

# Анализ позиции в отношении заданной темы с использованием дискурсивного анализа текста

Леонид Зайцев

Проектные задачи КЛ, 24.02.2025

## Предыдущие исследования:

### Использование дискурсивного анализа

- Анализ тональности на основе классификации каждого узла в дереве РС:
  - Bhatia Parminder et al. Better Document-level Sentiment Analysis from RST Discourse Parsing
  - Fu Xianghua et al. Long short-term memory network over rhetorical structure theory for sentence level sentiment analysis
  - Huber Patrick and Carenini Giuseppe. From Sentiment Annotations to Sentiment Prediction through Discourse Augmentation
  - Kraus Mathias and Feuerriegel Stefan. Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees

# Вдохновлено:

Анализом аргументации при помощи  
дискурсивного анализа (TPC) в

Чистова, Смирнов 2022. Discourse-aware text  
classification for argument mining

# Чистова, Смирнов 2022

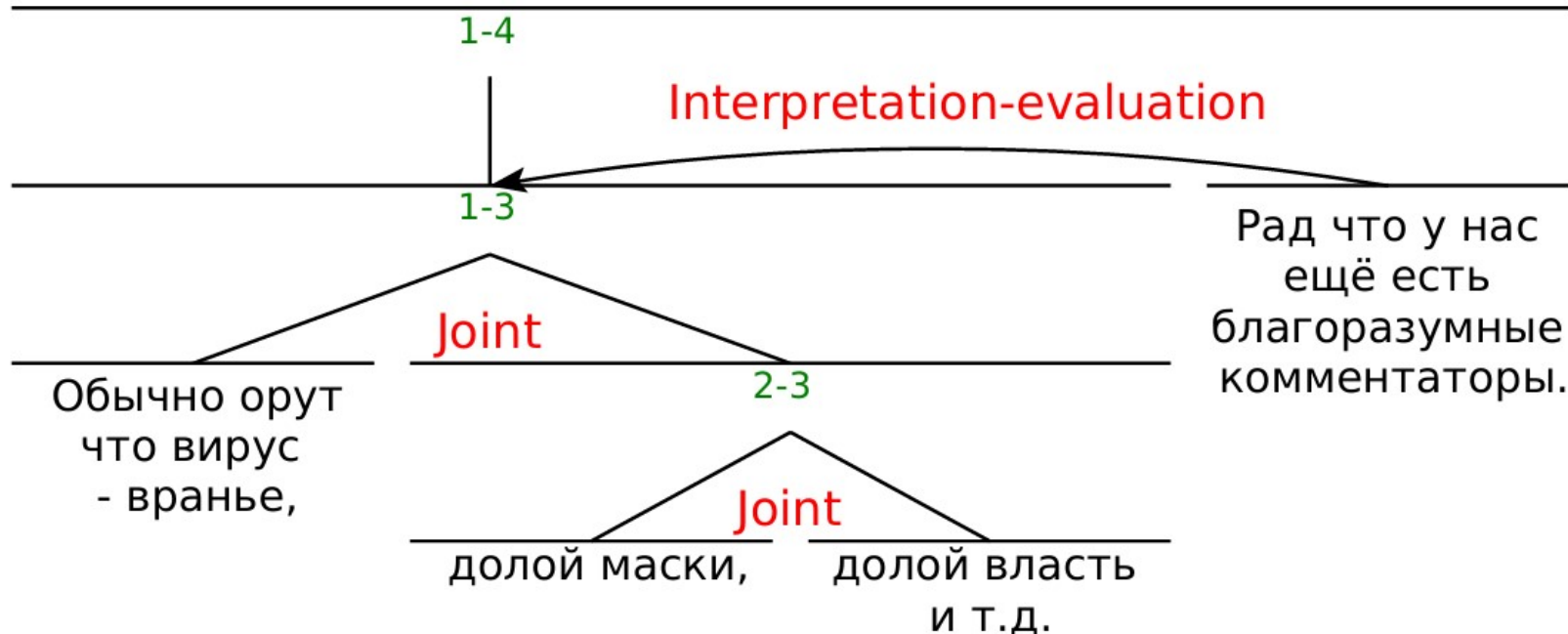
- Предварительная тонкая настройка ruBERT.
- Точное определение того, как именно риторическая структура будет использована:

# Чистова, Смирнов 2022

- Точное определение того, как именно риторическая структура будет использована:
  - «листы» дерева РС классифицируются замороженной моделью заранее
  - TreeLSTM дообучается определять по их меткам и РО класс «корня» — позицию автора текста

# Чистова, Смирнов 2022

- «листы» дерева РС классифицируются замороженной моделью заранее
- TreeLSTM дообучается определять по их меткам и РО класс «корня» — позицию автора текста



Что, если LLM, выполняющая инструкции, просто получит информацию о риторической структуре текста (с пояснением)?

# Корпус: RuArg-2022

## Составление

- Взяты три утверждения:
  - «Вакцинация полезна для общества».
  - «Введение и соблюдение карантина полезно для общества».
  - «Ношение масок полезно для общества».



# Анализ данных RuArg-2022

- «Из социальных сетей была собрана коллекция предложений – комментариев к постам из социальных медиа. В этих предложениях могут содержаться как высказывания, определяющие авторскую позицию относительно заданных утверждений, так и высказывания с доводами «за»/«против» этих утверждений.»

(<https://github.com/dialogue-evaluation/RuArg>)

# Анализ данных RuArg-2022

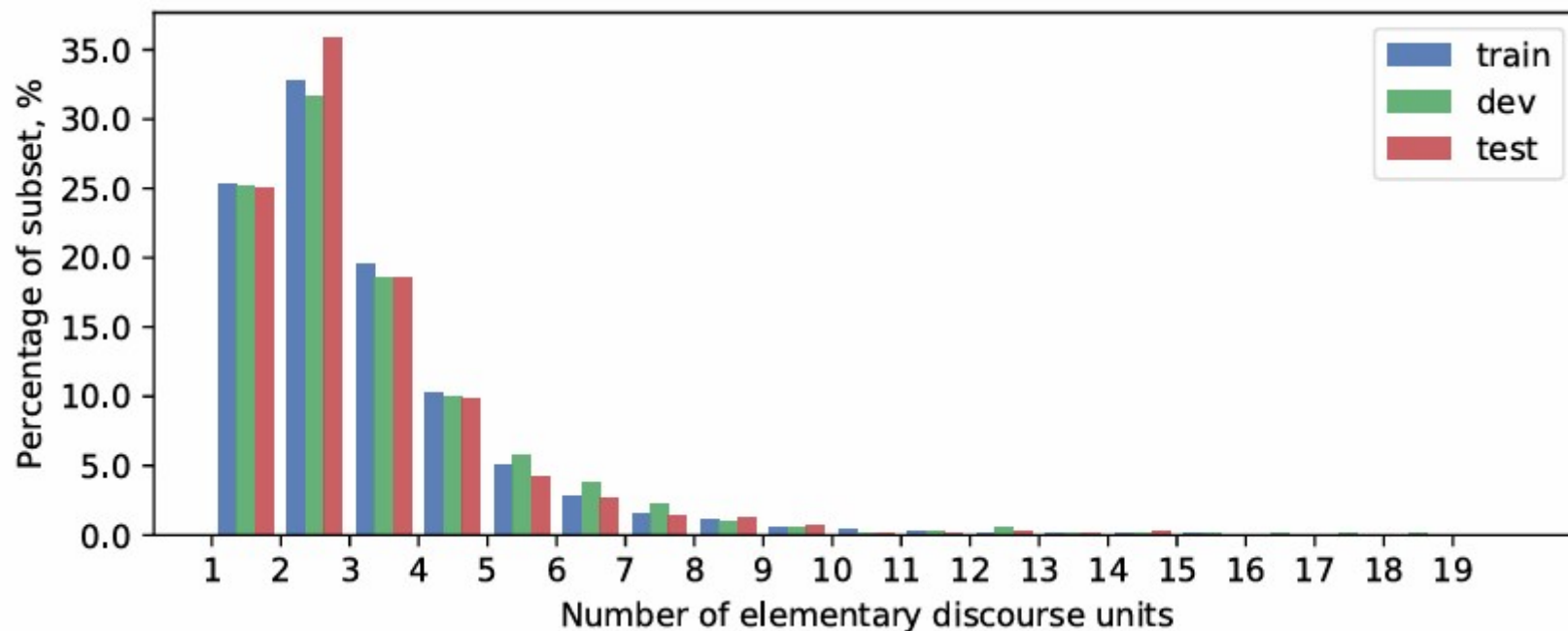
- Каждое предложение размечено по позиции и по доводу для каждого из трёх утверждений, то есть имеет шесть меток.
- Значения меток:
  - «за» (2)
  - «против» (0)
  - «прочее» (1)
  - «нерелевантно» (-1) (для данного утверждения)

# Анализ данных RuArg-2022

- Разметка произведена вручную, по детальной инструкции.
- Предложение рассматривалось как довод, если разметчик мог использовать его, чтобы убедить оппонента в правильности некоторого утверждения.

# Использование дискурсивного анализа

- Предпосылки: только четверть выборки состоит из «простых» предложений (из одной ЭДЕ)



# Корпус: RuArg-2022

- В данном исследовании были взяты предложения из тренировочной выборки RuArg-2022 по теме «маски» с разметкой только по позиции автора.

# Разметка РС

- Разметка проводилась вручную, с помощью редактора риторических структур rstWeb (<https://github.com/amir-zeldes/rstWeb>).

# Выборка

- Для ручной разметки были взяты первые сто примеров из тренировочной выборки RuArg-2022, из которых были отсеяны примеры из одной ЭДЕ и примеры настолько нечитаемые, что риторическую структуру для них установить невозможно.

=> 78 примеров.

# Выборка

- Баланс классов:
  - положительный: 29
  - нейтральный: 22
  - отрицательный: 27



# Модель

- LLM:

Mistral AI, которой подавались  
инструкция и сами предложения

# Промпт

«Для текста, приведённого после слова ТЕКСТ, оцени, как его автор относится к медицинским маскам, по шкале:

- плохо относится — 0,
- нейтрально относится, позиция неясна или противоречива — 1,
- хорошо относится — 2.

В качестве ответа верни только число.

ТЕКСТ:...»

# Промпт

Для аннотированных примеров добавлялась инструкция о дереве РС:  
«...После слова АННОТАЦИЯ приводится разметка риторических отношений между частями текста. Риторические отношения выражают связь содержания этих частей с замыслом автора и между собой...»

Добавлялся соответствующий пункт входных данных.

# Формат аннотации

- Файл XML, кодирующий все отношения в графе, читая который, rstWeb рисует дерево.
- Два вариант ввода: весь файл (с инвентарём РС) или только сам граф

# Формат аннотации

```
<rst>
  <header>
    <relations>
      <rel name="antithesis" type="rst"/>
      <rel name="background" type="rst"/>
      <rel name="cause" type="rst"/>
      <rel name="circumstance" type="rst"/>
      <rel name="concession" type="rst"/>
      <rel name="condition" type="rst"/>
      <rel name="contrast" type="multinuc"/>
      <rel name="elaboration" type="rst"/>
      <rel name="evaluation" type="rst"/>
      <rel name="evidence" type="rst"/>
      <rel name="joint" type="multinuc"/>
      <rel name="justify" type="rst"/>
      <rel name="motivation" type="rst"/>
      <rel name="preparation" type="rst"/>
      <rel name="purpose" type="rst"/>
      <rel name="restatement" type="multinuc"/>
      <rel name="restatement" type="rst"/>
      <rel name="result" type="rst"/>
      <rel name="sequence" type="multinuc"/>
      <rel name="solutionhood" type="rst"/>
    </relations>
  </header>
  <body>
    <segment id="1" parent="11" relname="joint">[USER], Потому что те, кто сейчас занёс ковид19 в список таких болезней</segment>
    <segment id="2" parent="10" relname="joint">как чума,</segment>
    <segment id="3" parent="10" relname="joint">сибирская язва и т. п.,</segment>
    <segment id="4" parent="11" relname="joint">объявил масочный режим,</segment>
    <segment id="5" parent="11" relname="joint">карантин,</segment>
    <segment id="6" parent="11" relname="joint">смерть малого и среднего бизнеса,</segment>
    <segment id="7" parent="11" relname="joint">принуждает всех вакцинироваться,</segment>
    <segment id="8" parent="12" relname="span">10-ю годами ранее очень "эффективно" оптимизировал медицину,</segment>
    <segment id="9" parent="8" relname="background">чуть не угробив вообще инфекционные мед. учреждения на западный манер.</segment>
    <group id="10" type="multinuc" parent="1" relname="elaboration"/>
    <group id="11" type="multinuc" parent="8" relname="elaboration"/>
    <group id="12" type="span" />
  </body>
</rst>
```

# Результаты

- Точность (accuracy) и F-мера для разметки с использованием аннотации РС повышаются на 2-3% относительно базовой.
- (И примерно настолько же понижаются, если предъявлять весь файл.)

# Результаты

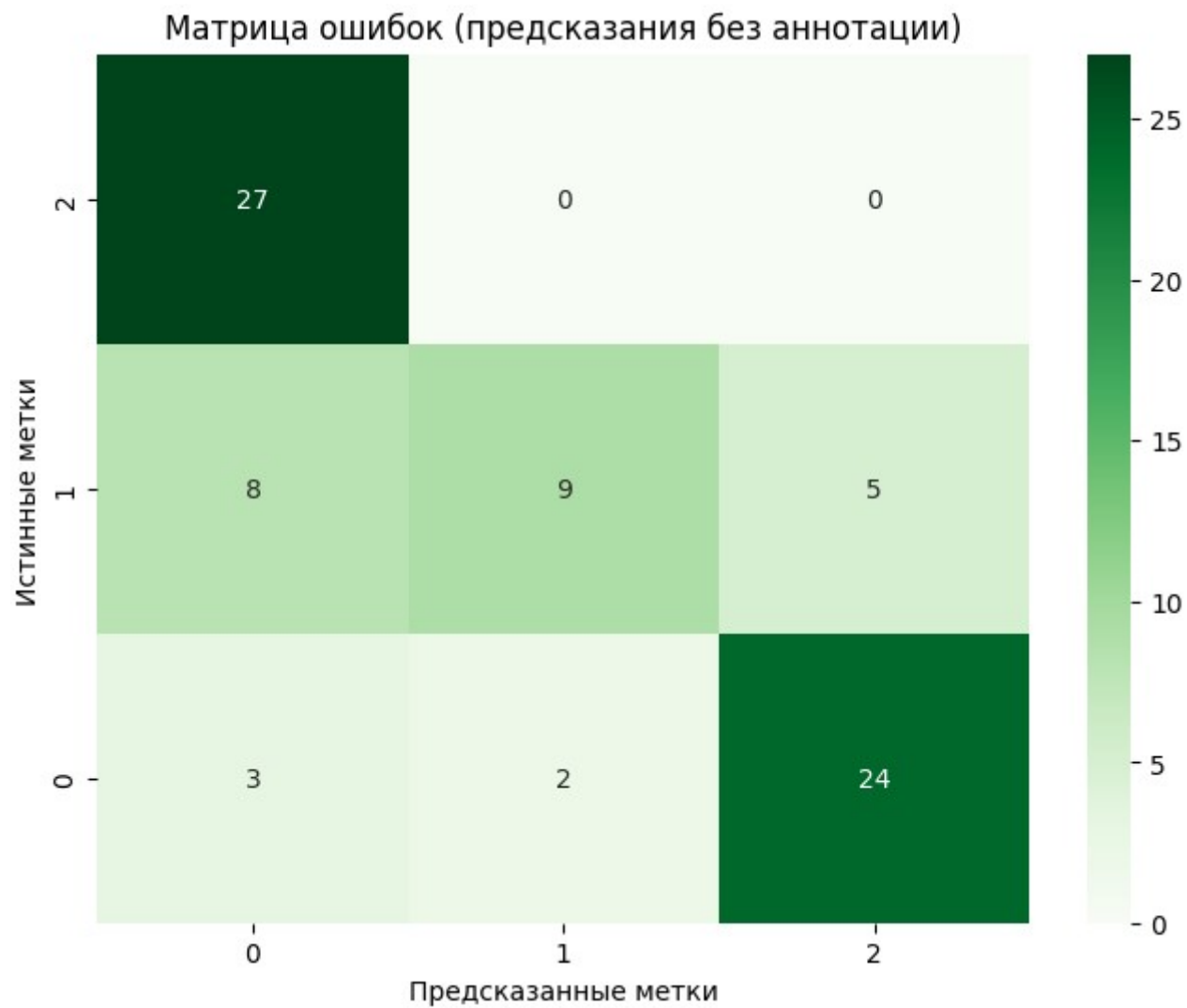
Метрика	Стимул без аннотации	Стимул с аннотацией РС
Accuracy	76.92	78.21
F1-score	73.46	76.53

# Результаты

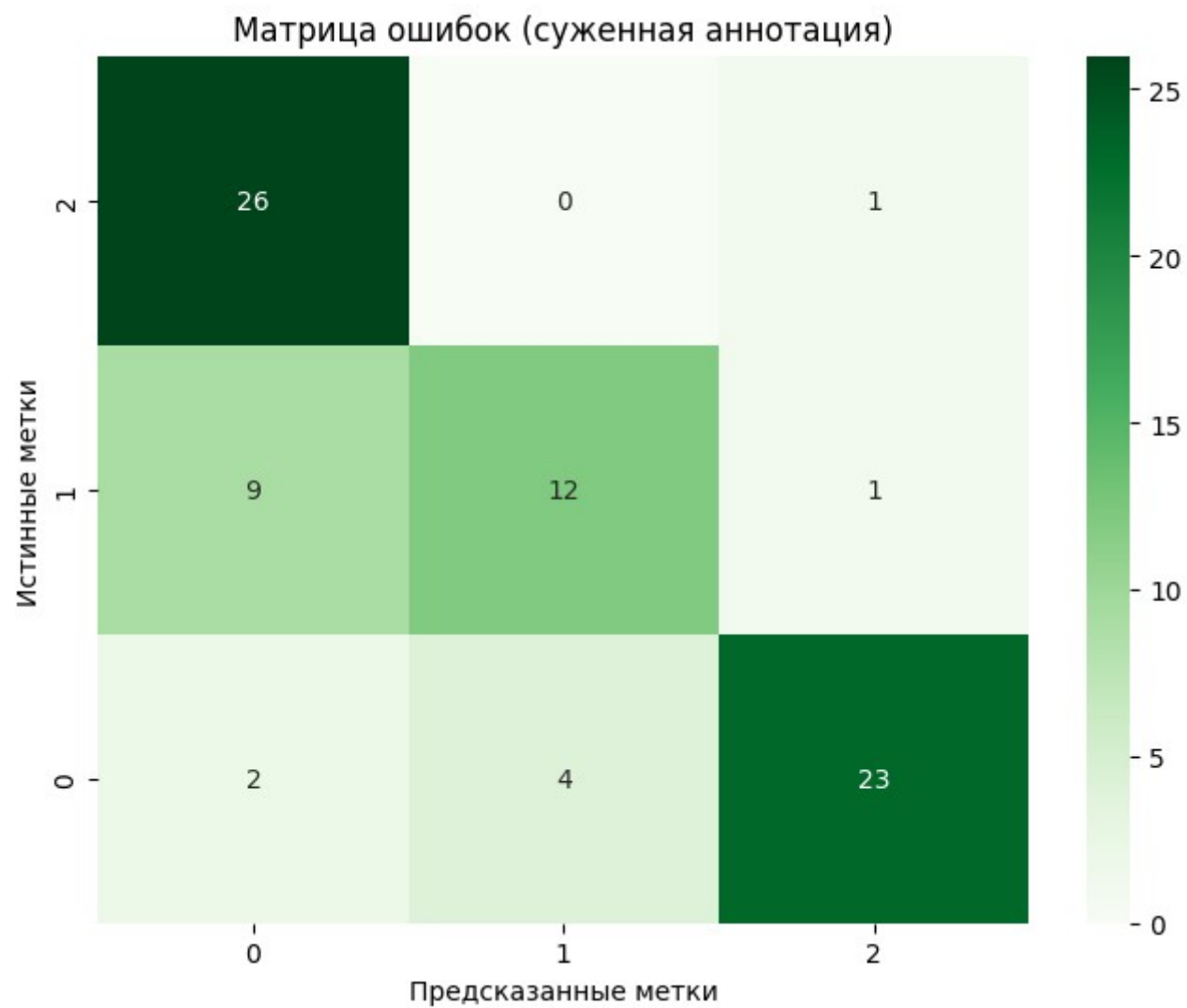
- С информацией о РС модель верно определила большее число нейтральных примеров
- Были исправлены метки и для других классов



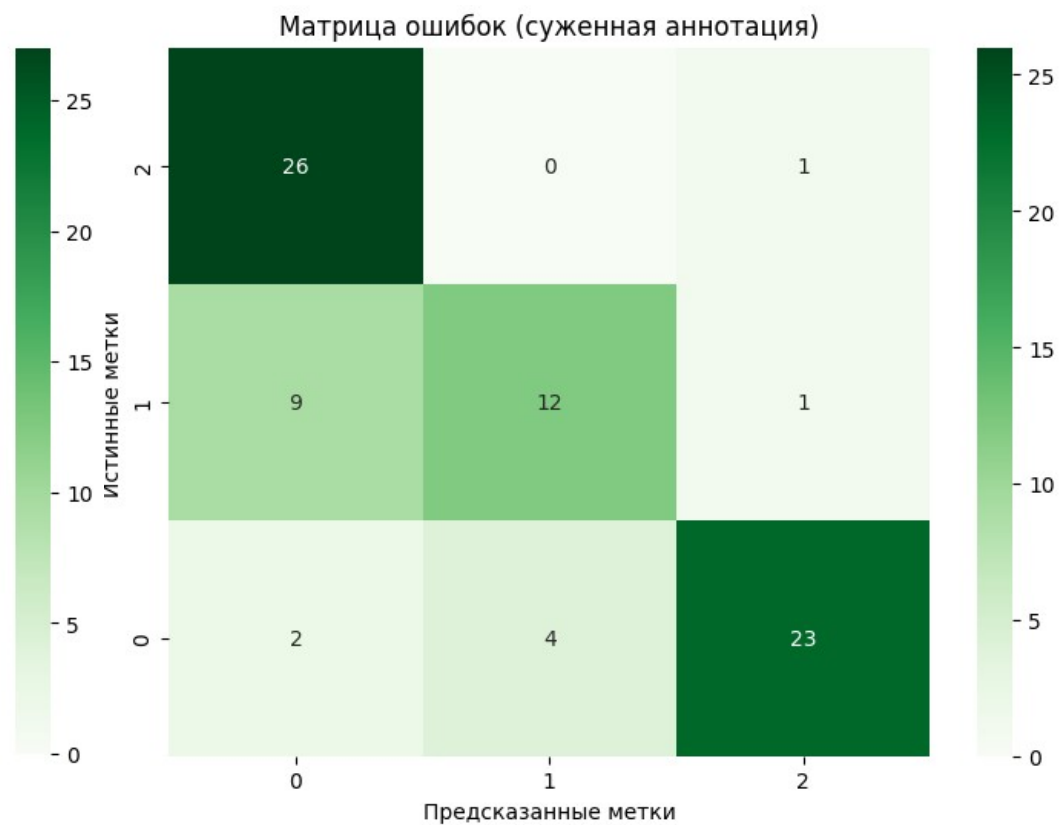
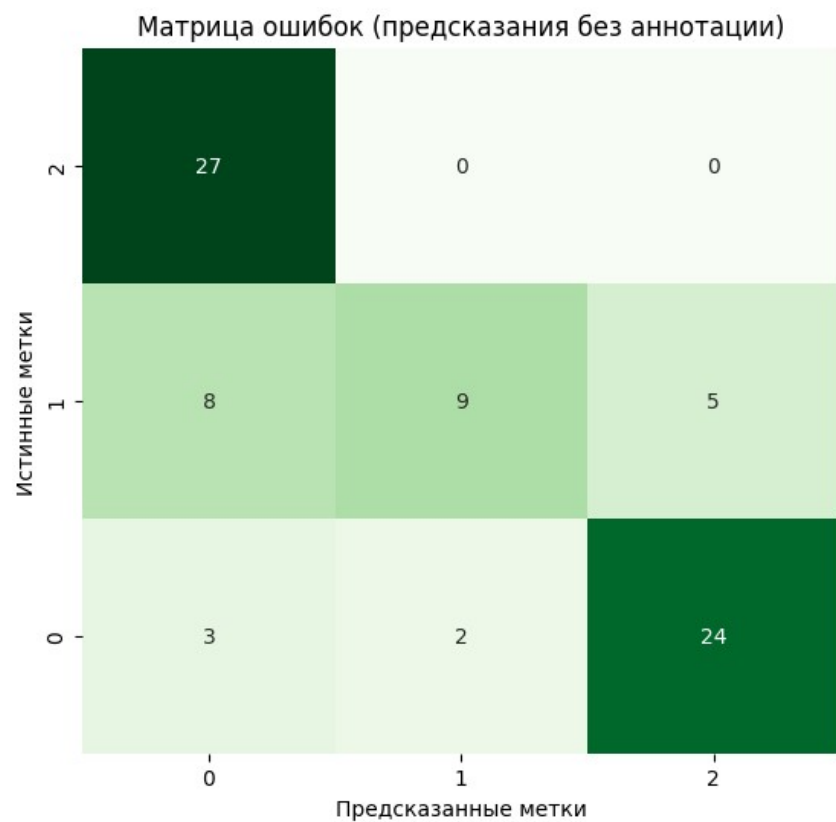
# Результаты



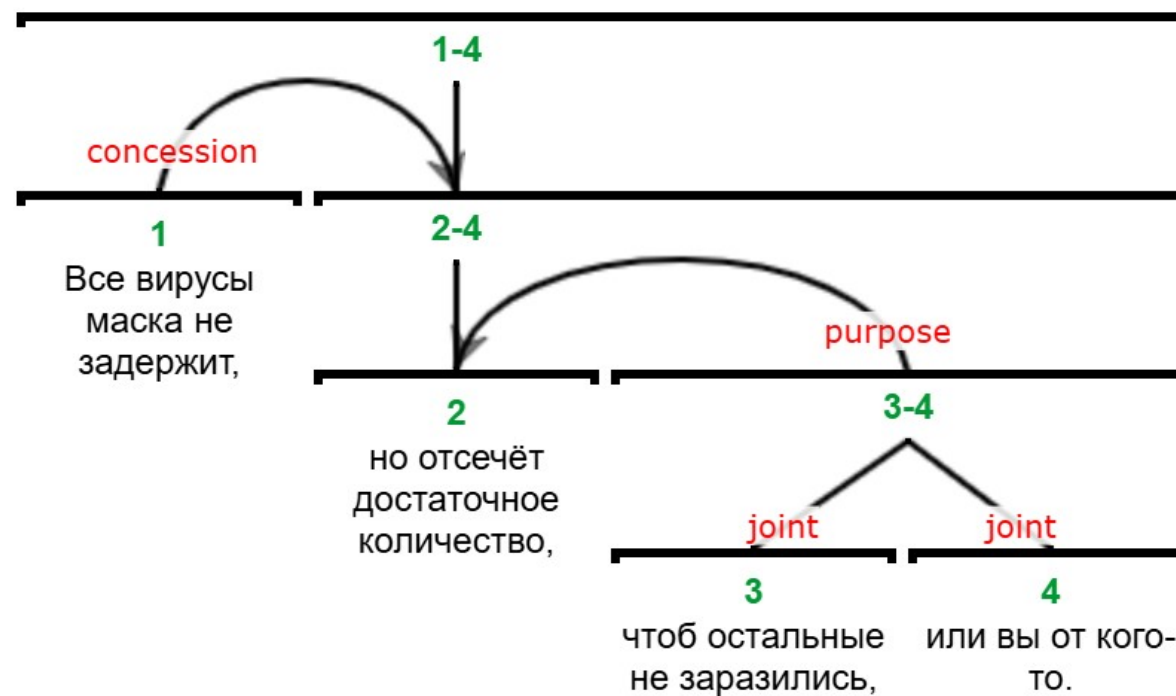
# Результаты



# Результаты



# Результаты



- Модель без аннотации: отрицательный класс
- С аннотацией: положительный — верно!

# Перспективы

- Проверка результатов на большой выборке, для которой требуется рабочий и качественный автоматический парсер
- Исследование возможностей LLM с использованием структуры дискурса в задаче собственно извлечения доводов

# Ссылки

- Elena Chistova and Ivan Smirnov. 2022. Discourse-aware text classification for argument mining // Computational Linguistics and Intellectual Technologies: papers from the Annual conference “Dialogue”.
- Fu Xianghua et al. Long short-term memory network over rhetorical structure theory for sentence level sentiment analysis // Asian conference on machine learning / PMLR. — 2016. — P. 17–32.
- Kraus Mathias and Feuerriegel Stefan. Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees // Expert Systems with Applications. — 2019. — Vol. 118.—P. 65–79.
- Huber Patrick and Carenini Giuseppe. From Sentiment Annotations to Sentiment Prediction through Discourse Augmentation // Proceedings of the 28th International Conference on Computational Linguistics. — 2020. — P. 185–197.
- Bhatia Parminder et al. Better Document-level Sentiment Analysis from RST Discourse Parsing // Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. — 2015. —P. 2212–2218.
- rstWeb — <https://github.com/amir-zeldes/rstWeb>
- Рабочий блокнот google.colab:  
[https://colab.research.google.com/drive/15MDWaEQBU3rsqbCPvNfJsnJoCGo39\\_n\\_a?usp=sharing](https://colab.research.google.com/drive/15MDWaEQBU3rsqbCPvNfJsnJoCGo39_n_a?usp=sharing)