимальное паросочетание состоит из 4 рёбер; несвязанными остаются 1 фрагмент из верхней разметки и 4 из нижней.

Задача поиска оптимального паросочетания решается по отдельности для каждой компоненты связности графа, что сильно сокращает перебор. Внутри каждой компоненты производится полный перебор паросочетаний. При переносе ребра из исходного графа в паросочетание компонента может распасться на ещё меньшие компоненты связности, которые в свою очередь можно обработать по отдельности, ещё сильнее сократив перебор. Эта идея реализуется рекурсивным алгоритмом полного перебора всех паросочетаний в связанных компонентах заданного двудольного графа. Для ускорения перебора можно перебирать рёбра в порядке убывания количества смежных с ними рёбер.

#### Алгоритм:

1. задан двудольный граф  $G$ ; паросочетание  $M$  пусто;  $Q(M) = n + m$ ;
2. перенести из  $G$  в  $M$  все рёбра  $(i, k)$ , для которых  $L[i, k] = 0$ ;
3.  $M_{\min} := M$ ;  $Q_{\min} := Q(M)$ ;
4. перебрать\_паросочетания  $(G, M)$ ;
5. вернуть наилучшее паросочетание  $M_{\min}$ ;

#### Функция перебрать\_паросочетания $(G, M)$ :

1. разбить граф  $G$  на связные компоненты;
2. для каждой связной компоненты  $G'$  из  $G$ :
  3. перебирать рёбра  $(i, k)$ , несмежные с  $M$ , по убыванию числа смежных рёбер;
  4. перенести ребро  $(i, k)$  из  $G'$  в  $M$ ;
  5. вычислить  $Q(M)$ ;
  6. если  $Q(M) < Q_{\min}$  то запомнить  $M_{\min} := M$ ;  $Q_{\min} := Q(M)$ ;
  7. перебрать\_паросочетания  $(G', M)$ ;
  8. вернуть ребро  $(i, k)$  из  $G'$  в  $M$ ;