



Всероссийская научно-практическая конференция с международным участием
“Язык, текст, культура через призму цифровых технологий”

Стандартизация разметки данных и оценивание языковых моделей для формализации гуманитарных знаний

Воронцов

Константин Вячеславович

д.ф.-м.н., профессор РАН,
Институт ИИ МГУ им. М.В. Ломоносова, Москва, Россия,
МФТИ, Москва, Россия
k.vorontsov@iai.msu.ru



Карабулатова

Ирина Советовна

д.филол.н., профессор,
Хэйлунцзянский университет, Харбин, КНР,
Институт ИИ МГУ им. М.В. Ломоносова,
radogost2000@mail.ru



Контент-анализ: обобщение и автоматизация

Обобщённый контент-анализ — четыре базовые операции с текстом:

- 1) выделить фрагмент
- 2) классифицировать (тегировать) фрагмент по рубриктору
- 3) *связать несколько фрагментов*
- 4) *дать комментарий (затекст) к фрагменту или связи*

Цель — автоматизировать контент-анализ больших текстовых массивов по небольшим размеченным корпусам, в любой предметной области

Три задачи построения обучаемой модели разметки:

- 1) разработка рубриктора и инструкций разметчика
- 2) дообучение большой языковой модели по разметке
- 3) оценивание качества, сравнение и выбор моделей

Примеры задач разметки текстов в NLP/ML

- распознавание наименований (named entity recognition, NER)
- распознавание частей речи (part of speech tagging, POS)
- выделение тональности объектов (sentiment analysis, SA)
- выделение синтаксических связей (syntax parsing)
- выделение семантических ролей (semantic role labeling, SRL)
- выделение текстовых полей данных (slot filling)
- выделение полей в библиографических записях
- сегментация научных или юридических текстов
- разрешение анафоры, кореферентности, эллипсиса

Пример: разметка наименований (NER)

Óним (имя собственное) служит для выделения именованного объекта среди других объектов, его индивидуализации и идентификации

Named entity — объект (сущность) реального мира, имеющий наименование и относящийся к определённой категории.

Примеры категорий:

- персона, организация, локация, время
- историческое событие, артефакт, документ
- товар, изделие, предмет искусства
- заболевание, симптом, препарат
- астрономический объект, телескоп

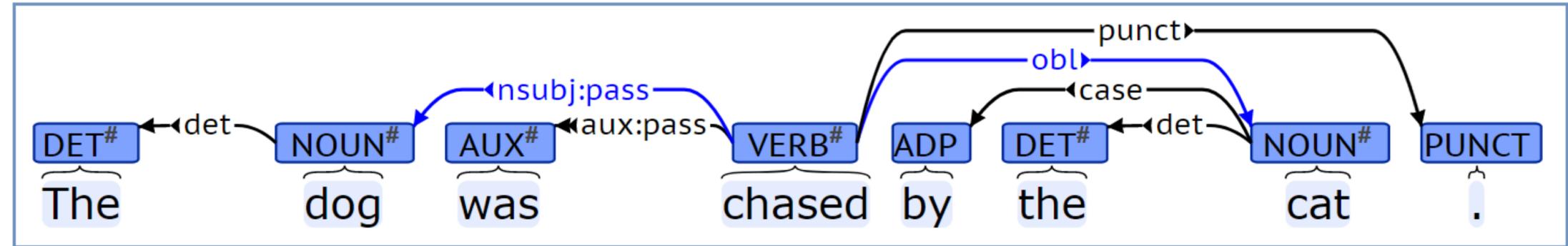
Legend: Person (p), Organization (o), Other (z), Location (l), Date (d)

Shinzō Abe is a Japanese politician serving as the 63rd and current Prime Minister of Japan and Leader of the Liberal Democratic Party (LDP) since 2012, previously being the 57th officeholder from 2006 to 2007. He is the third-longest serving Prime Minister in post-war Japan.[1]

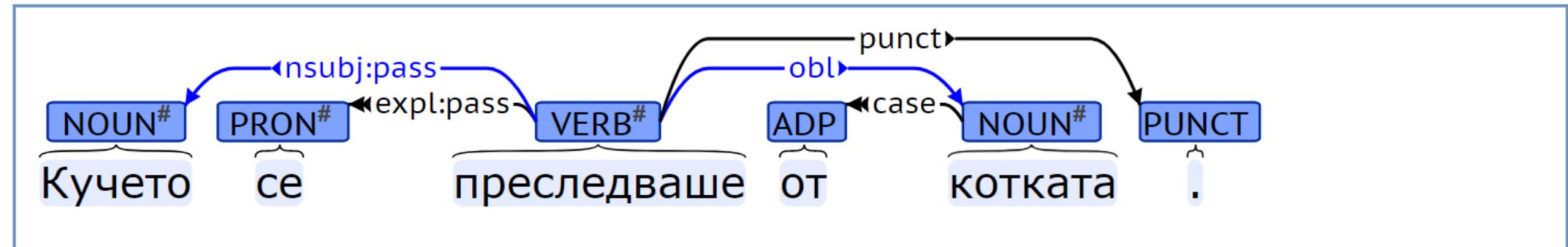
Abe comes from a politically prominent family and was first elected Prime Minister by a special session of the National Diet in September 2006. Then aged 52, he became Japan's youngest post-war Prime Minister and the first to have been born after World War II. Abe resigned on 12 September 2007 for health reasons. He was replaced by Yasuo Fukuda, the first in a

Пример: частеречная и синтаксическая разметка

английский:



болгарский:



теги
частей
речи

NOUN	noun	существительное	INTJ	interjection	междометие
PROPN	proper noun	имя собственное	ADP	adposition	предлог
ADJ	adjective	прилагательное	CONJ	conjunction	союз
VERB	verb	глагол	PART	particle	частица
ADV	adverb	наречие	PUNCT	punctuation	знак пунктуации
PRON	pronoun	местоимение	SYM	symbol	символ
NUM	numeral	числительное	X	other	иное

Пример: разметка семантических ролей (SRL)

Задача: найти в предложении *актанты* — именные группы, обозначающие участников ситуации и их *семантические роли*.

Примеры семантических ролей:

- **агенс:** одушевлённый инициатор и контролёр действия
- **пациенс:** участник, на которого направлено действие
- **бенефактив:** участник, получающий пользу или вред
- **адресат:** получатель сообщения (может быть бенефактивом)
- **инструмент:** посредством чего осуществляется действие
- **экспериенцер:** носитель чувств и восприятий
- **стимул:** источник восприятий
- **источник:** исходный пункт движения
- **цель:** конечный пункт движения

Конкурсы SemEval по детекции пропаганды

Базовая разметка: «фрагмент, метка класса»



Gallia est omnis divisa in partes tres, quarum unam incolunt Belgae, aliam Aquitani, tertiam qui ipsorum lingua Celtae, nostra Galli appellantur. Hi omnes lingua, institutis, legibus inter se differunt. Gallos ab Aquitanis Garumna flumen, a Belgis Matrona et Sequana dividit. Horum **omnium fortissimi** sunt Belgae, propterea quod a cultu atque humanitate provinciae longissime absunt, minimeque ad eos mercatores saepe commeant atque ea **quae ad effeminandos animos pertinent important**, proximique sunt Germanis, qui trans Rhenum incolunt, quibuscum continenter bellum gerunt. Qua de causa **Helvetii quoque reliquos Gallos virtute praecedunt**, **quod fere cotidianis proeliis cum Germanis contendunt**, cum aut suis finibus eos prohibent aut ipsi in eorum finibus bellum gerunt. Eorum una pars, quam Gallos obtinere dictum est, initium capit a flumine Rhodano, continetur Garumna flumine, Oceano, finibus Belgarum, attingit etiam ab Sequanis et Helvetiis flumen Rhenum, vergit ad septentriones. Belgae ab extremis Galliae finibus oriuntur, pertinent

Manipulative Wording: Loaded Language

Attack on Reputation: Smears

Manipulative Wording: Exaggeration

Justification: Appeal to Values



Commissio
PopulusQue
Europaea

Упрощённая разметка: «предложение, метка класса»

Обобщённая разметка: «фрагмент→мишень, метка класса»

- SemEval-2023 task 3. Detecting the genre, the framing, and the persuasion techniques in online news in a multi-lingual setup. <https://propaganda.math.unipd.it/semEval2023task3>
- G.Martino, P.Nakov et al. A survey on computational propaganda detection. 2020.
- F.Alam, P.Nakov et al. Overview of the WANLP 2022 shared task on propaganda detection in Arabic. 2022.

Соревнование RuTermEval-2024



Поиск и классификация терминов в русскоязычных научных статьях:

- specific term – термины, специфичные доменно и лексически
- common term – общеизвестные термины, специфичные только доменно
- nomen – номенклатурные наименования доменно специфичных объектов

Метрики качества:

- полное/частичное совпадение выделенных терминов

Особенности соревнования: **(дедлайн продлён до 1 февраля 2025)**

- вложенные термины, мультидоменная и мультижанровая постановка задачи
- разметка: 1150 русскоязычных аннотаций,
20 статей конференции Диалог 2000-2023 (домен компьютерной лингвистики)
250 аннотаций статей пяти других доменов

Конкурс ПРО//ЧТЕНИЕ

(<http://ai.upgreat.one>)



Задача: автоматическая разметка смысловых ошибок в сочинениях ЕГЭ по русскому яз., литературе, истории, обществознанию, английскому яз.

Период: декабрь 2019 — июнь 2022, три цикла испытаний.

Призовой фонд: ₹100М русский язык + ₹100М английский язык

Типов ошибок: 152

(р:70 л:16 о:23 и:20 а:23)

Подтипов ошибок: 236

(р:112 л:19 о:29 и:26 а:50)

Помимо выделения ошибок, надо давать их объяснения.

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

Конкурс ПРО//ЧТЕНИЕ

(<http://ai.upgreat.one>)

Up Great
technology
contest



comprehension technology

Сравнение двух разметок:

Алгоритмическая разметка

Нередко люди совершают плохие поступки, забывая о том, что, даже скрыв свой поступок от других, человек не скроется от своей совести. Что же такое безнравственный поступок? Безнравственный поступок - это поступок, не соответствующий моральным нормам.

Можно ли оправдать безнравственный поступок? Именно эту проблему В. Ф. Тендряков поднимает в своем тексте. Докажем сказанное примерами из представленного отрывка.

В тексте В. Ф. Тендряков говорит о том, что человек во благо себе может легко совершить низкий поступок, не испытав при этом чувство стыда. Человек сможет оправдать свой поступок перед самим собой, объяснив причину. В пример автор приводит поведение героя, который часто в жизни совершал безнравственные поступки. Он врал, дрался и крал. Мы видим, что до войны герой привык совершать плохие поступки. Он всегда оправдывался, потому что не хотел нести ответственность за свои действия, а значит не испытывал мучения совести. Мы знаем, что муки совести - это первое и самое сильное наказание, которое получает человек, совершивший плохой поступок. Но наш герой не получал никакого наказания и поэтому продолжал совершать безнравственные поступки. Проанализировав поведение главного героя, я убедилась в том, что человек обязан нести ответственность за свои поступки всегда, и поэтому я утверждаю, что нельзя оправдывать даже мелкие безнравственные поступки.

связь Р.ПОВТОР
Р.ПОВТОР Р.ЛИШН ПРОБЛЕМА
Р.ПОВТОР Р.ПОВТОР Р.ПОВТОР
Р.ЛИШН
Р.ПОВТОР
Р.ПОВТОР
Р.ПОВТОР
Р.ПОВТОР Г.ВИДОВР Р.ПОВТОР
Р.ПОВТОР Р.ПОВТОР
Р.ПОВТОР Р.ПОВТОР
Р.ПОВТОР Г.ВИДОВР Р.ПОВТОР
Р.ПОВТОР
Р.ПОВТОР

Экспертная разметка 2

Нередко люди совершают плохие поступки, забывая о том, что, даже скрыв свой поступок от других, человек не скроется от своей совести. Что же такое безнравственный поступок? Безнравственный поступок - это поступок, не соответствующий моральным нормам.

Можно ли оправдать безнравственный поступок? Именно эту проблему В. Ф. Тендряков поднимает в своем тексте. Докажем сказанное примерами из представленного отрывка.

В тексте В. Ф. Тендряков говорит о том, что человек во благо себе может легко совершить низкий поступок, не испытав при этом чувство стыда. Человек сможет оправдать свой поступок перед самим собой, объяснив причину. В пример автор приводит поведение героя, который часто в жизни совершал безнравственные поступки. Он врал, дрался и крал. Мы видим, что до войны герой привык совершать плохие поступки. Он всегда оправдывался, потому что не хотел нести ответственность за свои действия, а значит не испытывал мучения совести. Мы знаем, что муки совести - это первое и самое сильное наказание, которое получает человек, совершивший плохой поступок. Но наш герой не получал никакого наказания и поэтому продолжал совершать безнравственные поступки. Проанализировав поведение главного героя, я убедилась в том, что человек обязан нести ответственность за свои поступки всегда, и поэтому я утверждаю, что нельзя оправдывать даже мелкие безнравственные поступки.

Р.ПОВТОР T1
Р.ПОВТОР T1
Р.ПОВТОР T2 Р.ПОВТОР T1
ПРОБЛЕМА Р.ПОВТОР T2
ПРИМЕР Р.ПОВТОР T3
Р.ТАВТ T4 Р.ПОВТОР T1 Р.П.
Р.ПОВТОР T1
Р.ТАВТ T4
Р.ПОВТОР T1
Р.ТАВТ T4 Р.ПОВТОР T1
Р.ТАВТ T4 Р.ПОВТОР T1
Р.ТАВТ T4 Р.ПОВТОР T1
ПОЯСНЕНИЕ
Р.ПОВТОР T1
Р.ПОВТОР T1

Разметка текста: обобщённый контент-анализ

Пик научной фантастики (и советской, и западной) пришелся на 1960–1970-е годы. Однако в 1970-х годах этот жанр начал постепенно затухать и сходить на нет, уже в 1980-х на Западе начинает набирать силу жанр фэнтези. Конечно же, это неслучайно. Именно 1960-е годы стали пиком научно-технического прогресса в XX веке. К тому времени закончилась первая половина XX столетия, за эти полсотни лет было изобретено столько, что все казалось возможным, верилось, что прогресс будет нарастать по экспоненте. **1960-е — это мир безудержного социального и культурно-технического оптимизма.** Человек полетел в космос, запустил искусственные спутники и задумался об освоении других планет.

Но этот порыв человечества в будущее создавал определенную угрозу для власти имущих как на Западе, так и в Советском Союзе. И уже в 1960-е годы перед сотрудниками Тавистокского института изучения человека в Великобритании (причем по иронии судьбы он располагается в графстве Девоншир, рядом с дартмурскими болотами, где разыгрывалась мрачная драма «Собаки Баскервильей» Конан Дойля) **была поставлена задача притормозить научно-технический прогресс путем внедрения определенных информационно-психологических и организационных моделей.** В частности, стартовала работа по созданию молодежных и женских субкультур и движений (именно в это время как по заказу появились The Beatles, The Rolling Stones, стал развиваться экологизм).

Одна из главных задач, поставленных перед Тавистокком, звучала так: to stamp out the cultural optimism of the 1960s (искоренить, вырубить, вытравить культурный оптимизм 1960-х годов). А **научная фантастика, особенно советская, безусловно, была оптимистической по своему настрою.**

Некоторые менее оптимистические ноты (не могу их назвать пессимистическими, но они выглядели более сложными, чем просто оптимизм) прослеживались у ряда писателей в соцлагере, в частности в книгах Станислава Лема (достаточно почитать его «Астронавтов» и «Магелланово облако»). Однако общий настрой советской фантастики до середины 1960-х годов был преимущественно оптимистичным — это видно и по творчеству братьев Стругацких, и по романам Ивана Ефремова.

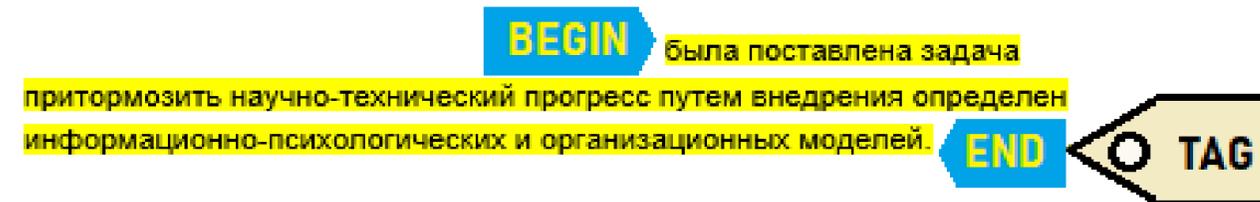
Первый доклад Римскому клубу (он создан в 1968 году) назывался «Пределы роста». В нем утверждалось, что человечество в своем индустриальном развитии достигло пределов, избыточно давит на природную среду, надо тормозить промышленно-экономическое развитие, перейдя к «нулевому росту». То есть 50 процентов всех средств должно идти на нейтрализацию негативных последствий, которые несет индустриальное развитие.

Разметка состоит из элементов

Элемент разметки — несколько взаимосвязанных фрагментов, затекстов и тегов

Теги (классы) выбираются из рубрикатора

Фрагмент задаётся началом и концом, может иметь один или несколько тегов:



Затекст — комментарий, объяснение, дополнительная информация и т.п., может иметь один или несколько тегов

Инструмент разметки

<https://markup.mlsa-iai.ru>

Весы войны Весь этот этический разлом дискуссии о военных действиях в Украине о том: стоит ли защита государства детской слезы? Стоит ли спасение некой организационной формы человеческого существования реально человеческой жертвы? Когда мои знакомые, скажу больше родственники, берут оружие и защищают свой город Львов - я понимаю их... Для них нет многосложных расчётов рисков России, для них нет геополитических вопросов и политики вообще. Есть город их детства, есть шум взрывов и солдаты другой страны... Когда Президент моей страны говорит о том, что перейдены все границы дозволенного, и мы вынуждены делать то, что делаем; а я смотрю на многолетние убийства и издевательства в ДНР; я вижу и знаю сколько денег и сил вложено в уничтожение России и подрыв наших ценностей - я понимаю это геополитическое решение. Проблема в том, что в пацифистском сознании любого современного образованного человека война - очень сложная и абстрактная химера. Когда сталкиваешься с ее реальностью становится страшно и хочется под лавку. Но России и нам россиянам под лавку больше нельзя. Мы сидели там долго и растеряли все. А на этой лавке сидели лорды и владельцы денег всех мастей. Никто не одобряет войну, об этом мы говорим все эти дни, но одуревший сосед, который ломится в НАТО и...

← Вернуться

Сохранить и выйти

Есть город их детства, есть шум взрывов и солдаты другой страны X

Есть город их детства, есть шум взрывов и солдаты другой страны X

+ добавить затекст

+ добавить затекст элемента разметки

добавить теги к элементу разметки

Тональность: отрицательная
Жилище
Жизнь
Чувство принадлежности (единство народов)
Этничность

сколько денег и сил вложено в уничтожение России X

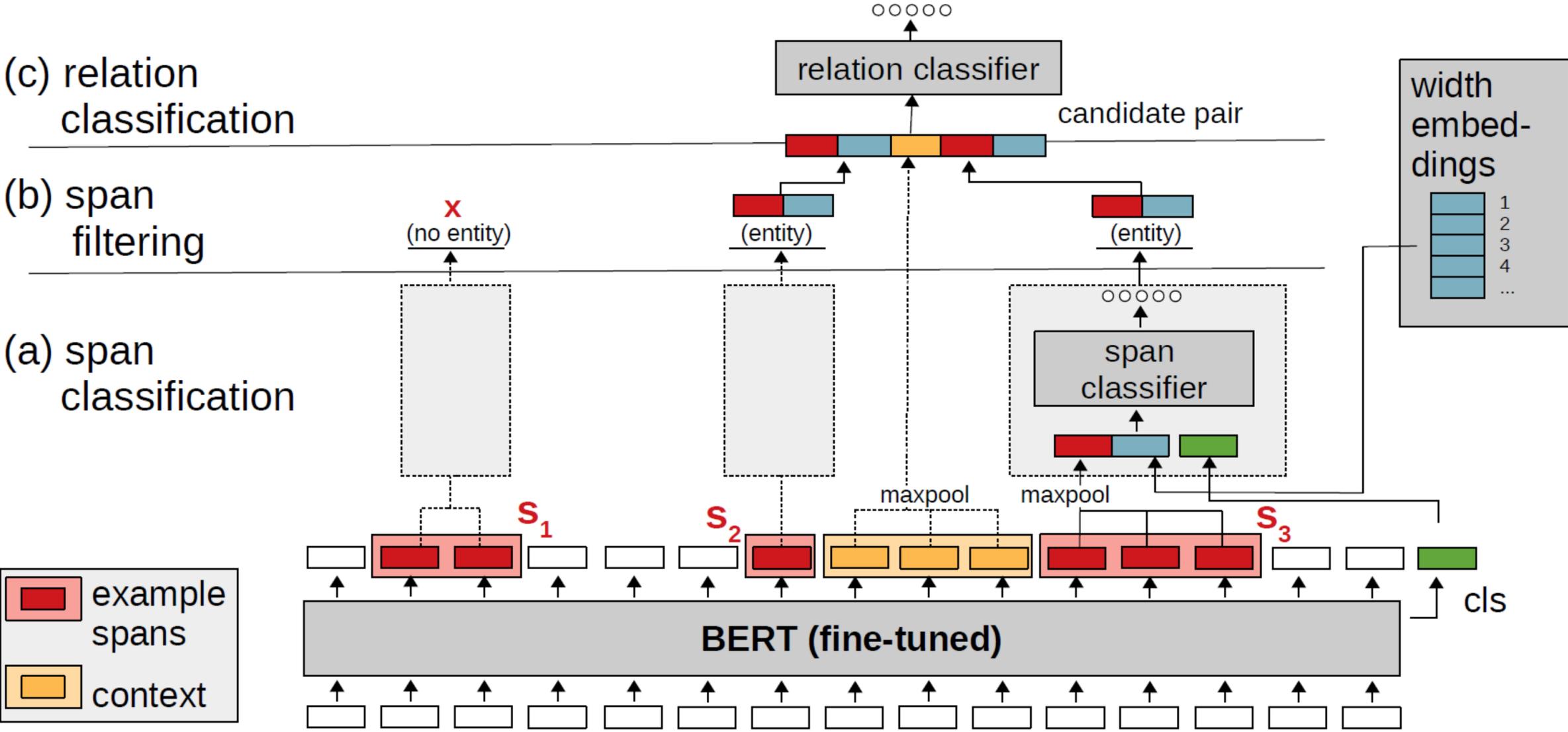
на этой лавке сидели лорды и владельцы денег всех мастей X

+ добавить элемент разметки

Введите имя тега

- ▶ Группа витальных ценностей
- ▶ Группа морально-нравственных ...
- ▶ Группа политических ценностей
- ▶ Группа религиозных ценностей
- ▶ Группа социальных ценностей
- ▶ Группа экзист. и познавательн...
- ▶ Группа эстетич. и гедонистичес...
- ▶ Пустое
- ▶ Служебные метки
- ▶ Тональность

Нейросетевые обучаемые модели разметки

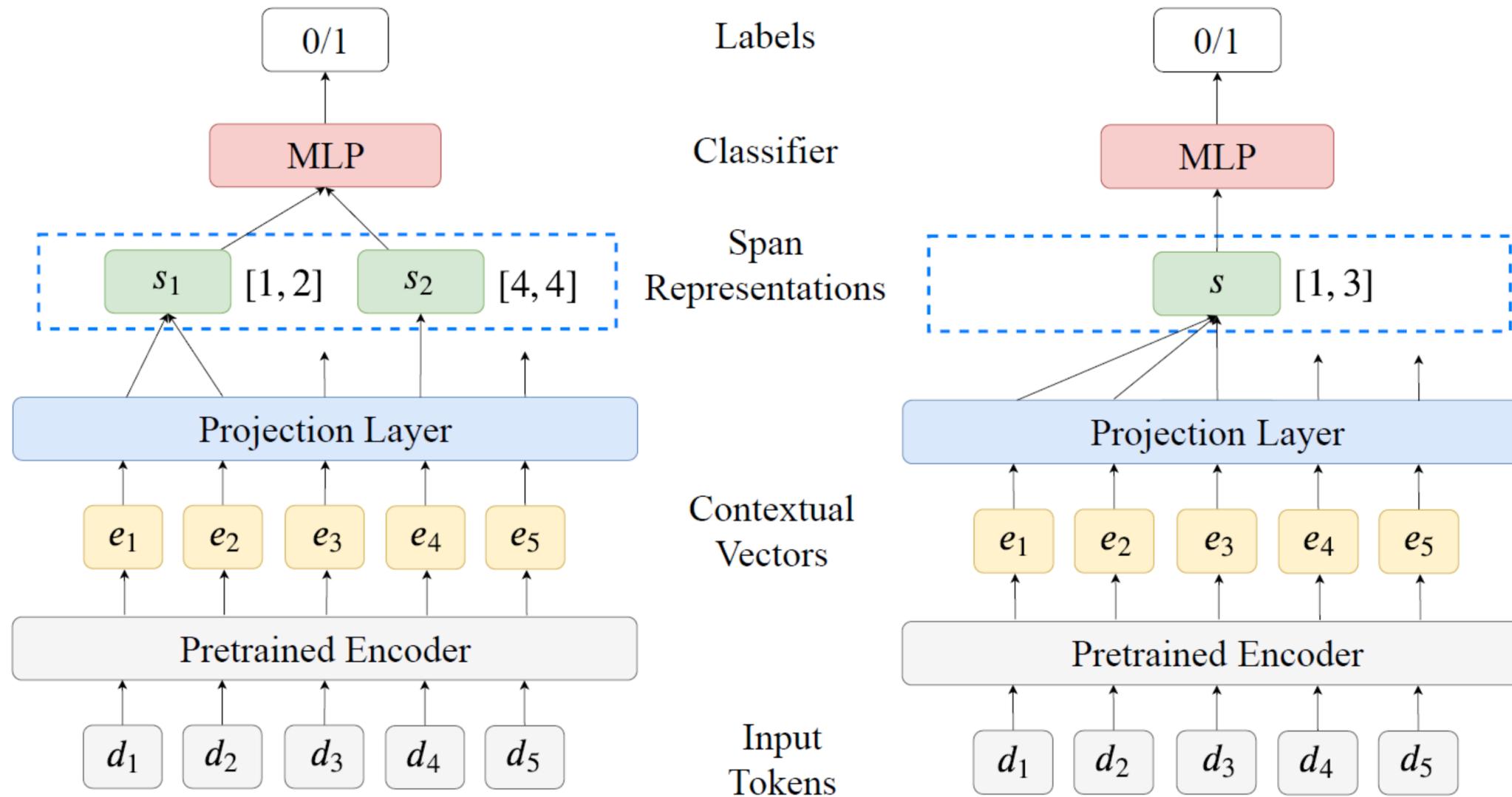


M.Eberts, A.Ulges. Span-based joint entity and relation extraction with transformer pre-training. 2020.

L.Anisiutin, T.Batura, N.Shvarts. Information extraction from news texts using a joint deep learning model. 2021.

Wayne Xin Zhao et al. A Survey of Large Language Models. ArXiv, 29 Jun 2023.

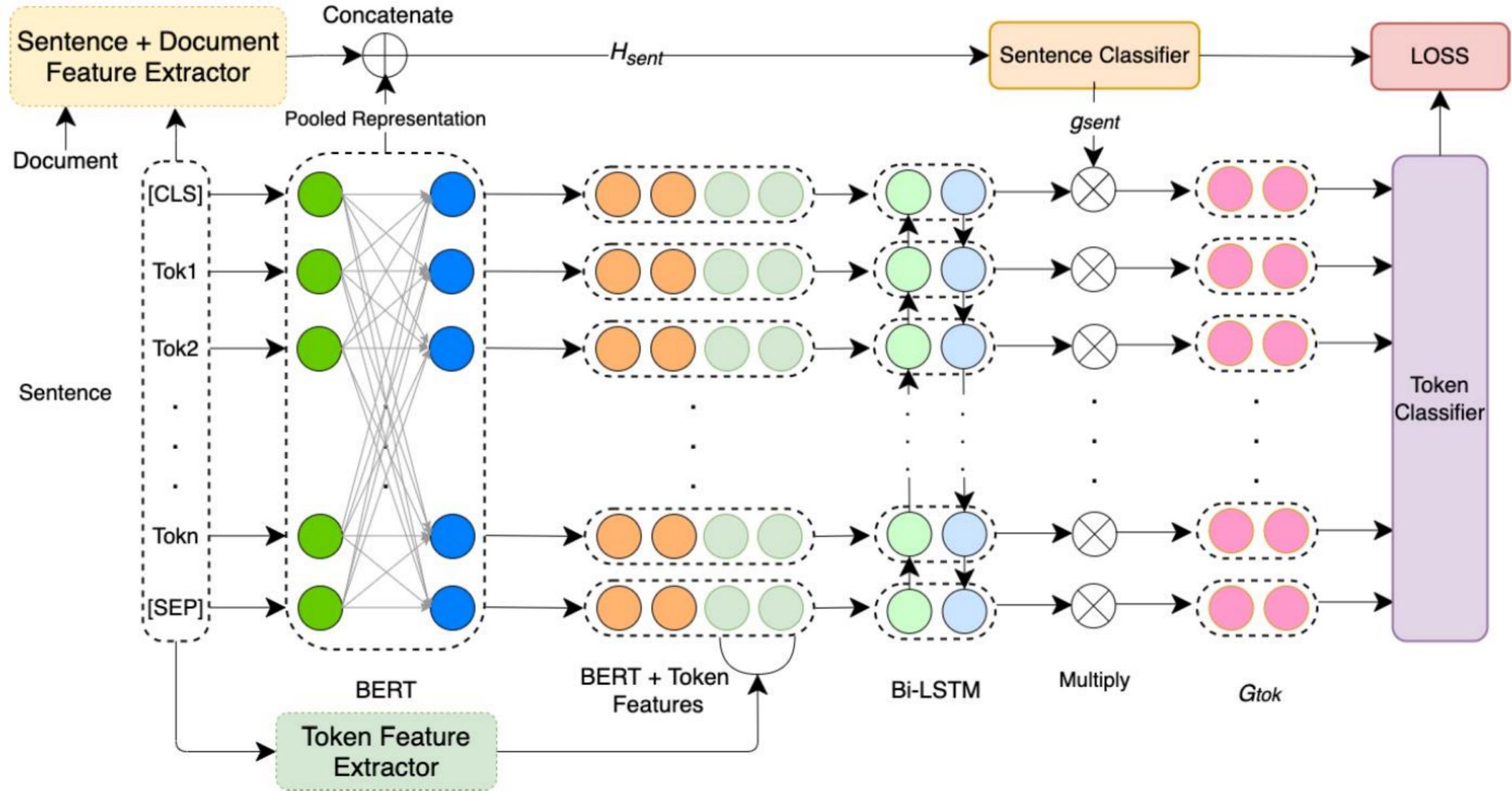
Нейросетевые обучаемые модели разметки



Xiaoya Li et al. A Unified MRC Framework for Named Entity Recognition. 2022.

S.Toshniwal et al. A Cross-Task Analysis of Text Span Representations. 2020.

Нейросетевые обучаемые модели разметки



Sopan Khosla et al. LTIatCMU at SemEval-2020 Task 11: Incorporating Multi-Level Features for Multi-Granular Propaganda Span Identification. 2020.

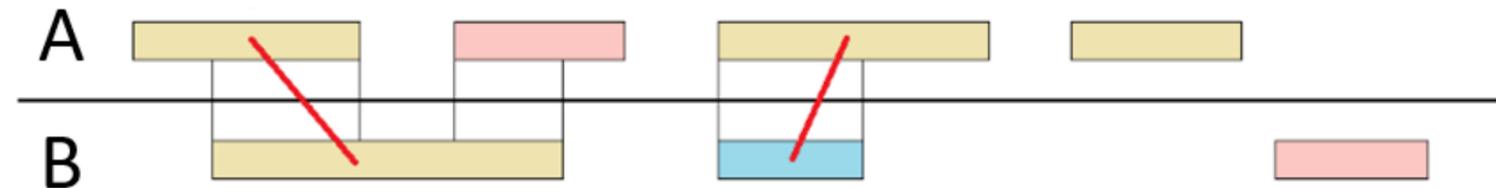
Оценивание алгоритмической разметки



- В основе методики — парное сравнение разметок текста:
«алгоритм ↔ эксперт», «эксперт-1 ↔ эксперт-2»
на основе оптимального сопоставления их элементов
- Вводятся меры согласованности пары разметок $C_n(A, B)$, $n = 1, \dots, N$
- Вводится их средневзвешенная согласованность $Con(A, B)$
- СТАР (Средняя Точность Алгоритмической Разметки)
— средняя по выборке $Con(A, E)$ разметок алгоритма A и эксперта E
- СТЭР (Средняя Точность Экспертной Разметки)
— средняя по выборке $Con(E_1, E_2)$ разметок двух экспертов, E_1 и E_2
- ОТАР = СТАР / СТЭР, если больше 100%, то модель лучше экспертов

Критерии согласованности разметок

Оптимальное сопоставление элементов разметок A и B



Критерии (числовые величины от 0 до 1; чем выше, тем лучше):

C_1 = доля фрагментов, для которых найдено сопоставление

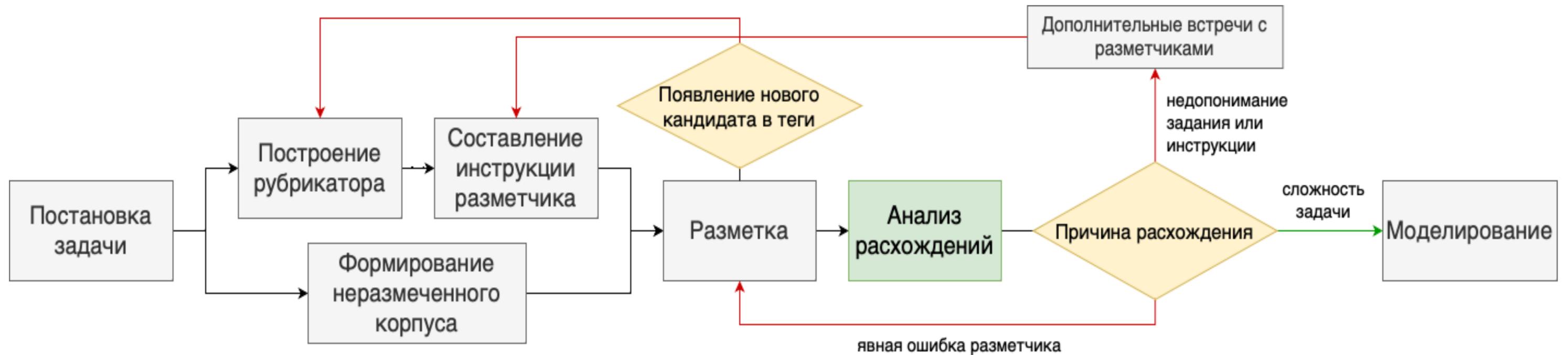
C_2 = точность наложения сопоставленных фрагментов

C_3 = точность совпадения тегов сопоставленных фрагментов

C_4 = точность совпадения связей сопоставленных фрагментов

C_5 = точность совпадения затекстов сопоставленных фрагментов

Организация процесса разметки



- каждый документ размечается несколькими экспертами (min 3)
- документы ранжируются по согласованности экспертов $Con(E, E')$
- наибольшие расхождения обсуждаются, вырабатывается консенсус
- происходит доработка инструкции и/или переразметка документов

Детекция ценностей социокультурного кода

Создание рубрикатора на основе кросс-дисциплинарного подхода

Исследователи	Предмет исследований
Милтон Рокич	Ценностные ориентации людей - психология, социология
Герт Хоффстеде	Культурные характеристики народов - социология
Шолом Шварц	Теория базовых человеческих ценностей - социальная психология
Рональд Инглхарт	Исследование мировых ценностей - политология, социология
Сэмюэл Хантингтон	Этнокультурное описание цивилизаций - политология, социология
Юрий Сергеевич Степанов	Концепты русской культуры - лингвистика
Александр Александрович Аузан	Культурные коды экономики - экономика

Указ президента Российской Федерации № 400 от 2-07-2021 «О стратегии нацбезопасности»

Указ президента Российской Федерации № 809 от 9-11-2022 «Об утверждении основ госполитики по сохранению и укреплению традиционных российских духовно-нравственных ценностей»

Какие ценности брать для рубрикатора: аксиоматический подход

- **Общественная значимость**
ценность — это то, что оказывает влияние на социальную жизнь
- **Индивидуальная значимость**
то, что влияет на принятие решений отдельными людьми
- **Субъективная измеримость**
то, что человек может принимать, отвергать или быть безразличным
- **Коммуникативность**
то, на отношение к чему можно повлиять в процессе коммуникации
- **Текстуальность**
то, что возможно описать, выразить текстом, фразой, историей
- **Атомарность**
то, что не сводится к набору других ценностей

Рубрикатор ценностей (1 из 2)

М. Рокич	Ш. Шварц	Г. Хофстеде
Ю.С. Степанов	Р. Инглхарт	Б.С. Ерасов

Группа социальных ценностей	
• Социальные ценности	• Личные границы
• Авторитет	• Материальные ценности
• Альтруизм	• Патриархальность
• Благородство происхождения / аристократизм	• Патриотизм
• Важность общественного мнения	• Пацифизм / мир во всём мире
• Воспитание	• Полезность (созидательный труд)
• Гендерное разнообразие	• Профессиональный успех
• Дети	• Репутация
• Долгосрочная ориентация	• Семья
• Дружба	• Социальное признание
• Избегание неопределённости	• Суеверия
• Индивидуализм	• Трудолюбие / продуктивность
• Интеллигентность	• Чувство принадлежности / единство народов
• Коллективизм	• Этничность
• Культура (нормы) поведения	• Язык

Группа витальных ценностей
• Витальные (необходимые) ценности
• Безопасность (личная)
• Время
• Еда
• Жизнь
• Жилище
• Здоровье
• Природа

Группа политических ценностей
• Политические ценности
• Власть
• Выборность власти (демократия)
• Институциональное доверие
• Историческая память и преемственность поколений
• Либерализм
• Национальная безопасность
• Права и свободы
• Правосознание (гражданская активность, гражданственность)
• Справедливость

Рубрикатор ценностей (2 из 2)

М. Рокич	Ш. Шварц	Г. Хофстеде
Ю.С. Степанов	Р. Инглхарт	Б.С. Ерасов

Группа религиозных ценностей

- Религиозные ценности
- Благочестивость
- Бог
- Религиозность
- Эзотерика

Группа эстетических и гедонистических ценностей

- Эстетические и гедонистические ценности
- Жизнь, полная впечатлений
- Красота
- Культура и искусство
- Наслаждение жизнью
- Потворство желаниям
- Творчество
- Эстетика

Группа экзистенциальных и познавательных ценностей

- Экзистенциальные и познавательные ценности
- Интеллект
- Критическое мышление
- Любовь
- Любознательность
- Мудрость
- Образование
- Перфекционизм
- Познание
- Принятие жизни
- Развитие
- Самостоятельность (выбор собственных целей)
- Смелость
- Смысл жизни
- Спокойствие (внутренняя гармония)
- Счастье
- Талант
- Твёрдая воля
- Целеустремлённость
- Широта взглядов

Группа морально-нравственных ценностей

- Морально-нравственные ценности
- Аккуратность
- Беспечность
- Вежливость
- Верность
- Взаимопомощь и взаимоуважение
- Гордость
- Гостеприимство
- Гуманизм
- Дисциплина
- Достоинство (самоуважение, самооценочность)
- Духовность (приоритет духовного над материальным)
- Заботливость
- Искренность
- Милосердие
- Ответственность
- Скромность
- Смирение (послушание, кротость)
- Совесть (нравственный закон, мораль)
- Терпение
- Уважение традиций
- Целомудрие
- Честность
- Чувство юмора

Ценностный ландшафт

Самые частотные теги
(по количеству элементов)

40%



1	Материальные ценности	1569
2	Жизнь	478
3	Историческая память и преемственность поколений	329
4	Правосознание (гражданская активность, гражданственность)	267
5	Политические ценности	246
6	Семья	243
7	Этничность	237
8	Культура и искусство	236
9	Права и свободы	235
10	Патриотизм	233

1. *Rink Olga, Lobachev Viktor, Vorontsov Konstantin*. Detecting human values and sentiments in large text collections with a context-dependent information markup: a methodology and math. HCI 2024. Lecture Notes in Computer Science series (in print). Cham: Springer.
2. *Vorontsov K.V., Gladchenko I.A., Lobachev V.A., Mamontova A.V., Rink O.L., Shabelskaya N.K.* Methodology for detecting human values in large text collections // Bulletin of St. Petersburg University. International relations. In Russian (in print)

Список публикаций (продолжение)

3. *Rink O. L., Vorontsov K. V., Shabelskaya N. K.* Uncovering positivism, negativism, and conflict in large text collections. Material values and the code of «noble maidens» // EDN ADEUQP, April 18–20, 2024. – P. 137-139.
4. *Vorontsov K.V., Gladchenko I.A., Lobachev V.A., Mamontova A.V., Rink O.L., Shabelskaya N.K.* Developing an Open Interdisciplinary Classifier of Human Values by means of Annotating Multi-fragments // Bulletin of St. Petersburg University. International relations. 2024 (in print)
5. *Maysuradze A.I., Rink O.L., Fedorov A.M., Tabachenkov A.M., Vorontsov K.V.* Does annotating multi-spans improve classification in large text collections? 2024ICSAI (China & IEEE). Lecture Notes in Computer Science series, 2024 (in print)
6. *Karabulatova I.S., Vorontsov K.V., Okolyshev D. A.* Problems of parametrization of the concept "Municipal Government" in the media space: stages of the development of an information and analytical system for assessing the social and communicative potencial of municipal authorities in the regions. Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie [Science Journal of Volgograd State University. Linguistics], 2024

Список публикаций (продолжение)

7. *Karabulatova I.S., Vorontsov K.V., Okolyshev D. A.* Problems of parametrization of the concept "Municipal Government" in the media space: stages of the development of an information and analytical system for assessing the social and communicative potencial of municipal authorities in the regions. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie [Science Journal of Volgograd State University. Linguistics]*, 2024
8. *Воронцов К. В., Карабулатова И. С., Катцина Т. А., Бородина Н. В., Сеницын М. В.* Проблемы разработки информационно-аналитической системы по истории социальной помощи в Енисейской губернии (XIX- первая четверть XX в.)// *Вопросы истории*, 2022, №12-3. С. 150 -161
9. *Карабулатова, И. С., Копнина Г. А.* Специфика параметризации деструктивного массмедийного текста с обесцениванием исторической памяти // *Медиалингвистика*. 2023. Т.10. №.3. С. 319-335.
10. *Kussepoova G. T., Karabulatova I. S., Kenzhigozhina K. S. Bakhus A. O., Vorontsov K. V.* Verification of communicative types in the judicial public space of media discourse in the USA, Kazakhstan and Russia as a psycholinguistic marker of fact-checking. *Amazonia Investiga*, 2023, 12(61), pp.131-144.
11. *Karabulatova I.S., Vorontsov K.V.* Digital linguistical migrationology: the possibilities of artificial intelligence in the study of migration processes // *Scientific and practical materials of the 2-nd world Congress in real and virtual mode "West-East: crossing cultures"*, 2-6 October 2019, Japan, Kyoto Sange University, volume 2, pp. 760-766.