

Автоматическое порождение аргументационных текстов в экономической сфере*

Е. В. Котельников¹, Д. Осадчий², И. Н. Фищева³

¹доктор технических наук, доцент, профессор кафедры прикладной математики и информатики, Вятский государственный университет. Россия, г. Киров;

старший научный сотрудник Национального центра когнитивных разработок, Университет ИТМО. Россия, г. Санкт-Петербург. ORCID: 0000-0001-9745-1489. E-mail: kotelnikov.ev@gmail.com

²магистрант факультета цифровых трансформаций, Университет ИТМО.

Россия, г. Санкт-Петербург. ORCID: 0000-0001-5131-7866. E-mail: da.osadchiy@gmail.com

³старший преподаватель кафедры прикладной математики и информатики, Вятский государственный университет. Россия, г. Киров. ORCID: 0000-0002-6941-2009. E-mail: fishchevain@gmail.com

Аннотация. Разработка больших и сверхбольших языковых моделей, таких как GPT-3, T5, Switch Transformer, ERNIE и др. позволила в последнее время значительно повысить качество генерации текстов. Одним из важных направлений в этой области является порождение текста с аргументами. Решение такой задачи может быть использовано при проведении деловых совещаний, в политических дебатах, в диалоговых системах, при подготовке студенческих эссе. Одной из основных предметных областей в указанных приложениях является экономическая сфера.

Ключевой проблемой при генерации аргументов для русского языка является дефицит корпусов, размеченных по аргументации. В настоящей работе мы используем переводные версии корпусов Argumentative Microtext, Persuasive Essays и UKP Sentential для обучения моделей на основе RuBERT и XGBoost. Далее построенные модели используются для разметки по аргументации корпуса экономических новостей. Затем размеченный корпус применяется для дообучения модели ruGPT-3, которая порождает аргументационные тексты. Результаты показывают, что такой подход позволяет повысить правильность генерации аргументов на 9 процентных пунктов (60 % против 51 %) по сравнению с исходной моделью ruGPT-3.

Ключевые слова: извлечение аргументации, генерация текстов, ruGPT-3, RuBERT, XGBoost.

Автоматическая генерация текста в последнее время достигла впечатляющих успехов в связи с разработкой больших и сверхбольших предобученных языковых моделей [7], таких как GPT-3 [2], T5 [13], Switch Transformer [3], ERNIE [22] и др. Эти модели позволяют осуществлять настройку на решаемую задачу при помощи обновления весов на небольшой обучающей выборке (дообучение) или без обновления весов в режимах few-shot learning (с несколькими обучающими примерами, как правило от 10 до 100), one-shot learning (один обучающий пример) и даже zero-shot learning [2].

Одним из важных направлений в генерации текста является порождение текста с аргументами [6; 8; 17]. Аргументы в этом случае либо находятся при помощи информационно-поисковой системы [8], либо генерируются предобученной языковой моделью [1; 6]. Аргумент – это совокупность высказываний, включающая утверждение и доводы [21]. Под утверждением понимается некоторое высказывание, выражающее потенциально спорную точку зрения. Доводами являются высказывания, подтверждающие или опровергающие данное утверждение.

Системы, позволяющие порождать текст с доводами относительно заданного утверждения, могут применяться в ходе деловых совещаний для оперативной генерации аргументов; в политических дебатах; в юриспруденции для поиска и генерации аргументов по законодательным актам и прецедентам; в диалоговых системах для подбора аргументов с целью убеждения собеседника; в образовании при анализе, генерации и оценке аргументации в студенческих работах. Одной из наиболее распространенных предметных областей для указанных приложений является экономическая сфера.

© Котельников Е. В., Осадчий Д., Фищева И. Н., 2021

* Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации и Германской службы академических обменов (DAAD) в рамках международного научно-образовательного сотрудничества по программе «Михаил Ломоносов» по теме: «Риторические модели порождения текста».

Для русского языка существует несколько общедоступных предобученных языковых моделей, в том числе RuBERT [10], SBERT [16] и ruGPT-3 [14]. Также в последнее время появилось несколько работ, посвященных анализу аргументации на русском языке [4; 5; 9; 15].

Однако до сих пор не было работ, посвященных генерации аргументативных русскоязычных текстов. В настоящей работе предпринимается попытка закрыть этот пробел и предлагается подход для построения языковой модели, позволяющей генерировать аргументативный текст на русском языке в экономической сфере.

Схема предлагаемого подхода для генерации аргументативных текстов включает три шага. На *первом шаге* существующие корпуса с аргументативной разметкой (Argumentative Microtext, PersEssays и UKP Sentential) используются для дообучения предобученной модели RuBERT [10] и для обучения модели XGBoost. Указанные модели показали наилучшие результаты в задаче классификации аргументов [5]. Рассматривается задача классификации «довод/не довод» на уровне предложения. На *втором шаге* обученные модели независимо друг от друга классифицируют предложения экономического корпуса. На *третьем шаге* осуществляется дообучение предобученной модели ruGPT-3 [14] в два этапа. На первом этапе с этой целью были использованы по 1000 предложений от каждой модели с наивысшими оценками вероятности отнесения к классу «довод». На втором этапе применялись 100 предложений, размеченных на предыдущем шаге вручную. Для оценки и сравнения качества исходной модели и дообученной использовались 10 затравок, для каждой из которых обе модели генерировали 10 потенциальных предложений-доводов. Полученные 200 предложений размечались вручную.

В исследовании использовались три существующих корпуса с аргументативной разметкой – Argumentative Microtext, PersEssays и UKP Sentential, а также корпус экономических новостей.

Argumentative Microtext Corpus был предложен в [12; 18]. Корпус включает 283 текста по разным темам (повышение пенсионного возраста, страхование здоровья, школьная униформа и т. п.). Каждый текст содержит одно утверждение относительно некоторой темы и от 2 до 10 ADUs (argumentative discourse unit), размеченных как доводы «за» или «против» данного утверждения. ADU – это фрагмент текста, который имеет единственное аргументационное значение [21, р. 63]. В Argumentative Microtext ADU может быть как целым предложением, так и частью предложения. В дальнейшем изложении мы считаем каждое отдельное ADU предложением.

Persuasive Essays Corpus (PersEssays) был введен в [20]. Он содержит 399 текстов по широкому спектру тематик (школьное образование, иммиграция, экономическая политика государства и т. п.). В качестве ADU выступают предложения. Разметка предложений выполнена по четырем типам: основное утверждение, утверждение, довод и нейтральный элемент. При формировании обучающего корпуса для классификации «довод/не довод» основные утверждения были исключены, просто утверждения и собственно доводы использованы как доводы, а нейтральные элементы – как «не доводы».

UKP Sentential Argument Mining Corpus (UKP Sentential) был предложен в [19]. Корпус включает 25492 предложения, размеченных как доводы «за», «против» или «не является доводом» по отношению к одной из восьми тем (аборт, клонирование, смертная казнь и т. п.). В корпусе имеются 489 предложений, разметка которых отличается для разных тематик. Такие предложения были исключены.

Фищева и Котельников [5] показали, что при машинном переводе англоязычного корпуса Argumentative Microtext на русский язык лучший результат среди систем Google Translate, Yandex.Translate и Promt продемонстрировал Google Translate. Поэтому все англоязычные корпуса в этой работе были переведены на русский язык с помощью Google Translate.

Для дообучения модели ruGPT-3 в сфере экономики использовался корпус русскоязычных экономических новостей, собранных с сайта banki.ru. Исходный корпус включал 7759 текстов за период с 01.06.2019 по 12.07.2021. Для текстов была осуществлена сегментация на предложения на основе библиотеки *Stanza*¹ (средний размер текстов – девять предложений) и токенизация с помощью библиотеки *nlTK*². По 10 % самых длинных и самых коротких предложений, а также повторяющиеся предложения были удалены. В результате предобработки было получено 68859 предложений со средней длиной 15 слов.

На первом шаге предложенный подход предполагает дообучение предобученной модели RuBERT и обучение модели XGBoost в задаче классификации предложений на два класса – «довод» и «не довод». Для обеих моделей использовалась одинаковая 5-кратная перекрестная проверка для получения оценок качества. Для подбора гиперпараметров для XGBoost применялась 4-кратная вложенная перекрестная проверка, для RuBERT – отложенная выборка (20 %):

¹ <https://stanfordnlp.github.io/stanza>.

² <https://www.nlTK.org>.

- XGBoost³: количество деревьев = [50, 150, **500**], максимальная глубина дерева = [2, 8, 20, 30];
- RuBERT: количество эпох = [3, **5**], размер батча = [4, **8**], скорость обучения = **10**⁻⁵.

Для XGBoost использовалось три типа признаков:

- лексические признаки – дискурсивные маркеры (“поэтому”, “следовательно”, “во-первых” и т. п.) и модальные слова (“следует”, “может”, “хочет” и т. п.), включая отрицания, всего 255 признаков;

- пунктуационные признаки – запятая, двоеточие, точка с запятой, вопросительный и восклицательный знаки, всего 5 признаков;

- морфосинтаксические признаки – N-граммы на основе частей речи (существительные, местоимения, глаголы, прилагательные и наречия), $N = \{2, 3, 4\}$, и грамматические признаки глаголов: время, наклонение, лицо; всего 783 признака. Морфологический анализ осуществлялся с помощью `mystem`⁴.

В отличие от [5] мы не добавили признаки предыдущего и следующего предложений, так как корпус UKP Sentential содержит только отдельные предложения, а не тексты.

Результаты кросс-валидации для объединенного корпуса (Argumentative Microtext \cup PersEsays \cup UKP Sentential) представлены в таблице 1.

Таблица 1

**Оценки качества для задачи классификации «довод»/«не довод»:
макроусредненная F1-мера, точность и полнота (среднее \pm стандартное отклонение)**

Модель	F1-мера	Точность	Полнота
XGBoost	0,6800 \pm 0,0066	0,6817 \pm 0,0065	0,6796 \pm 0,0066
RuBERT	0,7903 \pm 0,0051	0,7901 \pm 0,0051	0,7908 \pm 0,0050

XGBoost значительно уступает RuBERT в этой задаче, в отличие от результатов [4], где обе модели показали сопоставимое качество. Это связано с тем, что при обучении XGBoost в нашем исследовании не использовались контекстные признаки, являющиеся очень важными по результатам [4]. После оценки качества обе модели с подобранными оптимальными гиперпараметрами были обучены на всем объединенном корпусе. Построенные модели использовались на следующем этапе для классификации предложений экономического корпуса.

Для проверки качества классификации предложений на «довод»/«не довод» была проведена ручная разметка 50 предложений от каждой модели с наивысшими оценками вероятности отнесения к классу «довод» (всего 100 предложений). Задача разметки осложнялась тем, что для полученных доводов отсутствовали утверждения. Чтобы упростить разметку, аннотаторам предоставлялся заголовок новостной статьи, из которой было извлечено размечаемое предложение.

В разметке участвовало пять аннотаторов (носители русского языка). Перед ними была поставлена бинарная задача разметить предложения как содержащие или не содержащие довод (без указания того, какая модель их предоставила). Доводом следовало считать предложение, которое могло быть использовано для убеждения оппонента в каком-либо утверждении с учетом контекста, предоставляемого заголовком новостной статьи. Окончательное решение относительно предложения принималось на основании простого голосования оценок аннотаторов.

Для XGBoost точность классификации оказалась 56 % (28 из 50), для RuBERT – 36 % (18 из 50). Согласие аннотаторов, вычисленное по капле Флейса, равно 0,1734. Такой уровень согласия соответствует незначительному согласию по шкале Лэндиса и Коха [11]. Относительно низкий уровень согласия связан с тем, что аннотаторам не были предоставлены утверждения. При разметке доводов с утверждениями уровень согласия оказался значительно выше.

Несмотря на то, что XGBoost продемонстрировал более высокое качество при ручной разметке, было принято решение использовать для разметки экономического корпуса обе модели. Это связано с тем, что, во-первых, обе модели распознают в качестве доводов различные предложения; во-вторых, RuBERT показал себя лучше при автоматической оценке качества (см. Таблицу 1).

С целью построения модели генерации аргументативных текстов использовалось дообучение предобученной модели `ruGPT-3` [14] в два этапа. На первом этапе в качестве обучающих данных использовались по 1000 предложений из экономического корпуса, размеченные моделями XGBoost и RuBERT как «доводы» с наивысшими оценками уверенности (всего 2000 предложений).

Дообучалась версия `ruGPT3Large` с использованием Google Colab Pro (видеокарта NVIDIA Tesla P4) со следующими гиперпараметрами: количество эпох 5, размер батча 1, размер блока 128. На вход моде-

³ Оптимальные значения гиперпараметров выделены жирным – они оказались одинаковыми на всех разбиениях.

⁴ <https://yandex.ru/dev/mystem>.

ли подавались по отдельности предложения обучающего корпуса, которым приписывался префикс «потому что».

На втором этапе для обучения использовались 100 предложений, размеченных на предыдущем шаге вручную. Гиперпараметры и видеокарта были теми же. На следующем шаге обученная модель *ruGPT-3* тестировалась с помощью ручной разметки сгенерированных доводов.

Для тестирования обученной модели *ruGPT-3* использовались 10 затравок-утверждений (после каждого утверждения добавлялся суффикс «потому что»): «Банкам следует более широко использовать биометрию», «Государственные облигации являются одним из наиболее надежных видов ценных бумаг», «Деньги нужно вкладывать в акции», «Криптовалюты лучше фиатных валют», «Лучшей инвестицией является покупка недвижимости», «Не следует играть на валютном рынке», «Обучение финансовой грамотности зачастую приводит к необоснованной уверенности», «Покупка земельного участка является хорошей инвестиционной стратегией», «При оформлении кредитной карты стоит внимательно отнестись к выбору банка», «Сбережения следует хранить в валюте».

В качестве модели для сравнения применялась исходная модель *ruGPT3Large*. Для каждой затравки каждой моделью были сгенерированы 10 потенциальных предложений-дowodов (*zero-shot learning*). Для генерации использовались следующие параметры: $\text{Top-K}=50$, $\text{top-p}=0,92$.

Сгенерированные 200 предложений размечались вручную пятью аннотаторами. Аннотаторам предоставлялись пары <утверждение – предложение> без информации о том, какая модель сгенерировала данное предложение. Доводом следовало считать предложение, которое могло быть использовано для убеждения оппонента в заданном утверждении. Окончательное решение относительно предложения принималось на основании простого голосования оценок аннотаторов.

В результате ручной разметки точность дообученной модели *ruGPT-3* оказалась равна 60 % (60 доводов из 100 предложений), исходной модели – 51 (51 довод из 100 предложений).

Согласие аннотаторов, вычисленное по капле Флейсса, равно 0,4292. Такой уровень согласия соответствует умеренному согласию по шкале Лэндиса и Коха [11]. Уровень согласия оказался значительно выше, чем в предыдущей процедуре разметки отдельных доводов, в связи с тем, что аннотаторам предоставлялось утверждение.

Таким образом, в настоящем исследовании предложен подход для построения русскоязычной модели генерации доводов по заданному утверждению в экономической сфере. На наш взгляд, достигнутое качество генерации аргументативных текстов дообученными моделями позволяет с осторожностью говорить о возможности использования таких моделей на практике, например, для оперативной генерации доводов в ходе деловых совещаний.

В качестве перспективных направлений исследований следует выделить изучение влияния входного контекста на качество генерации аргументативных текстов моделей типа *GPT*, а также исследование ситуации переобучения в процессе дообучения моделей.

Список литературы

1. *Al-Khatib K., Trautner L., Wachsmuth H., Hou Y., Stein B.* Employing Argumentation Knowledge Graphs for Neural Argument Generation. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Pp. 4744–4754 (2021).
2. *Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J. et al.* Language Models are Few-Shot Learners. In: Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Pp. 1877–1901 (2020).
3. *Fedus W., Zoph B., Shazeer N.* Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. URL: <https://arxiv.org/abs/2101.03961> (2021).
4. *Fishcheva I., Goloviznina V., Kotelnikov E.* Traditional Machine Learning and Deep Learning Models for Argumentation Mining in Russian Texts. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog-2021". Pp. 246–258 (2021).
5. *Fishcheva I., Kotelnikov E.* Cross-lingual argumentation mining for Russian texts. In: Proceedings of the 8th International Conference "Analysis of Images, Social networks and Texts" (AIST 2019), Lecture Notes in Computer Science, 11832. Pp. 134–144 (2019).
6. *Gretz S., Bilu Y., Cohen-Karlik E., Slonim N.* The workweek is the best time to start a family – A Study of GPT-2 Based Claim Generation. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Pp. 528–544 (2020).
7. *Han X., Zhang Z., Ding N., Gu Y., Liu X., et al.* Pre-Trained Models: Past, Present and Future. AI Open. URL: <https://doi.org/10.1016/j.aiopen.2021.08.002> (2021).
8. *Hua X., Hu Z., Wang L.* Argument generation with retrieval, planning, and realization. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Pp. 2661–2672 (2019).
9. *Irina D., Kononenko I., Sidorova E.* On Developing a Web Resource to Study Argumentation in Popular Science Discourse. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog-2021". Pp. 318–327 (2021).

10. Kuratov Y., Arkhipov M. Adaptation of deep bidirectional multilingual transformers for Russian language. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog-2019". Pp. 333–340 (2019).
11. Landis J. R., Koch G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33(1). Pp. 159–174 (1977).
12. Peldszus A., Stede M. An annotated corpus of argumentative microtexts. In: Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation. Pp. 801–815 (2015).
13. Raffel C., Shazeer N., Roberts A., Lee K., Narang S. et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21. Pp. 1–67 (2020).
14. ruGPT-3 pretrained language model. URL: <https://sbercloud.ru/ru/warp/gpt-3>.
15. Salomatina N., Kononenko I., Sidorova E., Pimenov I. Identification of connected arguments based on reasoning schemes "from expert opinion". *Journal of Physics: Conference Series* 1715 (2021).
16. SBERT pretrained language model. URL: <https://developers.sber.ru/portal/services/sbert>.
17. Schiller B., Daxenberger J., Gurevych I. Aspect-Controlled Neural Argument Generation. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Pp. 380–396 (2021).
18. Skeppstedt M., Peldszus A., Stede M. More or less controlled elicitation of argumentative text: enlarging a microtext corpus via crowdsourcing. In: Proceedings of the 5th Workshop in Argumentation Mining. Pp. 155–163 (2018).
19. Stab C., Miller T., Schiller B., Rai P., Gurevych I. Cross-topic argument mining from heterogeneous sources. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2018). Pp. 3664–3674 (2018).
20. Stab C., Gurevych I. Annotating argument components and relations in persuasive essays. In: Proceedings of the International Conference on Computational Linguistics. Pp. 1501–1510 (2014).
21. Stede M., Schneider J. Argumentation Mining. *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers (2018).
22. Sun Y., Wang S., Feng S., Ding S., Pang C. et al. ERNIE 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation. URL: <https://arxiv.org/abs/2107.02137> (2021).

Automatic generation of argumentative texts in the economic sphere

E. V. Kotelnikov¹, D. Osadchiy², I. N. Fischeva³

¹Doctor of Technical Sciences, associate professor, professor of the Department of applied mathematics and computer science, Vyatka State University. Russia, Kirov; senior researcher at the National Center for Cognitive Development, ITMO University. Russia, St. Petersburg. ORCID: 0000-0001-9745-1489. E-mail: kotelnikov.ev@gmail.com

²master student of the Faculty of digital transformations, ITMO University.

Russia, St. Petersburg. ORCID: 0000-0001-5131-7866. E-mail: da.osadchiy@gmail.com

³senior lecturer of the Department of applied mathematics and computer science, Vyatka State University. Russia, Kirov. ORCID: 0000-0002-6941-2009. E-mail: fishchevain@gmail.com

Abstract. The development of large and ultra-large language models, such as GPT-3, T5, Switch Transformer, ERNIE and others, has recently significantly improved the quality of text generation. One of the important directions in this area is the generation of text with arguments. The solution of such a problem can be used during business meetings, in political debates, in dialogue systems, in the preparation of student essays. One of the main subject areas in these applications is the economic sphere.

The key problem in generating arguments for the Russian language is the shortage of corpora marked up by argumentation. In this paper, we use translated versions of Argumentative Microtext, Persuasive Essays and UKP Sentential corpora to train models based on RuBERT and XGBoost. Further, the constructed models are used for marking up according to the arguments of the economic news corpus. Then the marked-up corpus is used to retrain the ruGPT-3 model, which generates argumentative texts. The results show that this approach makes it possible to increase the correctness of argument generation by nine percentage points (60 % vs. 51 %) compared to the original ruGPT-3 model.

Keywords: argumentation extraction, text generation, ruGPT-3, RuBERT, XGBoost.

References

1. Al-Khatib K., Trautner L., Wachsmuth H., Hou Y., Stein B. Employing Argumentation Knowledge Graphs for Neural Argument Generation. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Pp. 4744–4754 (2021).
2. Brown T. B., Mann B., Ryder N., Subbiah M., Kaplan J. et al. Language Models are Few-Shot Learners. In: Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Pp. 1877–1901 (2020).
3. Fedus W., Zoph B., Shazeer N. Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity. Available at: <https://arxiv.org/abs/2101.03961> (2021).
4. Fischeva I., Goloviznina V., Kotelnikov E. Traditional Machine Learning and Deep Learning Models for Argumentation Mining in Russian Texts. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog-2021". Pp. 246–258 (2021).

5. *Fishcheva I., Kotelnikov E.* Cross-lingual argumentation mining for Russian texts. In: Proceedings of the 8th International Conference "Analysis of Images, Social networks and Texts" (AIST 2019), Lecture Notes in Computer Science, 11832. Pp. 134–144 (2019).
6. *Gretz S., Bilu Y., Cohen-Karlik E., Slonim N.* The workweek is the best time to start a family – A Study of GPT-2 Based Claim Generation. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Pp. 528–544 (2020).
7. *Han X., Zhang Z., Ding N., Gu Y., Liu X. et al.* Pre-Trained Models: Past, Present and Future. AI Open. Available at: <https://doi.org/10.1016/j.aiopen.2021.08.002> (2021).
8. *Hua X., Hu Z., Wang L.* Argument generation with retrieval, planning, and realization. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Pp. 2661–2672 (2019).
9. *Ilina D., Kononenko I., Sidorova E.* On Developing a Web Resource to Study Argumentation in Popular Science Discourse. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog-2021". Pp. 318–327 (2021).
10. *Kuratov Y., Arkhipov M.* Adaptation of deep bidirectional multilingual transformers for Russian language. In: Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialog-2019". Pp. 333–340 (2019).
11. *Landis J. R., Koch G. G.* The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33(1). Pp. 159–174 (1977).
12. *Peldszus A., Stede M.* An annotated corpus of argumentative microtexts. In: Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation. Pp. 801–815 (2015).
13. *Raffel C., Shazeer N., Roberts A., Lee K., Narang S. et al.* Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21. Pp. 1–67 (2020).
14. ruGPT-3 pretrained language model. Available at: <https://sbercloud.ru/ru/warp/gpt-3>.
15. *Salomatina N., Kononenko I., Sidorova E., Pimenov I.* Identification of connected arguments based on reasoning schemes "from expert opinion". *Journal of Physics: Conference Series* 1715 (2021).
16. SBERT pretrained language model. Available at: <https://developers.sber.ru/portal/services/sbert>.
17. *Schiller B., Daxenberger J., Gurevych I.* Aspect-Controlled Neural Argument Generation. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Pp. 380–396 (2021).
18. *Skeppstedt M., Peldszus A., Stede M.* More or less controlled elicitation of argumentative text: enlarging a microtext corpus via crowdsourcing. In: Proceedings of the 5th Workshop in Argumentation Mining. Pp. 155–163 (2018).
19. *Stab C., Miller T., Schiller B., Rai P., Gurevych I.* Cross-topic argument mining from heterogeneous sources. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2018). Pp. 3664–3674 (2018).
20. *Stab C., Gurevych I.* Annotating argument components and relations in persuasive essays. In: Proceedings of the International Conference on Computational Linguistics. Pp. 1501–1510 (2014).
21. *Stede M., Schneider J.* Argumentation Mining. *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers (2018).
22. *Sun Y., Wang S., Feng S., Ding S., Pang C. et al.* ERNIE 3.0: Large-scale Knowledge Enhanced Pre-training for Language Understanding and Generation. Available at: <https://arxiv.org/abs/2107.02137> (2021).