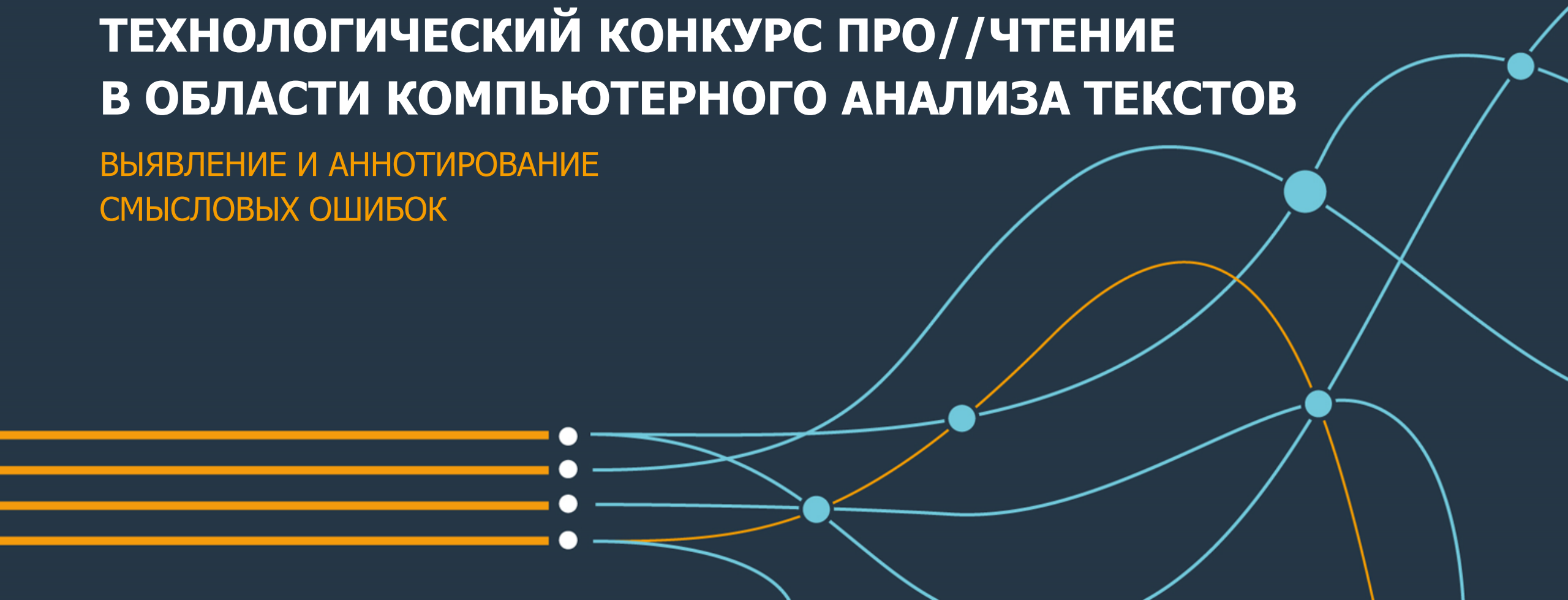




# ТЕХНОЛОГИЧЕСКИЙ КОНКУРС ПРО//ЧТЕНИЕ В ОБЛАСТИ КОМПЬЮТЕРНОГО АНАЛИЗА ТЕКСТОВ

ВЫЯВЛЕНИЕ И АННОТИРОВАНИЕ  
СМЫСЛОВЫХ ОШИБОК



# КОНКУРСЫ UP GREAT

ПРО//ЧТЕНИЕ

2

## Технологические конкурсы Up Great стартовали в 2018

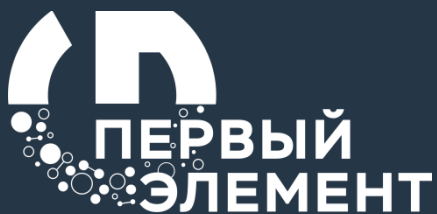
Команды инженеров и ученых пытаются преодолеть **технологические барьеры** в различных отраслях и найти решения **нерешенных нигде в мире** технологических задач.



Технологические конкурсы Up Great



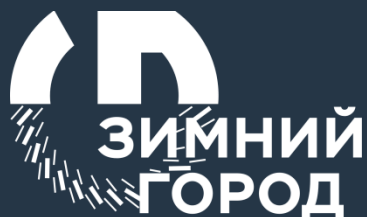
# ЗАВЕРШЕННЫЕ КОНКУРСЫ



## Первый элемент (2018-19)

Водородные топливные элементы для коптеров.  
Цель – создание чистой, тихой и невибрирующей альтернативы ДВС, обеспечивающей энергоплотность не ниже бензина.

Призовой фонд: 60 млн. руб.



## Зимний город (2018-19)

Масс-старт беспилотников по зимним городским дорогам: 10 автомобилей (5 финалистов-БПТС, 5 имитаторов городского движения)  
Сложные дорожные условия: заснеженная дорога, разметка, пешеходы, светофоры, нерегулируемые перекрестки, низкая температура.  
Цель – проехать 50 км за 3 часа без нарушений ПДД

Призовой фонд: 175 млн. руб.



## Ice Vision (2019)

Сателлит «Зимнего города»: молодежный конкурс по распознаванию дорожных знаков в условиях зимы на базе собранного к конкурсу крупнейшего открытого зимнего датасета.  
26 команд из 6 стран в финале, 300+ участников очной и заочной частей конкурса.

Призовой фонд: 3 млн. руб.



# ЦЕЛЬ КОНКУРСА

## СОЗДАНИЕ ТЕХНОЛОГИЙ ПОНИМАНИЯ СМЫСЛА ЕСТЕСТВЕННОГО ЯЗЫКА

Ключевая актуальная задача в области искусственного интеллекта – научить ИИ понимать смысл, рассуждать и самообучаться на малом объеме данных.

Когда ИИ научится понимать язык человека – начнется вторая волна ИИ-революции, которая позволит внедрить ИИ во все области человеческой деятельности, связанные со смысловой работой.

- ▶ **Three problems the community is working on:**
- ▶ **1. Learning with fewer labeled samples and/or fewer trials**
  - ▶ Self-supervised learning / unsup learning / learning to fill in the blanks
    - ▶ learning to represent the world before learning tasks
- ▶ **2. Learning to reason**, beyond “system 1” feed-forward computation.
  - ▶ Making reasoning compatible with gradient-based learning.
- ▶ **3. Learning to plan complex action sequences**
  - ▶ Learning hierarchical representations of action plans



**ЯН ЛЕКУН,**  
лауреат премии Тьюринга,  
VP Facebook, доклад 7.02.2020

# КОНКУРСНАЯ ЗАДАЧА

Создать **ИИ**, способный автоматически выявлять и аннотировать смысловые ошибки в текстах на естественных языках: русском и английском.

## ФАКТИЧЕСКАЯ ОШИБКА

автор высказывания А.Франц

В своем высказывании «Если человек зависит от природы, то и она от него зависит» Д. Мережковский **говорит** о необходимости защиты природы.

## ЛОГИЧЕСКАЯ ОШИБКА

тезис не обоснован

## ТЕХНОЛОГИЧЕСКИЙ БАРЬЕР:

создание стабильно работающего Программного комплекса для выявления фактических и смысловых ошибок в академических эссе, результат работы которого соответствует результату работы специалиста в условиях ограниченного времени.

ПРО//ЧТЕНИЕ

5





# КОНКУРСНАЯ ЗАДАЧА



|                             |                                                                                                                                                                                                                                                             |
|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Типы текстов                | Эссе и сочинения старшеклассников и студентов по неограниченному набору тем по предметам: русский язык, литература, обществознание, история, английский язык                                                                                                |
| Размер текста               | <12 000 символов                                                                                                                                                                                                                                            |
| Типы файлов                 | .txt файлы с вики-подобной разметкой                                                                                                                                                                                                                        |
| Ограничение по времени      | Не более 30 секунд на один файл                                                                                                                                                                                                                             |
| Аннотации                   | Ошибки классифицируются и аннотируются на естественном языке                                                                                                                                                                                                |
| Датасет                     | Десять тысяч проверенных, аннотированных и верифицированных эссе                                                                                                                                                                                            |
| Категории выявляемых ошибок | Десятки или сотни тысяч эссе, собранных в ходе краудсорса <ul style="list-style-type: none"><li>- грамматические, речевые (владение языком и соблюдение языковых норм)</li><li>- логические, фактические, этические (использование знаний о мире)</li></ul> |



# СТРУКТУРА ИСПЫТАНИЙ



## 1 КВАЛИФИКАЦИЯ

- Несколько десятков ранее не публиковавшихся эссе
- Испытания в течение недели
- Без временных ограничений
- Автоматический подсчет результатов
- Критерий прохождения: корректное выделение хотя бы одной ошибки

## 2 ОСНОВНЫЕ ИСПЫТАНИЯ

- 1000 неопубликованных, неразмеченных эссе
- Ограниченное время: каждую минуту публикуется новое эссе, среднее время на проверку не превосходит 30 секунд

## 3 ПРОВЕРКА

- После окончания испытаний эссе передаются на проверку экспертам
- Время на проверку одним экспертом: 10 минут
- Ручная проверка текстовых аннотаций, созданных ИИ

## 4 РЕЗУЛЬТАТЫ

- Автоматический подсчет результатов
- Аудит решений лучших 10 команд
- В случае, если команда показывает результат на уровне экспертов - победа



# СТРУКТУРА ИСПЫТАНИЙ



Решение ИИ сравнивается с решениями экспертов. Далее решения экспертов сравниваются друг с другом. Вычисляется отношение точности решения в первом случае и во втором. Если оно больше 100%, то технологический барьер пройден.



Испытания проводятся с периодичностью раз в полгода до завершения конкурса.



В остальное время открыта площадка для тестирования решений на закрытой тестовой выборке.



**ПРИЗОВОЙ  
ФОНД**

**РУССКИЙ  
100 МЛН. РУБ.**

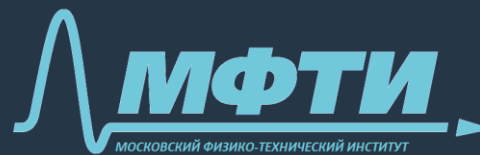
**АНГЛИЙСКИЙ  
100 МЛН. РУБ.**





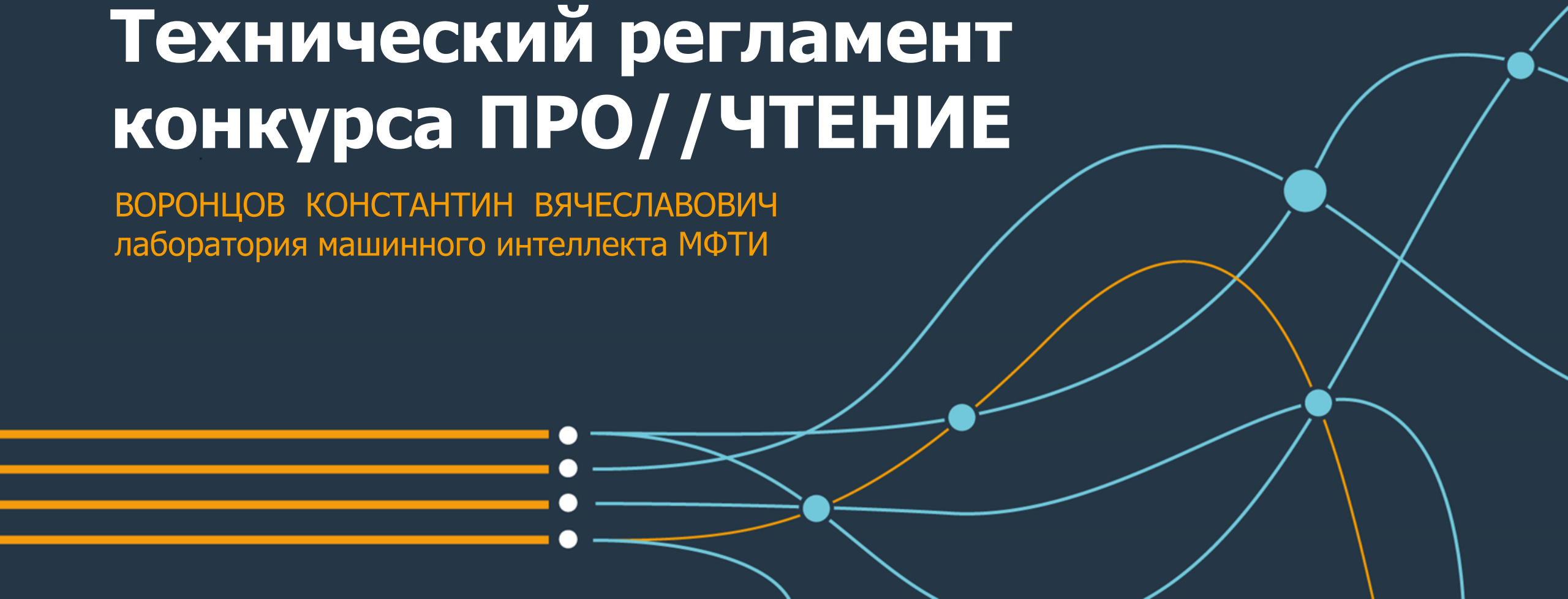
# ГРАФИК КОНКУРСА





# Технический регламент конкурса ПРО//ЧТЕНИЕ

ВОРОНЦОВ КОНСТАНТИН ВЯЧЕСЛАВОВИЧ  
лаборатория машинного интеллекта МФТИ





## КОНКУРСНАЯ ЗАДАЧА – СОЗДАТЬ ИИ ДЛЯ ПОИСКА СМЫСЛОВЫХ ОШИБОК В ТЕКСТЕ

## ДАНО

- Как устроена обучающая выборка
- Кто пишет эссе и кто их проверяет
- Как собрать большие данные

## НАЙТИ

- Какими бывают смысловые ошибки
- Как устроен язык разметки
- Как будет обучаться искусственный интеллект

## КРИТЕРИЙ

- Как сравнить разметку с эталонной
- Как будут проводиться испытания
- Каков критерий преодоления технологического барьера

# ДАННЫЕ

ПРО//ЧТЕНИЕ

12

## Неразмеченные тексты

Сочинения старшеклассников по неограниченному набору тем.  
ЕГЭ по предметам: русский язык, литература, обществознание, история, английский язык, ... (список расширяемый).

## Размеченные тексты

Сочинения, проверенные преподавателями.  
ЕГЭ: проверку делают два независимых эксперта;  
в случае существенных расхождений назначается третий.

## Обучающая выборка

Размеченные тексты, используемые для обучения моделей.

## Тестовая выборка

Размеченные тексты, разметка которых не сообщается участникам.  
Используются для сравнения моделей в квалификационных испытаниях.

## Финальная выборка

Неразмеченные тексты, используемые для финальных испытаний.  
Эксперты размечают их после получения решений от участников.

# ПРИНЦИПЫ РАЗМЕТКИ ДАННЫХ



- Язык разметки должен быть удобен как для человека, так и для машины



- Разметка выделяет фрагменты двух видов: ошибка и смысловой блок
  - **ошибка имеет:** *тип, подтип, комментарий, исправление, пояснение, тег*
  - **смысловой блок имеет:** *тип, тег*



- Оценка за эссе однозначно вычисляется по выделенным фрагментам
  - **наличие ошибки снижает оценку**
  - **наличие смыслового блока повышает оценку**



- Фрагменты могут быть вложенными, например, внутри смыслового блока могут выделяться ошибки



# ЯЗЫК РАЗМЕТКИ ДАННЫХ

ПРО//ЧТЕНИЕ

14

(**\*** **Коды** \ фрагмент \ **Комментарий** :: **Пояснение** >> **Исправление** # **Тег** **\***)

**Коды** – один или несколько идентификаторов *типа смыслового блока, типа ошибки или подтипа ошибки*

**Фрагмент** – неизменённый фрагмент исходного текста эссе, локализирующий ошибку или смысловой блок

**Комментарий** – объяснение ошибки непредусмотренного подтипа, без обращения к контексту сочинения

**Пояснение** – развёрнутый комментарий для учащегося, объясняющий суть ошибки в контексте сочинения

**Исправление** – предлагаемый экспертом правильный вариант фрагмента

**Тег** – строка из букв или цифр для связывания нескольких фрагментов, относящихся к одной и той же ошибке



# ПРИМЕРЫ РАЗМЕТКИ

ПРО//ЧТЕНИЕ

15

- Грамматическая ошибка из сочинения по русскому языку:

Все удивлялись его (\* Г.упр \ силой >> силе \*).

- Ошибка из мини-сочинения по обществознанию: понятие определено, но не связано с основной идеей

(\* ПОНЯТИЕ О.теорсвязь идея \ Деятельность – это процесс целенаправленной активности людей.  
:: Для раскрытия смысла высказывания следовало бы рассмотреть особенности процесса познания. Более общее понятие деятельности уводит в сторону от основной идеи высказывания. \*)

- Из исторического сочинения: связанные смысловые блоки «причина-следствие» и ошибка в понятии

(\* 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 \*)

# ПРИМЕР РАЗМЕТКИ (ПОЛНЫЙ ТЕКСТ)

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

**Предмет:** обществознание

**К1: 1 К2: 2 К3: 1 К4: 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, 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.\*)

# МЕТРИКИ СРАВНЕНИЯ ДВУХ РАЗМЕТОК

ПРО//ЧТЕНИЕ

17

## Сравнение алгоритмической разметки участника с экспертной разметкой того же эссе

- Вычисление оценки за эссе в баллах согласно правилам ЕГЭ для обеих разметок
- Поиск оптимального соответствия между фрагментами двух разметок одного и того же эссе
- Метрики *парной точности* по двум разметкам одного эссе (0%--100%):

M1 = точность предсказания оценки за эссе

M2 = точность и полнота поиска фрагментов

M3 = точность предсказания ошибок и смысловых блоков

M4 = точность предсказания комментариев

M5 = точность локализации фрагментов

M6 = точность исправлений ошибок

M7 = точность пояснений (экспертная оценка)

- Итоговая метрика *парной точности* (0%--100%) для заданной пары разметок одного и того же эссе:

$$M = (M1 + \dots + M7) / 7$$

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 действия некоего) 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.

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. (\*РОЛЬ Также можно выделить самого Александра III, (и действие некоего) 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 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.

# КРИТЕРИИ ТОЧНОСТИ РАЗМЕТКИ

Точность разметки оценивается *по выборке* размеченных эссе

- СТАР = Средняя Точность Алгоритмической Разметки  
при сравнении алгоритмических разметок с экспертными
- СТЭР = Средняя Точность Экспертной Разметки  
при сравнении экспертных разметок друг с другом
- Относительная Точность Алгоритмической Разметки:  
$$\text{ОТАР} = \text{СТАР} / \text{СТЭР} * 100\%$$
- **ОТАР  $\geq 100\%$  -- условие преодоления Технологического барьера**



# ЭТАПЫ КОНКУРСА НА КАЖДОМ ЦИКЛЕ

ПРО//ЧТЕНИЕ

19



## ОТБОРОЧНЫЙ ЭТАП

– рассмотрение заявок Участников Оргкомитетом Конкурса



## КВАЛИФИКАЦИОННЫЙ ЭТАП

– многократное тестирование алгоритмов Участников в двух режимах:

- *алгоритмическая квалификация*
  - оценивание точности алгоритмов на Тестовой выборке
- *техническая квалификация*
  - тренировка в режиме реального времени, оценивание надёжности



## ЭТАП ФИНАЛЬНЫХ ИСПЫТАНИЙ

– однократное тестирование алгоритмов в режиме реального времени

# ЭТАП ФИНАЛЬНЫХ ИСПЫТАНИЙ

ПРО//ЧТЕНИЕ

20

## Сценарий обмена данными на этапе финальных испытаний и технической квалификации

- У:** Запрос на открытие сессии (должен поступить *не позднее 10 секунд* после старта).
- С:** Проверка Участника и, в случае успеха, передача разрешения на открытие сессии.
- У:** Запрос на получение следующего файла неразмеченного эссе (должен поступить *не позднее 20 секунд* после открытия доступа к данному файлу).
- С:** Передача файла неразмеченного эссе или сообщения о завершении выборки.
- У:** Отправка размеченного файла того же эссе (должна произойти *не позднее 30 секунд* после отправки неразмеченного файла).
- С:** Замер времени отклика, сохранение разметки, оценивание качества разметки.
- С:** По завершении выборки генерация и сохранение аналитического отчёта.

---

\*У – программный комплекс участника

\*С – система проверки решений конкурсной платформы



# НЕСКОЛЬКО НЕОЖИДАННЫХ ИНСАЙТОВ ;)

ПРО//ЧТЕНИЕ

21

- Чтобы оценить сочинение, в нём надо выделять не только ошибки (минусы), но и смысловые блоки (плюсы).
- Разметка структурирует мышление у проверяющих и уменьшает число разногласий между ними благодаря необходимости выделять границы фрагментов и указывать тип ошибки.
- Даже очень простой (\* язык \ разметки \ текста :: может стать >> кошмаром \*) для гуманитария. Вначале разметчики испытывают трудности с закрывающими скобками, вложенными фрагментами и тегами.
- Оценка за сочинение однозначно вычисляется по разметке для всех предметов ЕГЭ... кроме английского языка.





# Юрий Молодых

Директор по развитию Up Great



[Molodykh.YO@rvc.ru](mailto:Molodykh.YO@rvc.ru)



[www.ai.upgreat.one](http://www.ai.upgreat.one)  
[ai@upgreat.one](mailto:ai@upgreat.one)

