Вячеслав Мустакимов, Алексей Мустакимов

КонтрПлагиат методом перефразирования и рерайта для антиплагиат ВУЗ:

как повысить оригинальность текста за несколько часов и пройти проверку с первого раза

Обобщен многолетний опыт практического рерайта, все лучшее собрано и представлено в подробном изложении, с примерами. Технология КонтрПлагиат позволяет повышать уникальность академических текстов - бесплатно, со скоростью 1000 знаков в минуту, исключает рутину, оставляя время для настоящего научного творчества.



Вячеслав Мустакимов

КонтрПлагиат методом перефразирования и рерайта для антиплагиат ВУЗ. Как повысить оригинальность текста за несколько часов и пройти проверку с первого раза

«Издательские решения»

Мустакимов В. А.

КонтрПлагиат методом перефразирования и рерайта для антиплагиат ВУЗ. Как повысить оригинальность текста за несколько часов и пройти проверку с первого раза / В. А. Мустакимов — «Издательские решения»,

ISBN 978-5-00-643068-6

Обобщен многолетний опыт перефразирования и рерайта академических текстов для успешного прохождения антиплагиат-проверки.В монографии собрано и подробно представлено все лучшее из практики перефразирования научных текстов, с примерами.Технология КонтрПлагиат позволяет повышать уникальность академических текстов — бесплатно со скоростью 1000 знаков в минуту, исключает рутину, статус генеративного, «Подозрительного документа», оставляя время для настоящего научного творчества.

ISBN 978-5-00-643068-6

© Мустакимов В. А.
© Издательские решения

Содержание

Введение 6
1. Соотношение понятий КонтрПлагиата и антиплагиата 8
1.1. Процесс КонтрПлагиата для антиплагиата 8
1.2. Методы которые не работают в антиплагиат ВУЗ 15
2. Википедия перефразирования для антиплагиат ВУЗ: работающие (Р) и частично работающие методы (ЧР) 20
Конец ознакомительного фрагмента. 63

КонтрПлагиат методом перефразирования и рерайта для антиплагиат ВУЗ Как повысить оригинальность текста за несколько часов и пройти проверку с первого раза

Вячеслав Алексеевич Мустакимов
Алексей Вячеславович Мустакимов

© Вячеслав Алексеевич Мустакимов, 2024
© Алексей Вячеславович Мустакимов, 2024

ISBN 978-5-0064-3068-6
Создано в интеллектуальной издательской системе Ridero

Введение

Актуальность настоящей работы обусловлена тем, что сегодня многие авторы письменных работ обладают умением находить в Интернете нужные тексты, копировать найденное и вставлять в свои работы. С оптимизационной точки зрения автора, желательно копировать большие куски, в идеале, целые параграфы. Авторская «оптимизация» написания работ, в итоге, приводит к дефициту оригинальности текста. По нашим статистическим заключениям, средний процент уникальности написанной работы колеблется в диапазоне 10—30%, а для успешного прохождения антиплагиат проверки требуется 70—80%. Исходя из этого, процесс написания любой академической работы содержит выраженные стадии: сборка работы (компиляция) и повышение уникальности (перефразирование, рерайт). Данная стадийность условная, но для цели нашего исследования вполне приемлемая.

Понятие «рерайт» не является сакральным, однако понимание «легкого», «поверхностного» и «глубокого» рерайта, перефразирования – ускользает, нам не удалось найти исчерпывающего объяснения такого отличия, причем, выраженного числовыми, измеримыми показателями, раз уж мы говорим о экспертизе научных работ на плагиат.

Для выявления уровня оригинальности академических и научных работ имеется масса сервисов, готовых заработать на поиске плагиата: Copyscape, Grammarly, HelioBLAST, iThenticate, PlagScan, PlagTracker, Turnitin, Unicheck, StrikePlagiarism, ETXT, text, РуКонтекст, antiplagiat и т. д. В силу отсутствия измеримости показателя рерайта, как отличия текста источника и текста после рерайта, каждый сервис изобретает собственные критерии, и клиентская общность оперирует такими понятиями, как «жесткая» и «мягкая» проверка. Эта «экспертная» сумятица вносит определённый хаос, т.к. блестяще пройденная проверка на плагиат в одном сервисе, покажет «уникальную несостоятельность» работы в другом сервисе.

Поиск текстовых заимствований в России – повторение опыта зарубежных коллег, в 2005 году в нашей стране был введен надзорный инструмент под названием «Антиплагиат», который сформировал динамично растущий рынок рерайтинга (перефразирования). Сегодня антиплагиат известен под разными именами, само понятие антиплагиат – явление проверки на уникальность (оригинальность), а антиплагиат, расположенный на домене ру, он же ВУЗ, – это упоминание конкретного сервиса antiplagiat, на указанном домене.

Уточняя терминологию, отметим, что КонтрПлагиат – это специальные меры, направленные на исключение из текстов плагиата (заимствований), включающих перефразирование, глубокий рерайт и копирайтинг, а также комплекс нетрадиционных для рерайта методов, отграниченных от обычной практики, принятой в «научном письме». Главное отличие КонтрПлагиата от традиционного рерайта и копирайтинга – его доказательность и прогнозируемость результатов, другими словами, текст изменяется не с «поверхностным, мягким» или «глубоким» отличием, а на нормированное значение, и в статистическом большинстве это нормированное изменение текста дает высокий и необходимый процент при проверке в системах антиплагиат.

КонтрПлагиат оперирует двумя терминами, шингл – словосочетание из двух слов, например – «Внимание, документ подозрительный: в документе присутствует сгенерированный текст», фраза содержит следующие шинглы: «Внимание, документ», «документ подозрительный», «подозрительный в», «в документе», «документе присутствует», «присутствует сгенерированный», «сгенерированный текст». Иллюстрация понятия шингла утрирована, т.к. шингл учитывает текстовое содержание без предлогов и стоп-слов, в лемматизированной форме.

N-грамма, в упрощенном понимании, это словосочетание из нескольких слов, например: «Внимание, документ», «подозрительный: в», «документе присутствует», «сгенерированный текст». В примере выделены биграммы, которые следуют друг за другом, триграммы будут состоять из трех слов и т. д.

КонтрПлагиат, не новация, предусматривает активное использование как традиционных методов, так и современных информационных решений и технологий, таких как генеративный искусственный интеллект (ИИ), способный генерировать текст в ответ на подсказки (инструкции). Наравне с перечисленными методами используются малоизвестные методики, бесплатные, но эффективные, позволяющие выдавать большую уникальность текста, нежели этого требуют «жесткие» проверки. В отличии от сервисов поиска заимствований, КонтрПлагиат поясняет содержание и критерии понятия «жесткая» проверка, позволяет измерять параметры локально, а при достижении критериев документ отдается на антиплагиат проверку, что позволяет пройти ее с первого раза.

В практике высшей школы написание письменных работ тоже не новация, и этим занимаются веками как студенты, так и их преподаватели. Классика компилятивного метода (см. Умберто Эко) гласит, что работа составляется из материалов, добросовестно найденных в литературе, это могут быть факты, цитаты, определения и т. д. Однако, всё, что написано в учебной литературе, давно является достоянием Интернета – плагиатом. Библиотеки из источника знаний превратились в источник плагиата. В этой связи, согласно рекомендациям Высшей школы, все использованные текстовые материалы должны перефразироваться, излагаться «другими», академическими словами. К сожалению, высшая школа молчит о критериях такого перефразирования. Данное молчание подтверждает ретроспектива статей, опубликованных на elibrary.ru, за последнее десятилетие, согласно контексту публикаций, отечественный научный мир, на исходе двух десятилетий существования антиплагиата, пытается понять, нужен антиплагиат высшей школе или нет.

Заметное разнообразие в «рефлексию» elibrary.ru вносят работы Н. В. Авдеева, В. В. Домакова [1], Т. А. Иващенко [3], В. П. Лукьяненко [4], Н. И. Мартишиной, В. В. Николаева [12], Н. Н. Пунчик, В. Г. Свечкарёва [3], В. И. Сперанского, Е. С. Стегниенко, В. В. Тимофеева и др.

Не смотря на попытки научного осмысления «антиплагиат-реальности», сегодня имеется несколько миллионов студентов, которые решают проблемы повышения оригинальности текстов, обращаясь к сервисам антиплагиат десятки миллионов раз за год. Растущий спрос на уникальные тексты породил сферу услуг – рерайт и перефразирование академических, научных текстов. Спустя почти два десятилетия феномена антиплагиата в России спрос на услуги академического рерайта не снижаются, а нарастают, несмотря на появление ИИ, который грозился оставить рерайтеров без работы. В этой связи назрела необходимость научного переосмысления реального положения вещей, обобщения, накопленных в сфере академического рерайта, практических наработок, способствующих успешному прохождению проверки в антиплагиат, версии ВУЗ.

Настоящая монография, посвящена одному вопросу – «как сделать рерайт, перефразирование в короткие сроки и успешно пройти проверку в антиплагиат. ВУЗ», основана на практическом опыте авторов и предназначена всем, кто стремится повысить уникальность текстов с наименьшими затратами и в сжатые сроки. В монографии не рассматриваются суррогатные методы обхода антиплагиата, предпочтение отдается рекомендуемому ВУЗами методу перефразирования, однако он представлен в совершенно ином аспекте, с позиции нового процесса, который мы называем КонтрПлагиат.

Работа представлена в авторской лексике и пунктуации, редактура осуществлена с применением инструкций в среде GPT, что демонстрирует современный уровень развитие ИИ решений.

1. Соотношение понятий КонтрПлагиата и антиплагиата

1.1. Процесс КонтрПлагиата для антиплагиата

Идея проверки текстов на плагиат не нова. Её прообразом выступают решения, применяемые в поисковых системах. Благодаря этому, нельзя выйти в топ поиска скопировав материалы с другого сайта. Копии таких текстов будут расцениваться поисковыми системами как плагиат, и, если он систематичен, сайт может попасть в серый или черный список, рис. 1.



Рисунок 1 – Как поисковые системы видят тексты

В погоне за уникальностью авторы контента прибегают к таким известным методам как рерайт (латентный, сокрытый плагиат – текст переписывается своими словами с сохранением смысловой нагрузки), или копирайтинг (написание нового, похожего текста). Это не всегда помогает, так как технологии развиваются и некоторые алгоритмы анализируют смысл текста, а не просто последовательность n-грамм (последовательность из n элементов, слогов, слов).

На уровне сервиса, предназначенного для проверки академических текстов на наличие заимствований, одним из первых, в 1997 г., появился Turnitin (сегодня – более 30 млн студентов, 15 000 организаций и учреждений). Turnitin – это интернет-сервис для проверки уникальности текстов, который управляется одноименной американской компанией, дочерним подразделением Advance Publications. Организация продает лицензии университетам и обра-

зовательным организациям, отличным от высшей школы. Эти субъекты образовательного процесса используют специализированный веб-сайт Turnitin в качестве портала (окна входа) услуги SaaS, что дает возможность сравнения представленных документов с содержимым базы данных Turnitin, таким образом, выявляются заимствования – плагиат.

Turnitin использует специализированные «секретные» алгоритмы для сравнения представленных работ с множеством текстов баз данных, чтобы проверить возможное наличие неоригинального контента.

В начале 2023 года Turnitin запустил новую функцию, направленную на обнаружение контента, созданного приложениями искусственного интеллекта, такими как ChatGPT. Однако, точность обнаружения контента, созданного искусственным интеллектом, остается предметом спора, в том числе и потому, что OpenAI, разработчик GPT, официально признал, что не умеет выявлять генеративные тексты.

Turnitin – сервис, который активно критикуется, это роднит его с отечественным антиплагиатом, который РУ и ВУЗ одновременно. С учетом опыта иностранной демократии некоторые зарубежные студенты отказываются от отправки работ на проверку, утверждая, что требование проверки подразумевает наличие вины, а действие презумпции невиновности никто не отменял. В отечественном, российском праве, по мнению О. В. Бобкова и С. А. Давыдова, плагиат является самостоятельным гражданским правонарушением, хотя нормы гражданского права не содержат в себе его состава в том виде, в котором в уголовном законе закреплен состав плагиата как преступления. «Наше право» предусматривает презумпцию невиновности в уголовном праве и презумпцию виновности в гражданском – это предположение, которое считается истинным до тех пор, пока ложность такого предположения не будет бесспорно доказана. Применительно к нашей теме, доказательством наличия или отсутствия плагиата выступает отчет сервиса по проверке работ на плагиат, который считается «первой, новационной, экспертной» системой, способной выносить заключение.

Некоторые зарубежные критики утверждают, что использование антиплагиат сервисов нарушает образовательную конфиденциальность и международное законодательство об интеллектуальной собственности, а также сохраняет работы авторов в частной базе данных антиплагиат сервиса для коммерческих целей.

Несмотря на все спорные вопросы, системы поиска плагиата процветают, выполняя десятки миллионов проверок. В марте 2019 года, Advance Publications приобрела Turnitin за 1,75 миллиарда долларов. В 2021 году, Turnitin приобрела конкурирующую компанию-разработчика программного обеспечения Ouriginal, которая сама по себе является результатом слияния Urkund и PlagScan.

Как подчеркивает Т. В. Хованская, система «Антиплагиат» имеет много общего с обычной поисковой системой. Однако, её «уникальность» заключается в том, что она способна анализировать не отдельные фразы, а целый документ. Размер этого документа может варьироваться от нескольких абзацев до тысяч страниц и быть практически любого текстового формата. В результате обработки запроса, система выдает детальный отчет, где заимствованные фрагменты отмечаются тегами. Кроме того, в правой части отчета появляется панель со списком источников и расчетом «процента оригинальности». Эта панель с перечнем источников является интерактивным инструментом, облегчающим подробный анализ заимствованного контента после машинного анализа, рис. 2. Система «Антиплагиат» также предоставляет возможность вносить изменения в полный отчет и пересчитывать результаты. К основным функциям редактирования отчета относятся отключение заимствованных источников, исключение отдельных заимствованных блоков и изменение типа источника с заимствованного на цитируемый и наоборот.

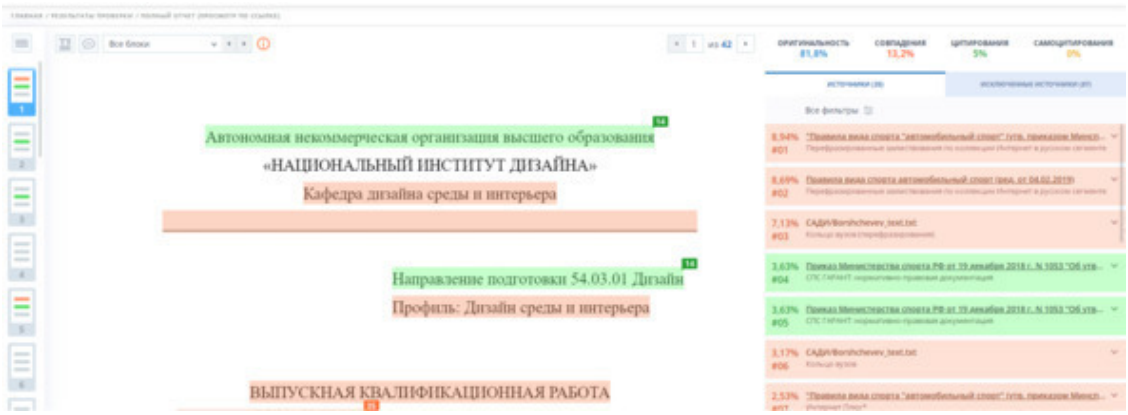


Рисунок 2 – Пример интерфейса с антиплагиат. ВУЗ-отчетом

Уточняя понимание Т. В. Хованской, отметим, что антиплагиат, в основе, является поисковой системой, поисковая система учитывает весь контекст, как целостность, отличие антиплагиата, в специализированной форме выдачи поискового результата.

В своей статье, И. Б. Стрелкова обращает внимание на то, что в 2005 году в Российской Федерации была создана специализированная поисковая система «Антиплагиат». Эта система была разработана с целью обнаружения текстовых заимствований и оценки их корректности. По мнению И. Б. Стрелковой, эту систему не следует рассматривать как «коммерческий продукт, навязанный научному и университетскому сообществу через искусственно созданный спрос». По ее мнению, антиплагиат, как явление, является необходимым элементом логики текущего этапа развития науки, на котором условия существования науки радикально трансформировались.

В контексте оценки системы «Антиплагиат», ее достоинств и недостатков, И. Б. Стрелкова приводит высказывание Н. И. Мартишиной: «Учитывая складывающуюся в стране ситуацию, использовать какую-либо систему проверки на плагиат (даже со всеми ее недостатками) – необходимо».

Ниже рассмотрим, как работает отечественная система антиплагиата, но не с точки зрения гуманитариев, а с точки зрения специалистов. Показатель уникальности текста представляет собой количественную меру, обычно выраженную в процентах. Как показывает практика, этот показатель вызывает интерес как у авторов, так и у их преподавателей. Данный показатель рассчитывается системами антиплагиата с помощью специально разработанных алгоритмов, которые держатся в строжайшем секрете и которые анализируют и исследуют текстовую комбинацию, отправленную на проверку. В простейшей интерпретации возможен следующий подход к толкованию процентов. Весь текст работы = 100%, если в работе 75 тыс. знаков, то 1% текста = 750 знакам. Если отчет о проверке показал 40% заимствований, то это означает, что работа содержит 40% * 750 зн. в 1% = 30 тыс. знаков текста плагиата.

Написание уникального текста, будь то реферат, курсовая или ВКР, может показаться на первый взгляд несложным процессом: грамотный подбор синонимов, перестановка слов и словосочетаний – и работа уникальна и готова к сдаче! Однако на практике оказывается, что реальность не так проста. Процесс проверки текста в системе антиплагиата проходит несколько этапов:

– Первичная обработка текста – удаление стоп-слов, знаков препинания и неалфавитных символов. Что может относится к стоп-словам: предлоги, а также слова не несущие смысловой нагрузки – без, быть, все, вы, для, его, есть, или, как, когда, кто, меня, мне, так, там, уже, чего, что, чтобы, этой, этом, этот, анализ, исследование, метод, результаты, данные, модель, проблема, теория, гипотеза, область, фактор, переменная, эффект, показатель, доказательство, вывод и т. д.

– Лемматизация – приведение слов к их нормальной форме, например, фраза: «Учитывая складывающуюся в стране ситуацию, использовать какую-либо систему проверки на плагиат (даже со всеми ее недостатками) – необходимо», примет вид «Учитывать складываться страна ситуация, использовать система проверка на плагиат (недостаток) – необходимо».

– Хеширование слов – каждое лемматизированное слово хешируется (предобразование в выходную битовую строку установленной длины) с помощью хеш-функции для унификации длины и упрощения сортировки.

– Формирование хеша шинглов (словосочетания из n-слов, n-граммы, рис. 3) – последовательность хешей слов представляется в виде значений, то есть перекрывающихся последовательностей из n хешей заданной длины шингла. Как правило, первично, используется шингл, состоящий из двух слов.



Рисунок 3 – Шинглы, состоящие из двух слов

– Хеширование значений шинглов.

– Запись хеша последовательностей с идентификатором текста и местоположением в тексте записывается в специальный файл, называемый индексом. Значения записываются в отсортированном порядке, что позволяет осуществлять двоичный поиск в индексном файле.

Из анализа алгоритма проверки уникальности текстов сервисом антиплагиат можно сделать следующие выводы:

– проверкой учитываются слова, несущие смысловую нагрузку. Введение в текст водянистых терминов – «по нашему мнению», «исходя из анализа», «подводя итог» и т. д. уникальности не прибавляет;

– изменение шинглов из двух слов (биграмм) приводит к повышению уникальности текстов, так как изменяется хеш шинглов;

– уникальность текста нужно повышать не фрагментарно, а целостно, так как это изменяет общее хеш-значение.

С учётом особенности процесса антиплагиат-проверки для создания качественного уникального научного текста используются три основных приёма:

– Обычный (легкий, поверхностный) рерайтинг. Для антиплагиат-сервисов, как правило, он не подходит, так как достигаемое отличие текстов (при сверке по шинглам из 2 слов – Ш2) не даёт возможности преодолеть порог срабатывания модуля перефразирования, преодоление которого возможно при показателе отличия текстов Ш2 = 80% и более процентов. Примером обычного рерайта является однократный перевод текста на иностранный язык и обратно на русский, при этом показатель Ш2 лежит в пределах 50—80%.

– Глубокий рерайтинг, текст «пересказывается» с использованием уникальных текстовых комбинаций, что существенно больше, чем перестановка предложений и замена синонимов.

Глубокий рерайт даёт отличие текстов по методу сверки Ш2 на уровне 80 и более процентов. Пример глубокого рерайта – последовательный перевод на венгерский-русский + финский-русский + китайский-русский. При выборе языков желательно ориентироваться на объем искажений, получаемых в процессе перевода, наибольшее отличие рерайта от текста донора дадут лексически отдаленные языки, рис. 4, например финский.

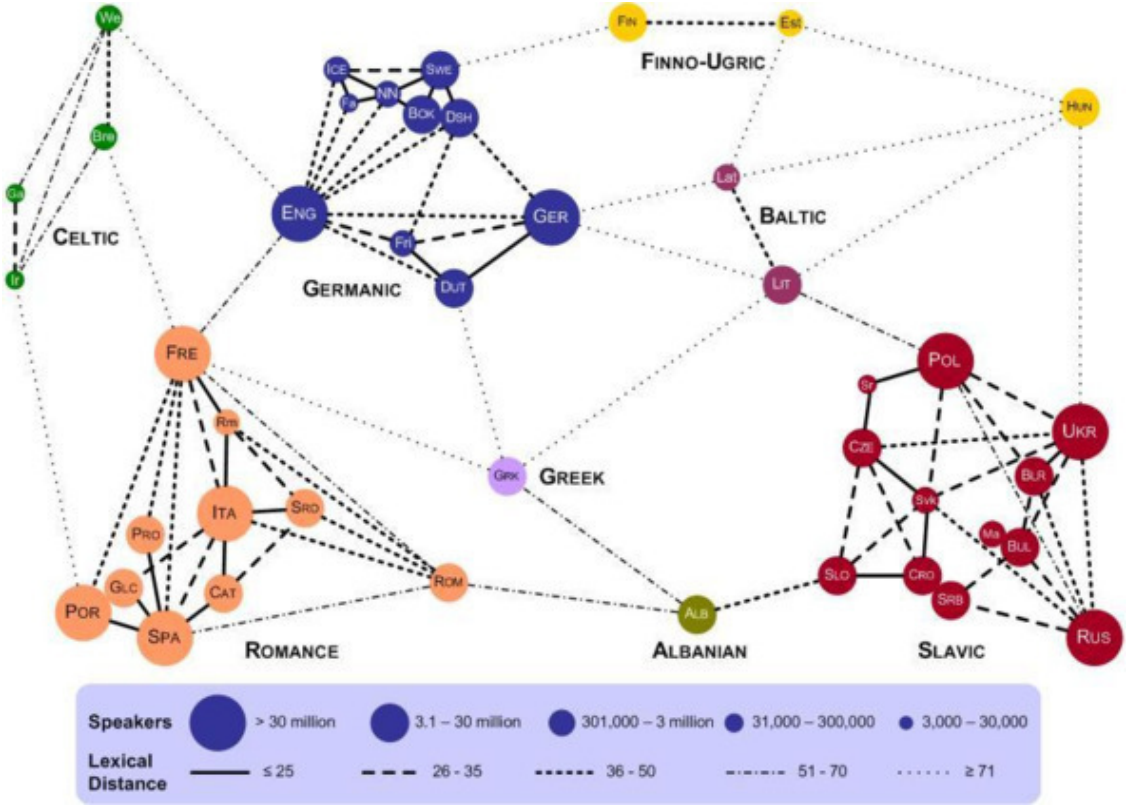


Рисунок 4 – Лексическое расстояние между европейскими языками

– Копирайтинг, оригинальный материал пишется «из головы», на основе глубокого понимания ранее полученного материала. Текст после копирайтинга отличается от источника при сверке по показателю Ш2 более чем на 80%, однако, нужно быть готовым к тому, что уникальность будет как у обычного или глубокого рерайтинга. Данная проблема, связанная с тем, что современные системы антиплагиат в погоне за «жесткостью» проверок давно уже грешат объективностью, а копирайтинг не предусматривает изменение n-грамм, которые относятся к терминам.

Рерайтерам и копирайтерам необходимо знать, что понятие уникальности материала по своей сути субъективно, его понимание зависит от множества внешних факторов. Каждый сервис антиплагиата по-своему понимает значение уникальности, поэтому во избежание кривотолков крайне важно исходить из потребности в проценте уникальности и учитывать, где проверяется текст:

Чтобы пройти проверку в StrikePlagiarism, достаточно получить отличия – Ш5 – 90—95%, Ш25 – 1—2%. Такое отличие дает перевод текста: венгерский-русский + китайский-русский.

Успешно пройти проверку в антиплагиат ВУЗ, потребует достижения показателя Ш2 – 90 и более процентов, рис. 5. Такое отличие возможно получить с помощью длинного цепочного перевода, но дефекты текста потребуют существенных трудозатрат на его редактирование.

РуКонтекст – отличие при сверке по показателю Ш2 – ваш показатель, который вы получите при проверке.

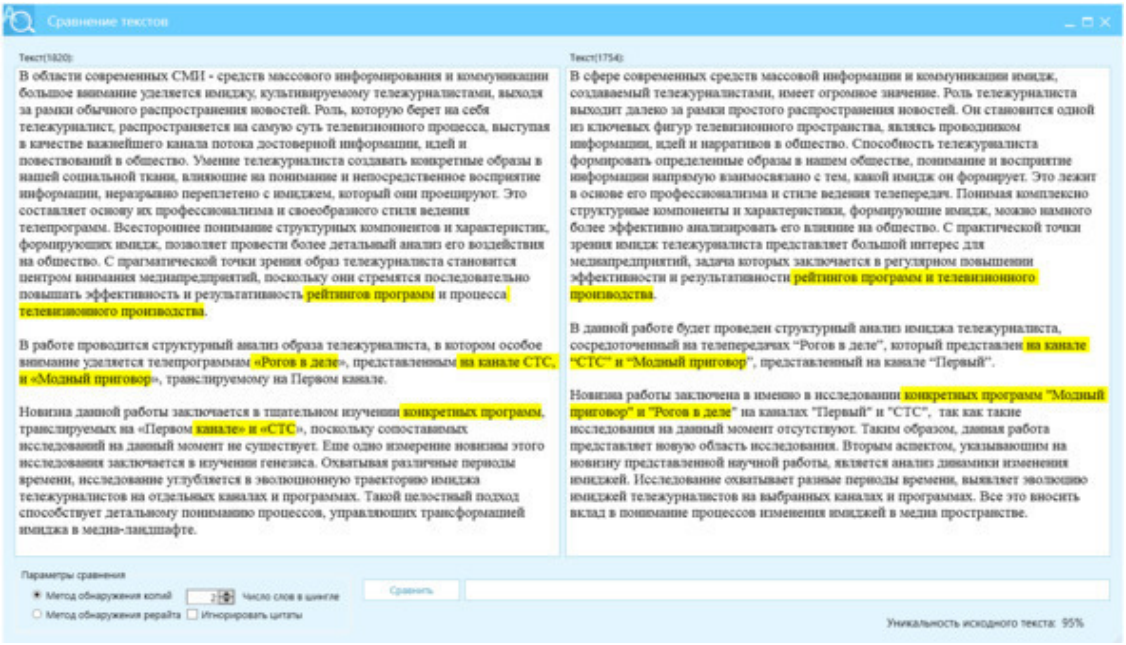


Рисунок 5 – Сверка текстов на отличие по показателю Ш2, левое окно – текст рерайта, правое окно – текст источник (донор), желтым выделены n-граммы, которые не изменились, показатель отличия см. внизу, справа = 95%

Что такое КонтрПлагиат? Как показано выше, рерайт (в частности, простой и глубокий) и копирайтинг не дают однозначного результата при проверке в антиплагиат, который РУ и ВУЗ. КонтрПлагиат свободен от этих недостатков, так как, применяя систему методов, он позволяет повышать уникальность с высоким отличием и позволяет проводить проверку текстов локально, на компьютере, бесплатно, за считанные секунды. Если показатель Ш2 больше 95%, то текст с высокой вероятностью пройдет любую проверку на антиплагиат в любой системе.

Разбивка шинглов (биграмм) в КонтрПлагиате – самостоятельный метод, его задача не оставлять в тексте желтых блоков, рис. 5, например вместо «Конституция Российской Федерации» – «Основной Закон нашей страны».

Диапазоны значений изменения текстов по показателю Ш2 могут варьироваться в достаточно большом диапазоне, что связано с объемом изменяемого текста. Если предложение содержит набор терминов, которые считаются неизменяемыми, то процесс перефразирования затрудняется и показатель Ш2 низок, например дан текст донор: «Государственную власть в РФ осуществляют Президент РФ, ФС (СФ и ГД), Правительство РФ, суды РФ. Государственную власть в субъектах РФ осуществляют образуемые ими органы государственной власти».

Имея текст с дефицитом изменяемой части, мы отходим от принципов рерайта и занимаемся чем-то иным, разбивая шинглы из 2 слов и придумывая слова и выражения, которые в обычной языковой практике, тем более научной не используются. Данный процесс называется КонтрПлагиатом. КонтрПлагиат может выглядеть следующим образом: «Функции органов, относимых к государственным, реализуют – высшее должностное лицо нашей страны – Президент, Собрание, на уровне Федерации, Совет, уровня Федерации и коллегиальный орган – Дума, государственного уровня, высший исполнительный орган Российской Федерации – Правительство, органы судебной власти РФ. Власть, уровня государства в субъектах,

территориях нашей страны, реализуют образуемые специализированные органы, относимые к власти государства».

Ваш текст с правильно расставленными запятыми выглядит следующим образом:

Таким образом, для цели настоящей работы, предлагаются следующие константы:

– рерайт, способ изменения текста донора методом перефразирования «другими словами» для цели достижения уникальности;

– шингл, n-грамма – последовательность слов, биграмма – 2 слова (шингл из 2 слов, Ш2), триграмма – 3 слова (шингл из 3 слов, Ш3);

– легкий, поверхностный рерайт – перефразирование текста с уникальностью до 80%, при сверке по шинглам, состоящим из 2 слов;

– глубокий рерайт – перефразирование текста с уникальностью более 90%, при сверке по шинглам, состоящим из 2 слов;

– КонтрПлагиат – система методов, применяемых для перефразирования текста с уникальностью более 95%, при сверке по шинглам, состоящим из 2 слов;

– мягкая антиплагиат проверка – процент, полученный при локальной сверке биграмм (шинглов из 2 слов), = проценту уникальности антиплагиат проверки (например, РуКонтекст, антиплагиат ру – бесплатная версия);

– жесткая антиплагиат проверка, при нормальной заспамленности, процент, полученный при локальной сверке биграмм (шинглов из 2 слов), = проценту уникальности антиплагиат проверки (например, антиплагиат. ВУЗ) – минус 10%. Приведем пример: локальная проверка отличия текстов показала отличие по показателю Ш2 = 97%, ожидаемый результат оригинальности в отчете антиплагиат ВУЗ 87%, рис. 6;

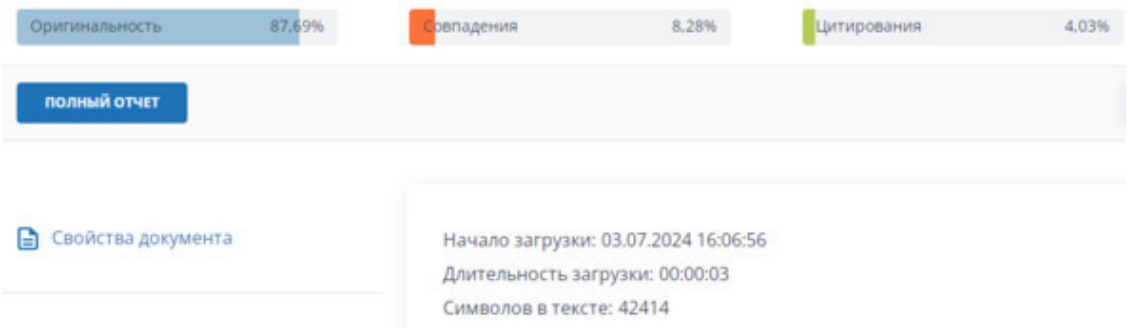


Рисунок 6 – Пример отчета в Антиплагиат, версии ВУЗ

– заспамленность, частотность текста – численная мера, показывающая популярность текста, например, в педагогических работах, посвященных младшему школьному возрасту, теоретическая часть содержит возрастную характеристику детей данного возраста, сколько работ о младшем школьном возрасте написано за последние 20 лет?

Таким образом, КонтрПлагиат – система принудительных мер, сочетающих в себе вариативность инструментов глубокого рерайта и копирайтинга, усиленных принудительным изменением шинглов (биграмм), дополненного методом сверки текста рерайта с текстом донора по показателю Ш2 (сверка шинглов, состоящих из 2 слов). КонтрПлагиат изменяет весь текст, в том числе и термины, при показателе Ш2 более 95% цель КонтрПлагиата достигается – работа набирает высокий показатель уникальности.

КонтрПлагиат применяется для успешного прохождения проверки в антиплагиат, версии ВУЗ, как правило для других систем поиска заимствований КонтрПлагиат не нужен, там достаточно глубокого рерайта.

1.2. Методы которые не работают в антиплагиат ВУЗ

Ниже обобщены методы, в которых присутствует как размытая, так и четка грань, связанная со студенческой «оптимизацией» рерайта, например методы некоторые способы демонстрируют элемент легальности, а использование технического кодирования, содержат состав мошеннических действий.

Изменение размера предложений и замена знаков препинания

Метод предполагает манипуляции с предложениями – простые объединить в сложные, сложные разделить на простые, метод – практически не влияющий на оригинальность текста, т.к. шинглы не изменяются. Максимальное повышение оригинальности, которое можно достичь с помощью этого метода, составляет всего 1—2%. Кроме того, такой подход требует значительных временных затрат. Применение данного метода в сочетании с рерайтом – эффективно, т.к. точка является своеобразным маркером, наравне с запятой.

Удаление абзацев и приведение текста к массиву на несколько страниц

Данный подход не изменяет шинглы, следовательно не является способом повышения уникальности. Изменяет общий хеш абзацев, но из наших пояснения выше явствует, что начальным анализом является анализ хеша шинглов а не хеша абзацев, поэтому изменение размера абзацев не влияет на уникальность при проверке в АП ВУЗ.

Принудительные и автоматические переносы

Принудительный перенос – постановка дефиса без учета требований русского языка (прогр-амма, пров-ерки, заимс-твован-ие). Метод может применятся в таблицах, на небольших участках текста до 500 знаков, с отключением функции проверки документа.

Автоматические переносы, хотя и представляют собой рабочую идею, дают результаты прибавки оригинальности лишь в небольшом объёме. Кроме того, в некоторых учебных заведениях запрещено использование автоматической расстановки переносов, это требуют соответствующие методические рекомендации.

Стоит отметить, что расстановка переносов в программе Word выполняется практически мгновенно, что позволяет воспользоваться этим методом без трудозатрат. Ещё лучший результат даёт копирование текста с переносами в блокнот, а затем вставка его обратно, в текст работы. Правая граница текста выравнивается с помощью пробелов между словами строки текста.

Мертворожденный метод – замена кириллицы на буквы другого алфавита (латиница, греческое письмо)

Когда то, лет 20 тому назад АП ВУЗ общался со своими клиентами на равных. Именно тогда нами был предложен в чате метод замены кирилистической буквы «о» на греческий, похожий символ. Это давало 100% уникальность в бесплатной версии проверки на плагиат.

Разработчики предложение не прокомментировали, но метод перестал работать в течении недели.

Замена букв русского языка на буквы других языков, таких как латиница или греческое письмо, все еще используется студентами, этим методом грешат зарубежные студенты, которых у нас становится все больше. Метод эффективен только при использовании в очень небольших объемах, например, в таблицах.

Следует отметить, что преподаватель может заметить нетипичные буквы, так как они отличаются от обычного написания. Если использовать этот метод в большем объеме, антипла-

гиат может выявить его использование и присвоить статус подозрительного документа. С учетом лишь 1—3 процентов повышения оригинальности, нет смысла тратить свои временные ресурсы и репутацию на этот метод.

Невидимый скрытый текст (белый или уменьшенный, или за пределами документа)

Скрытый текст использовался в прошлом в различных вариантах. Например, одним из способов было вставлять абзац размером несколько строк и делать его белым, а также уменьшать размер шрифта до единицы. Данные методы дискредитируют проверяемый файл, и вы получите сообщение о подозрительном документе – наличие вставок и т. д.

Другой подход заключался в создании надписей, в которых содержался текст, а сами надписи перемещались за пределы документа. Сегодня использование такого метода считается неприемлемым, т.к. антиплагиат ВУЗ все это подметит и свое слово в отчете скажет, выделив его красным транспарантом. Впрочем, насколько можно судить, некоторые методы вставок все еще пользуются популярностью и относятся к методам кодирования.

Вставка знаков, символов отличных от текста или кириллицы

Техническое кодирование, именуется по-разному, знаки в работе, символы и т. д. Метод технического кодирования применяет разные подходы, его цель – обман скрипта проверки системы антиплагиат. Любой метод технического «повышения оригинальности» текстов, с течением времени, становится известен и пресекается системами проверок, проверяемому документу присваивается статус подозрительного документа.

Практически все системы антиплагиата выявляют кодирование, да и сам проверяющий может наглядно увидеть элементы кодировки, см. рис. 7.

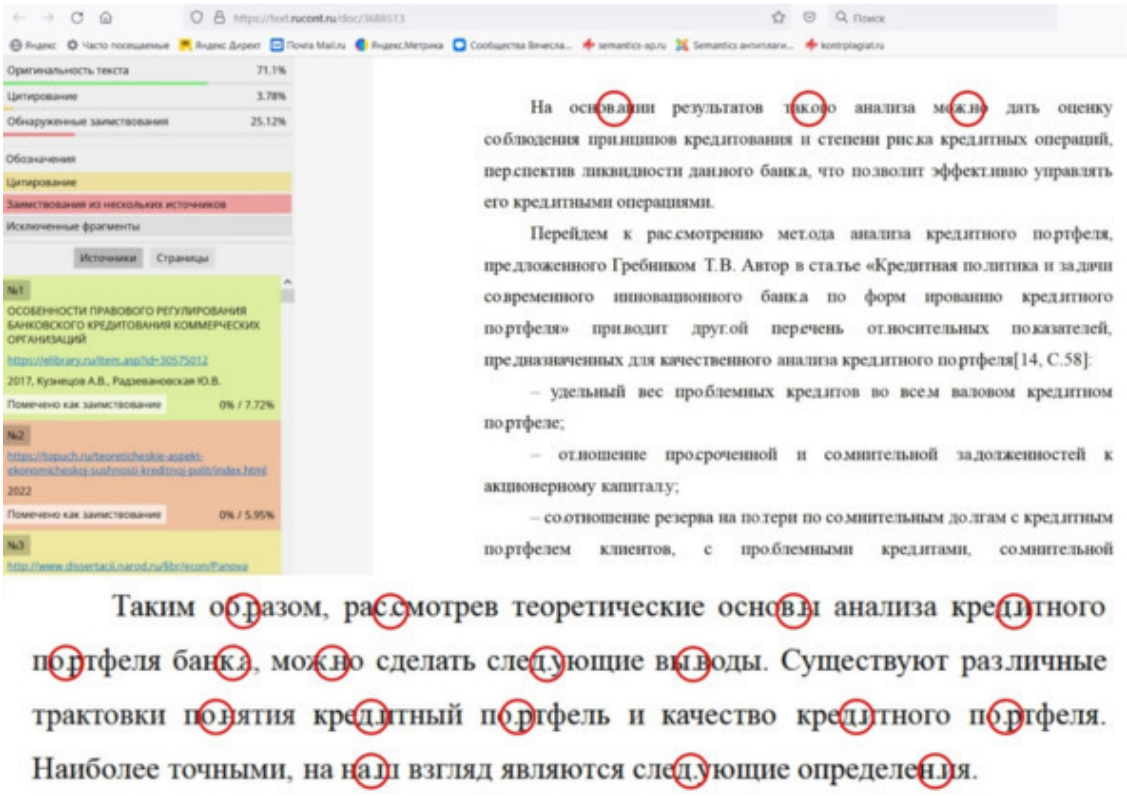


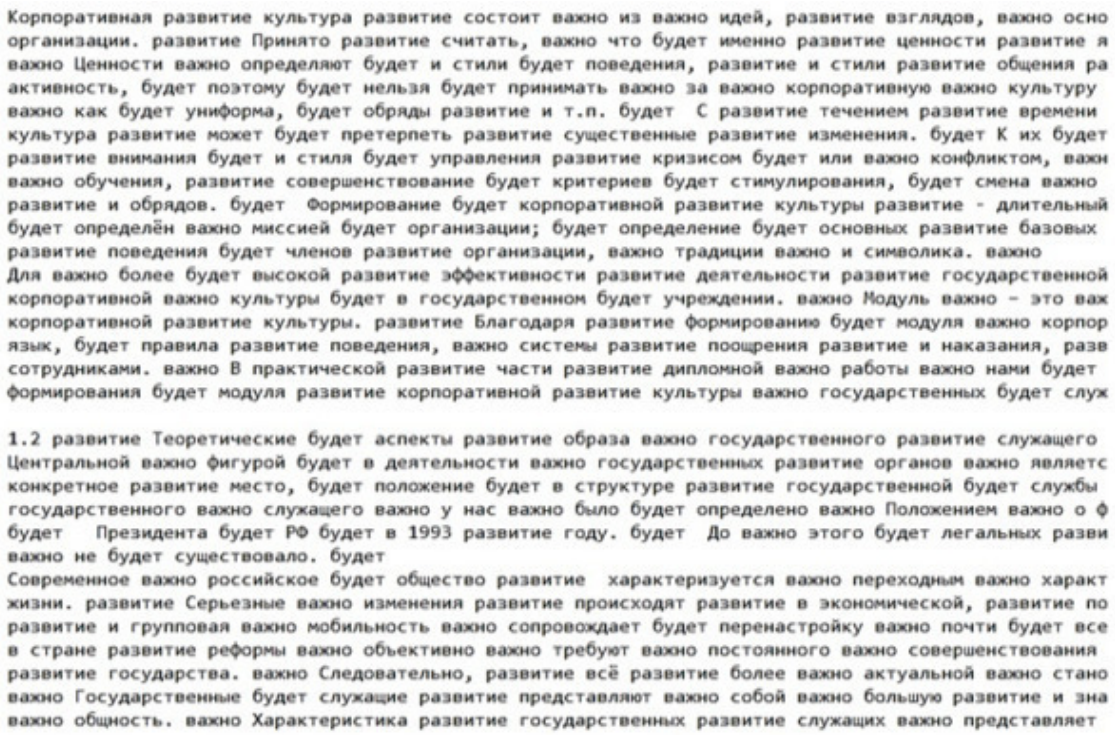
Рисунок 7 – Пример технического повышения оригинальности текстов, кодирование текста точками

Итак, все еще находящим применение методом для повышения оригинальности текста является вставка в слова невидимых в WORD символов. Метод называется кодированием, его преследуют все ВУЗы. Например, вместо слова «покакзатель» можно написать «пока затель». Если не злоупотреблять такими словами, это повышает оригинальность текста.

Техническое кодирование текста на сайтах

Знаки, символы, кодирование, к которому прибегают если осталось совсем немного времени до сдачи работы особенно опасно, рисунки выше это демонстрируют.

Кодирование можно убрать в блокноте, или, если применено перемешивание слов, рис. 8., путем распознавания сконвертированного в PDF документа.


Корпоративная развитие культура развитие состоит важно из важно идей, развитие взглядов, важно осно
организации. развитие Принято развитие считать, важно что будет именно развитие ценности развитие я
важно Ценности важно определяют будет и стили будет поведения, развитие и стили развитие общения ра
активность, будет поэтому будет нельзя будет принимать важно за важно корпоративную важно культуру
важно как будет униформа, будет обряды развитие и т.п. будет С развитие течением развитие времени
культура развитие может будет претерпеть развитие существенные развитие изменения. будет К их будет
развитие внимания будет и стиля будет управления развитие кризисом будет или важно конфликтом, важн
важно обучения, развитие совершенствование будет критериев будет стимулирования, будет смена важно
развитие и обрядов. будет Формирование будет корпоративной развитие культуры развитие - длительный
будет определён важно миссией будет организации; будет определение будет основных развитие базовых
развитие поведения будет членов развитие организации, важно традиции важно и символика. важно
Для важно более будет высокой развитие эффективности развитие деятельности развитие государственной
корпоративной важно культуры будет в государственном будет учреждении. важно Модуль важно – это вах
корпоративной развитие культуры. развитие Благодаря развитие формированию будет модуля важно корпор
язык, будет правила развитие поведения, важно системы развитие поощрения развитие и наказания, разв
сотрудниками. важно В практической развитие части развитие дипломной важно работы важно нами будет
формирования будет модуля развитие корпоративной развитие культуры важно государственных будет служ

1.2 развитие Теоретические будет аспекты развитие образа важно государственного развитие служащего
Центральной важно фигурой будет в деятельности важно государственных развитие органов важно являетс
конкретное развитие место, будет положение будет в структуре развитие государственной будет службы
государственного важно служащего важно у нас важно было будет определено важно Положением важно о ф
будет Президента будет РФ будет в 1993 развитие году. будет До важно этого будет легальных разви
важно не будет существовало. будет
Современное важно российское будет общество развитие характеризуется важно переходным важно характ
жизни. развитие Серьезные важно изменения развитие происходят развитие в экономической, развитие по
развитие и групповая важно мобильность важно сопровождает будет перенастройку важно почти будет все
в стране развитие реформы важно объективно важно требуют важно постоянного важно совершенствования
развитие государства. важно Следовательно, развитие всё развитие более важно актуальной важно стано
важно Государственные будет служащие развитие представляют важно собой важно большую развитие и зна
важно общность. важно Характеристика развитие государственных развитие служащих важно представляет

Рисунок 8 – Кодирование методом перемешивания слов (метод не повышает уникальность при проверке в АП ВУЗ)

Если вы попадете на фоновое обновление ПО антиплагиата, техническое кодирование будет обнаружено, и вам придется потратить еще больше времени на переписывание работы. Кроме того, кодирование всегда можно обнаружить простым способом, скопировав чистый текст в обычный блокнот, где будут видны невидимые символы, перемешивание слов или пропуски букв и слов.

Мы не рекомендуем использовать этот метод никогда, хотя с экономической точки зрения это может показаться оправданным и позволяет быстро повысить оригинальность документа хоть до 100%.

Обработка с помощью макросов

Переработка текста с помощью макросов фактически является методом технического кодирования. Разница заключается в том, что кодирование обычно выполняется сервисами, в то время как макросы могут быть скачаны и запущены на компьютере в среде Word.

Возможно, вам повезет, и вы найдете рабочий макрос с надежным способом обхода антиплагиат, однако, с учетом отечественного рынка рерайта академических работ, все, что работает приносит доход, поэтому в интернет нет рабочих макросов.

Любой внимательный научрук обнаружит попытку обхода за несколько минут. Поэтому наш совет – не стоит использовать такие методы для повышения оригинальности работы и не стоит тратить время на поиск граального макроса.

Разное содержание работы для проверки и в реальности

Различные варианты содержания работы для проверки и в реальности представляют собой явно неприемлемый и морально неэтичный подход к повышению оригинальности, который следует исключать во всех случаях.

Такой подход позволяет успешно пройти проверку на Антиплагиат, но скорее всего несоответствие будет обнаружено достаточно быстро, т.к. помимо научного руководителя за качество студенческих работ отвечают назначенные специалисты, которые просматривают работы и отчеты АП ВУЗ достаточно дотошно. В лучшем случае, придется представить другую работу для проверки. В худшем случае, это может привести к наказаниям, вплоть до отчисления.

Специализированный софт для быстрого и «качественного» повышения уникальности текстов

Такого софта не существует, а в ближайшее время не появится в силу того, что современные, бесплатные онлайн-переводчики способны дать достойный результат по показателю Ш2=70—85, а нейросети могут повысить уникальность до 100%. Сравните объемы инвестиций в эти сервисы, например в GPT. Указанный сервис бесплатен в версиях GPT-4o, GPT-3.5, а в ближайшие месяцы бесплатным станет и GPT-4. В чем смысл покупки сомнительно работающих программ?

Добавление ошибок (слияние слов, пробелы в словах, опечатки)

Насыщение текста ошибками, редко применяемая практика, но на самом деле все еще является популярным способом «повышения» уникальности текста. Аксиома проста, чем больше ошибок в тексте, тем выше его уникальность.

Следует помнить, что использование большого количества ошибок не рекомендуется – не более 1—2 опечаток на страницу. Если текст будет подчеркнут красным, возможно он и прибавит проценты, но точно не пройдет проверку глазами человека.

Проверка текстов в антиплагиат. ВУЗ перед вставкой в работу

Логично предположить, что если собрать в файл несколько тысяч листов текста, то проверка покажет, что является плагиатом, а что является уникальным текстом.

Метод имеет право на жизнь, с учетом специфики:

– антиплагиат учитывает совокупность, а не фрагментарность, поэтому нет 100% гарантии, что после включения «уникального» текста в работу он не станет плагиатом;

– накидать в файл проверки дипломные и курсовые работы можно, но как правило все эти работы уже отметились в индексе АП ВУЗ, здесь может повезти, вы найдете свежую работу, которая еще не ушла в кольцо ВУЗов.

Выбор редких тем работ

Решение выбрать сложную и непопулярную тему не дает предсказуемого результата, что связано с дефицитом материалов. Возможно текст, посвященный такой теме, обладает высокой степенью оригинальности, однако необходимость в длительном поиске и анализе материалов может значительно затруднить процесс написания работы.

Существует вероятность, что часть необходимых исходных данных будет недоступна или сложна для обработки. Поэтому, у нас нет однозначной рекомендации что лучше, выбрать сложную тему, по которой нет материалов или выбрать тему, по которой масса материала, но он уже использован не одной тысячей авторов. Примером заспамленной темы является возрастная характеристика детей, как правило младшего, дошкольного возраста, ее содержит почти каждая педагогическая работа.

Таким образом, в миллиардно базе индекса АП ВУЗ есть все, даже самые экзотичные тексты, экзотичных тем, при выборе темы исследования стоит лишь один вопрос, насколько высокочастотной является тема. Если ее писали миллионы раз, то вам может понадобиться добиваться отличия текста рерайта от источника на уровне Ш2 = 100%.

2. Википедия перефразирования для антиплагиат ВУЗ: работающие (Р) и частично работающие методы (ЧР)

ЧР – Лёгкий, поверхностный рерайт и перефразирование

Поверхностный, легкий рерайт представляет собой упрощённый метод модификации текста, путем перефразирования, который заключается в замене слов и фраз синонимами, изменении грамматических конструкций и перефразировании предложений. Показатель Ш2 для легкого рерайта находится на уровне до 80%, при сверке по шинглам, состоящим из двух слов, рис. 9.

Поверхностный рерайт можно делать используя метод пересказа, или, используя приложение 1 разбивать n-граммы текста авторским контентом. Ниже приведена иллюстрация процесса автоматизации этого метода.

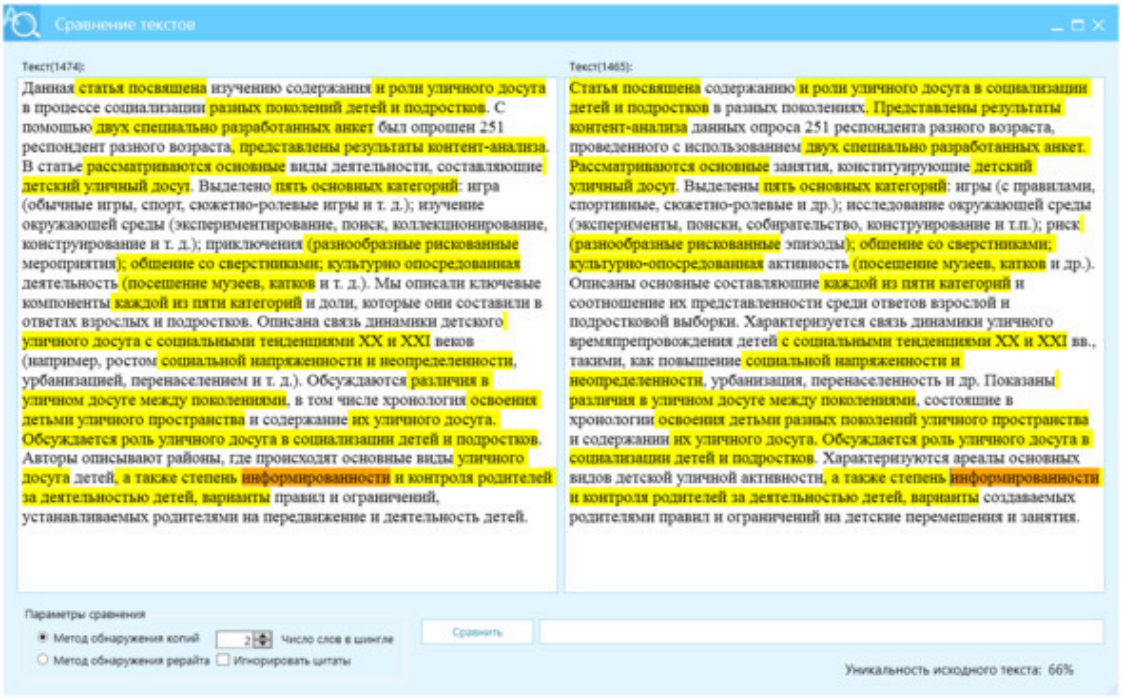


Рисунок 9 – Пример сверки двух текстов по методу Ш2, левое окно, текст переведенный на китайский и затем на русский, правое окно – текст источник, показатель отличия см. внизу, справа

Лёгкий рерайт не требует глубоких знаний стилистики и лингвистики, что делает его доступным для широкого круга пользователей. Поверхностное перефразирование может быть достаточно эффективным подходом для повышения оригинальности текста в системах проверки заимствований, основанных на элементарном сравнении слов и фраз (РуКонтекст, антиплагиат ру, бесплатная версия).

Антиплагиат ВУЗ использует более сложные алгоритмы анализа текста, основанные на методе Ш2, поэтому легкий рерайт не преодолеет порог срабатывания модуля поиска перефразирования и модуля поиска рерайта, который лежит на границе выше 80%.

Ниже приведены сравнения эффективности двух онлайн-переводчиков, примененных для поверхностного рерайта текстов, сверка отличия осуществлена по параметру Ш2 (сверка шинглов, состоящих из двух слов), табл. 1.

Таблица 1 – Показатель отличия текстов Ш2 для онлайн-переводчиков DEEPL и translate.google

Язык перевода	Отличие текстов по показателю Ш2 DEEPL, %	Отличие текстов по показателю Ш2 translate.google, %	Абсолютное отклонение: translate.google – DEEPL
английский (британский)	35	44	9
болгарский	47	55	8
венгерский	63	58	-5
испанский	49	46	-3
итальянский	50	48	-2
китайский (традиционный)	66	68	2
китайский (упрощенный)	65	67	2
латышский	57	56	-1
литовский	56	50	-6
немецкий	49	57	8
польский	50	59	9
португальский	56	56	0
румынский	55	51	-4
словацкий	57	59	2
словенский	53	54	1
украинский	10	13	3
финский	65	72	7
французский	56	51	-5
чешский	50	54	4
эстонский	64	64	0
японский	63	62	-1
	1 116	1 144	9

Из таблицы 1 видно, что большее отличие текстов, по сумме балов, достигает translate.google, абсолютными лидерами, дающими наибольшее отличие выступают: финский, китайский (традиционный), китайский (упрощенный), эстонский, японский языки.

Переводчик DEEPL демонстрирует не плохие результаты отличие текста по показателю Ш2 на следующих языках: китайский (традиционный), китайский (упрощенный), финский, эстонский, венгерский. Как заверяют разработчики DEEPL, переводчик самый точный в мире.

Производительность – в среде файлового перевода translate. google – 100 тыс. знаков в минуту, в интерфейсе – 10—15 тыс. в минуту. В интерфейсе DEEPL, 50 тыс. знаков в минуту.

В целом, для поверхностного рерайта можно рекомендовать финский, китайский и эстонский языки. Уровень поверхностного рерайта лежит в границах менее 80% отличия текстов по показателю Ш2.

ЧР – Глубокий рерайт и перефразирование

Глубокое перефразирование текстов актуально для всех сервисов проверки на уникальность, несмотря на значительную трудозатратность. Основная задача такого процесса заключается в систематичном внедрении новых слов после второго-четвёртого слова или замене существующих 2, 3, 4 слов. Что и как менять, какой это дает результат, показывает Приложение 1. Аксакалы традиционных школ рекомендуют текст прочитать и напечатать его своими словами по памяти, пересказать, другими словами. Метод имеет право на жизнь, хотя при наличии в доме пылесоса редкий эстет согласиться подметать пол зубной щеткой.

Глубокий рерайт подразумевает отличие текстов Ш2 на уровне более 80%. Исходя из таблицы 1, справедливо задаться вопросом, как онлайн-переводчики способны помочь в глубоком перефразировании. Обратимся к DEEPL и выполним цепочный перевод – русский – китайский (традиционный) – финский – эстонский – венгерский, это дает показатель отличия Ш2 – 85%.

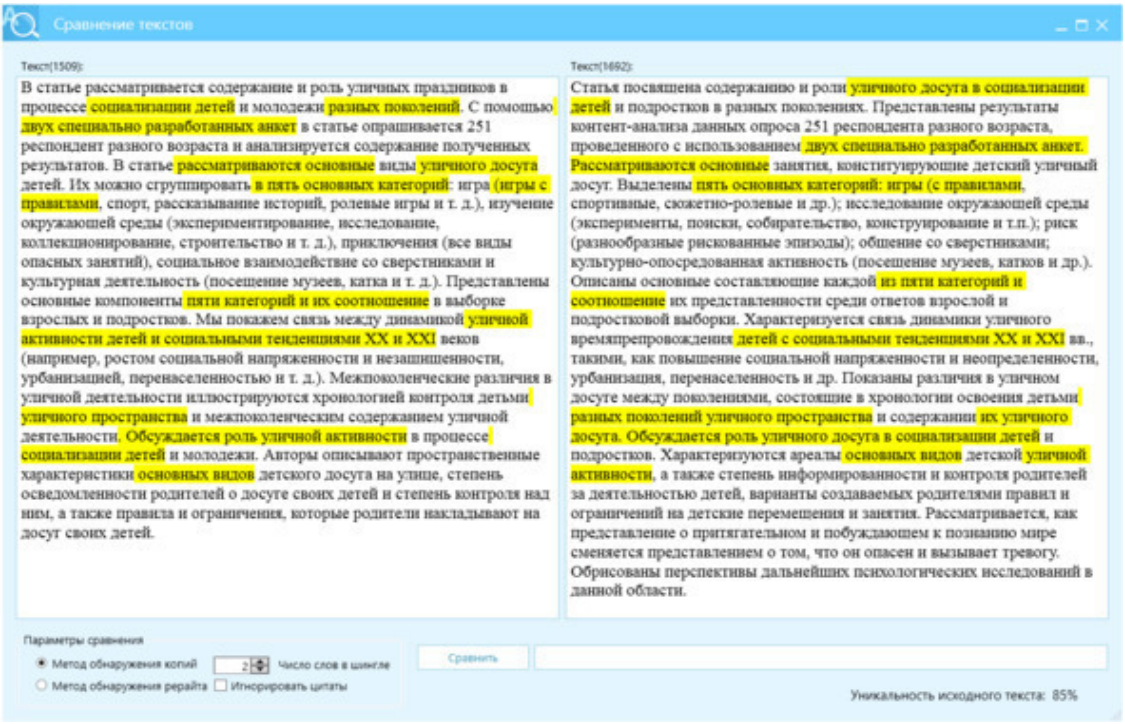


Рисунок 10 – Цепочный перевод – русский – китайский (традиционный) – финский – эстонский – венгерский, это дает показатель отличия Ш2 – 85%

Как видно из рисунка 10, текст рерайта требуют вычитывания, так как содержит неточности, например, текст рерайта содержит фразу – «уличных праздников», вместо «уличного досуга». В целом текст читабелен и не составит особого труда его подправить.

В следующем эксперименте выполним круговой перевод – русский – китайский (традиционный) – русский – финский – русский – эстонский – русский – венгерский – русский, рис. 11.

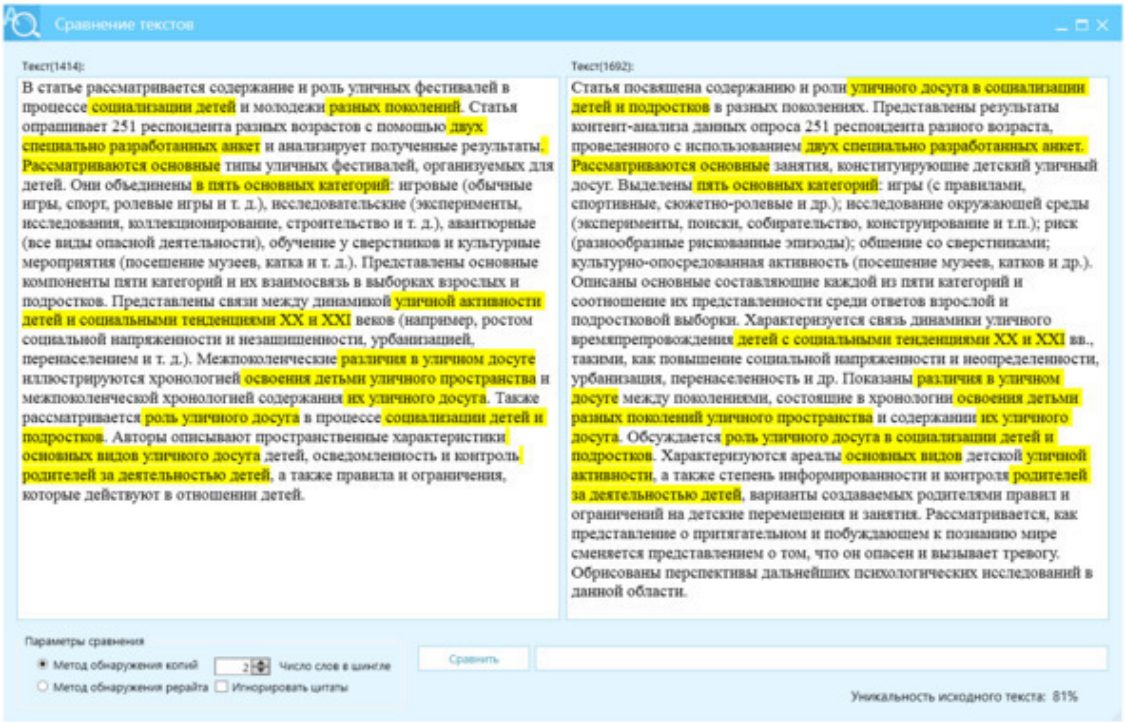


Рисунок 11 – Пример последовательного перевода, ин. язык-русский + ин. язык – русский…

Из рисунка 11 видно, что вместо «уличных праздников» появилась фраза «уличных фестивалей», текст обладает меньшим отличием, Ш2 = 81%, однако читабельность текста выше.

Производительность – в среде файлового перевода translate. google – 50 тыс. знаков в минуту, в интерфейсе – 5—10 тыс. в минуту. В интерфейсе DEEPL, 10 тыс. знаков в минуту.

Помимо переводчиков, сегодня только ленивый не пользуется услугами генеративных моделей ИИ-GPT. Протестируем результаты трех сервисов. В своем исследовании, с целью доказательности приведем скриншоты окна с текстом исходником, командой, ответом сервиса и результатом сравнения параметра Ш2.

ЧР Применение ChatGPT, Copilot, Gemini

Всем известен посыл маркетологов относительно перспектив рерайтеров, «они останутся без работы». Ниже мы проанализируем так ли это, протестировав три нейросети мирового уровня: ChatGPT, Copilot, Gemini.

Сходства протестированных ИИ заключено в следующем:

– Все три инструмента используют передовые языковые модели для обработки и генерации текста. ChatGPT и Copilot основаны на моделях GPT от OpenAI, а Gemini использует аналогичные технологии от Google.

– ChatGPT, Copilot и Gemini доступны через веб-интерфейсы и мобильные приложения, что делает их удобными для использования в любом месте и в любое время.

Как видно из анализа сходств, различие используемых генеративных моделей следующее:

– ChatGPT, известен своей универсальностью и креативностью. Он может генерировать текст на основе широкого спектра запросов и часто используется для создания контента, написания эссе и научных статей.

– Copilot, интегрирован в продукты Microsoft, такие как Office и Visual Studio, что делает его идеальным для пользователей этих платформ.

– Gemini, используя отличную генеративную модель является альтернативным выбором для пользователей, известен своей способностью помогать в создании статей и других текстов.

Согласно анализу качества текстов:

– Copilot – показал наивысшую точность в интерпретации данных и выполнении задач, связанных с анализом текста, другими словами, текст, достаточно часто, изменяется на недостаточно высокий процент, т.к. решение «боится отойти» от текста контекста.

– ChatGPT, отличается высокой степенью креативности, например вместо «полегших сортов пшеницы» мы можем получить фразу – «не прямостоячие сорта пшеницы».

– Gemini, удовлетворительно справляется с базовыми задачами рерайта, однако насыщает текст предельным количеством маркеров генеративности.

Как видно, ChatGPT, Copilot и Gemini предлагают определенные возможности для рерайта академических текстов, каждый инструмент имеет свои сильные и слабые стороны и особенности.

ChatGPT

GPT (Generative Pre-trained Transformer) – это серия моделей искусственного интеллекта, разработанных компанией OpenAI. Модели известны своей способностью генерировать текст, который, как считают разработчики, трудно отличить от написанного человеком. Наиболее известные версии включают GPT-3 и GPT-4. Эти модели обучены на огромных объемах текстовых данных и способны выполнять широкий спектр задач, от написания эссе до создания кода.

GPT используется в различных областях, включая создание контента, автоматизацию общения, перевод текстов, написание кода и многое другое. Модель также может использоваться для анализа данных и предоставления рекомендаций. GPT использует архитектуру трансформеров, что позволяет модели эффективно обрабатывать и генерировать текст.

Как правило пользователи GPT относится к генеративным возможностям с излишней наивностью, формирую примитивные инструкции, например «Выполни глубокий рерайт на русском языке», рис. 12, примитивный запрос порождает примитивный ответ с примитивным результатом, рис. 14.

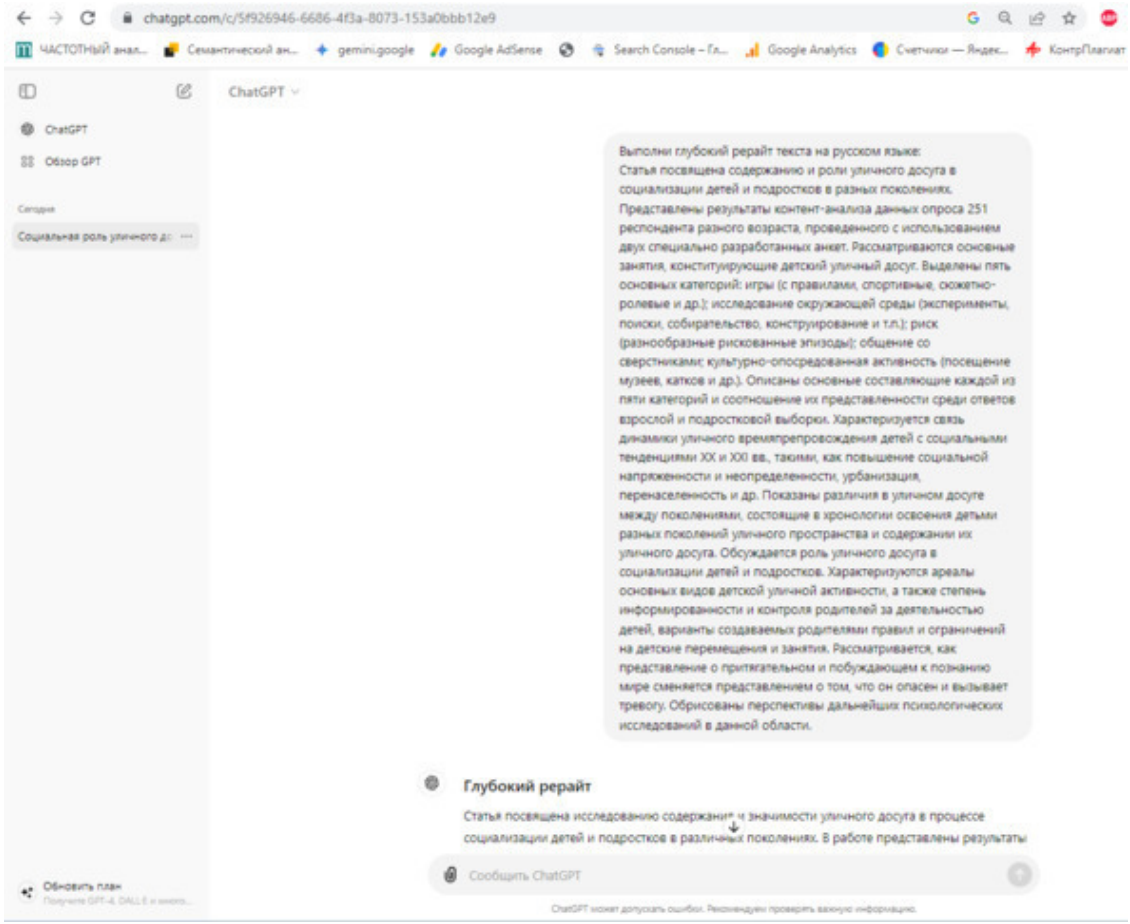


Рисунок 12 – Оправка текста на GPT-рерайт с инструкцией: выполни глубокий рерайт текста на русском языке

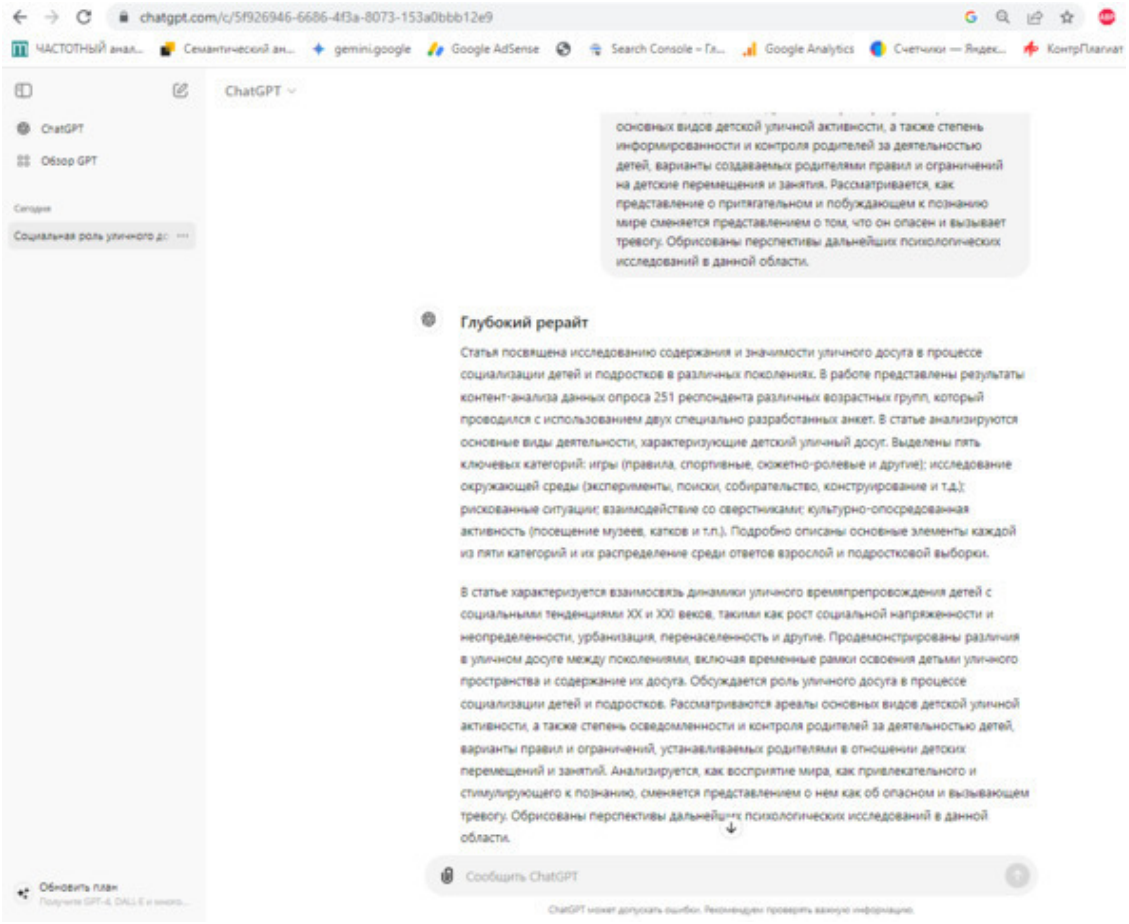


Рисунок 13 – Ответ ChatGPT

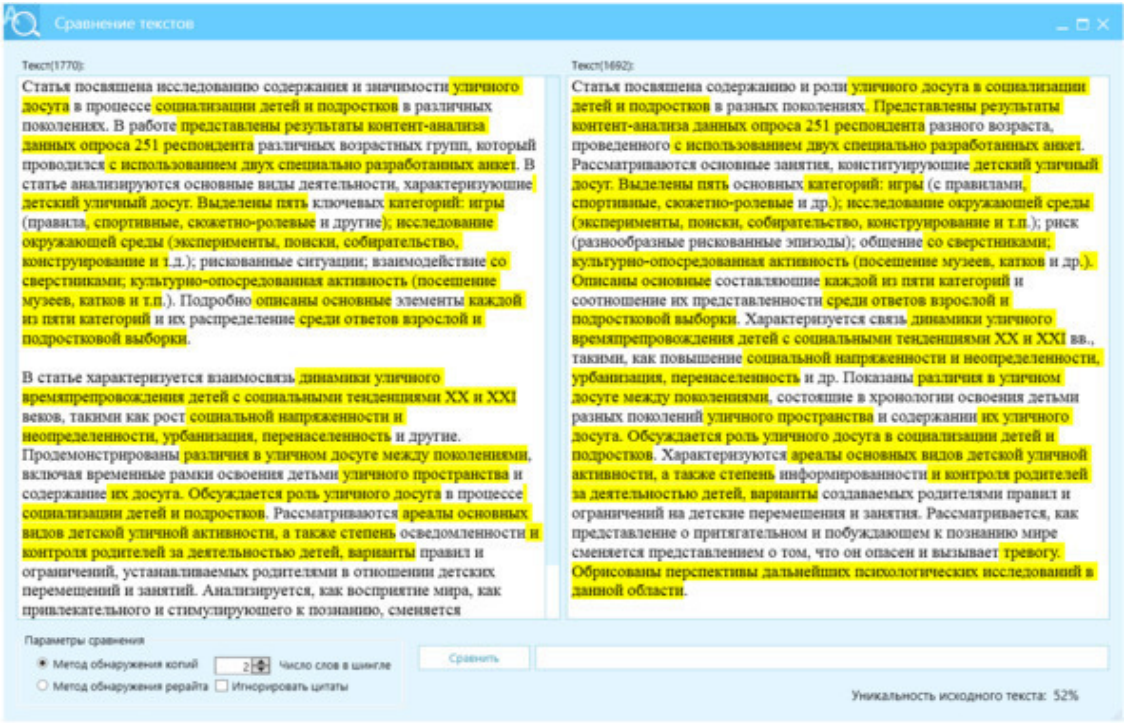


Рисунок 14 – Сверка контекста (правое окно) с GPT-рерайтом (левое окно) по методу Ш2=52

Как видно из рисунка 14, показатели «глубокого перефразирования» GPT находятся на достаточно низком уровне, Ш2 = 52%, в этом GPT проигрывает онлайн-переводчикам, которые показали, на китайском языке Ш2 = 68%, на финском 72%. Как видно из рисунка 14, текст практически не содержит неточностей, о которых мы упоминали, понятие «уличный досуг» изложено без ошибки.

Copilot

По мнению Microsoft, Copilot – это инновационный инструмент, разработанный для повышения продуктивности и улучшения взаимодействия с пользователями.

Copilot тесно интегрирован с приложениями Microsoft 365, такими как Word, Excel, PowerPoint и Outlook. Это позволяет пользователям получать интеллектуальные подсказки и автоматизировать рутинные задачи прямо в привычных приложениях.

Copilot использует модели искусственного интеллекта для анализа данных и предоставления рекомендаций. Copilot может автоматизировать множество задач, таких как создание отчетов, анализ данных и подготовка презентаций.

Проделаем аналогичные действия с ИИ Copilot, рис. 15—17.

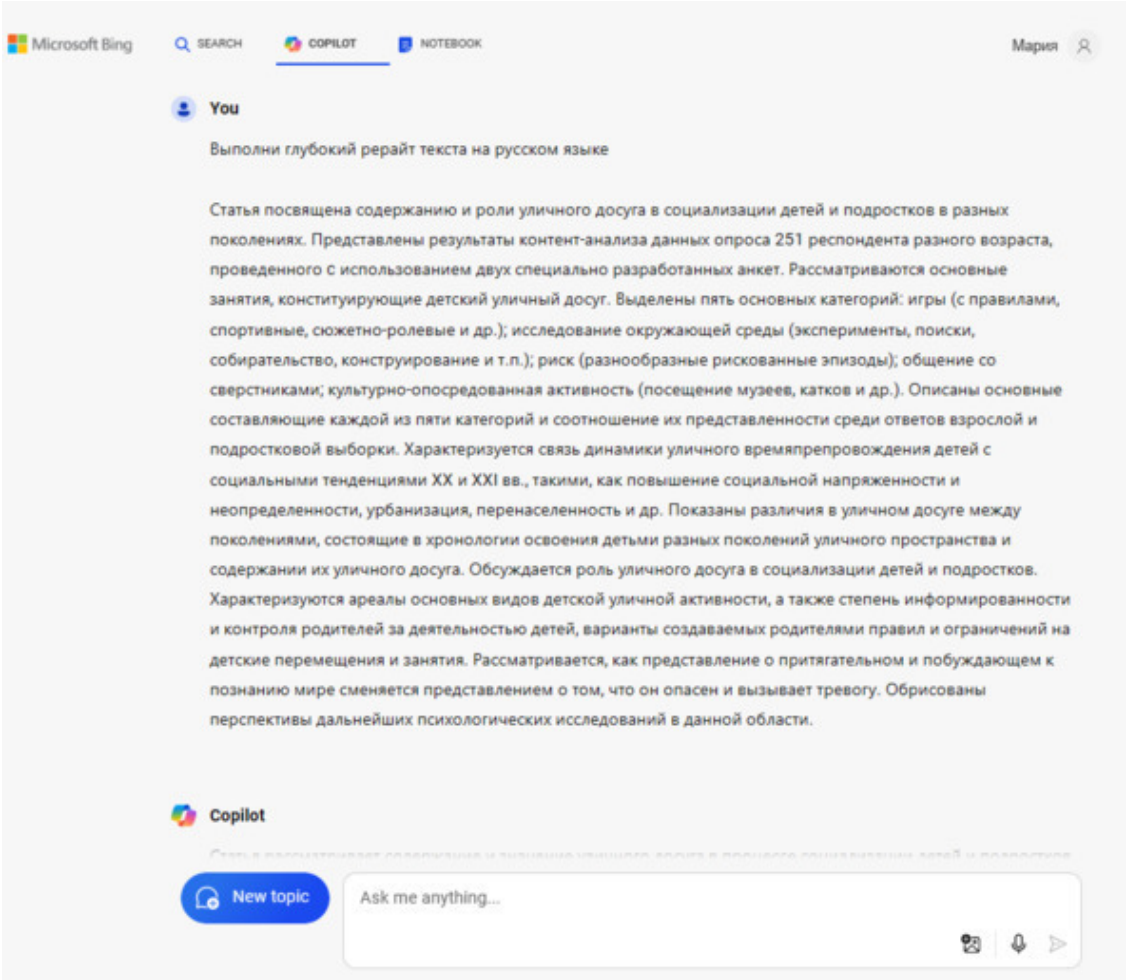


Рисунок 15 – Отправка инструкции и контекста в Copilot

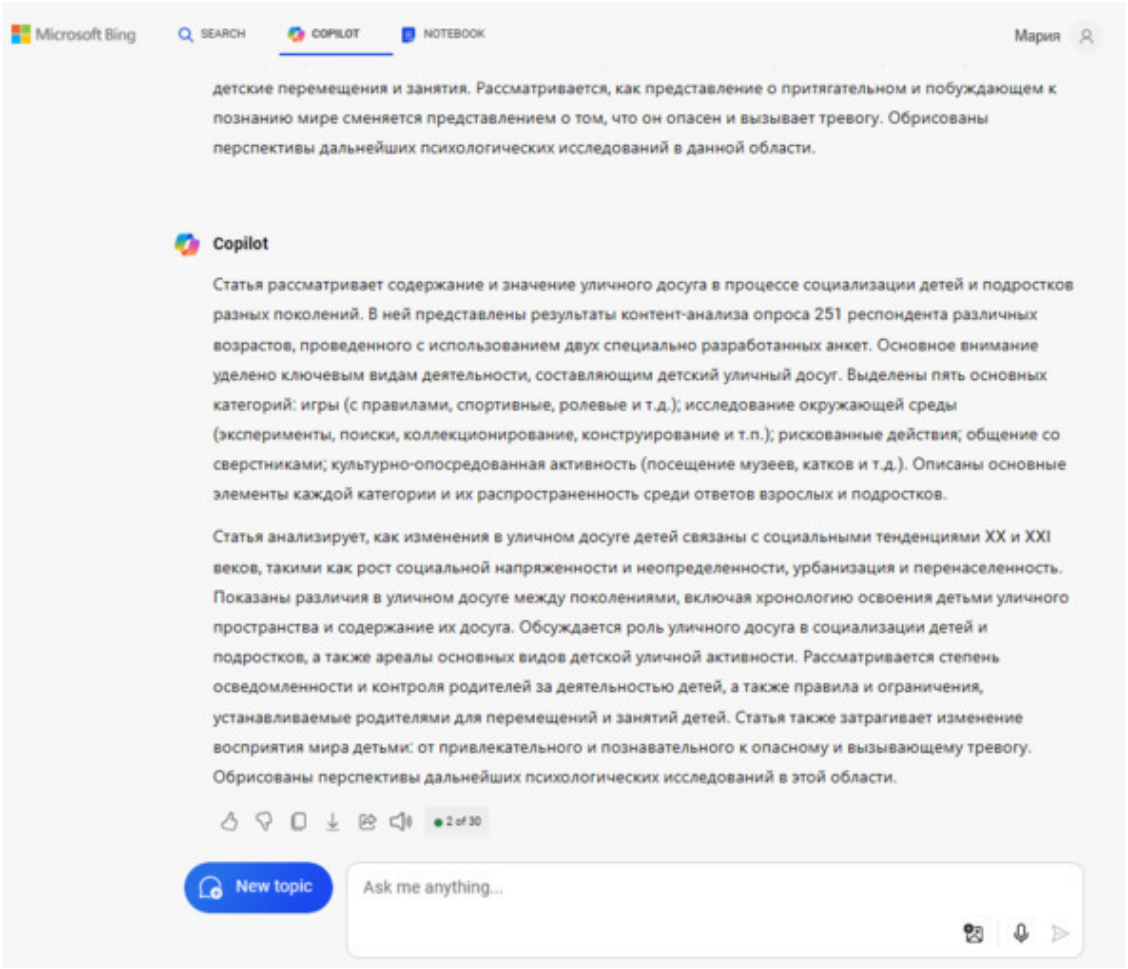


Рисунок 16 – Получение ответа, в виде текста рерайта

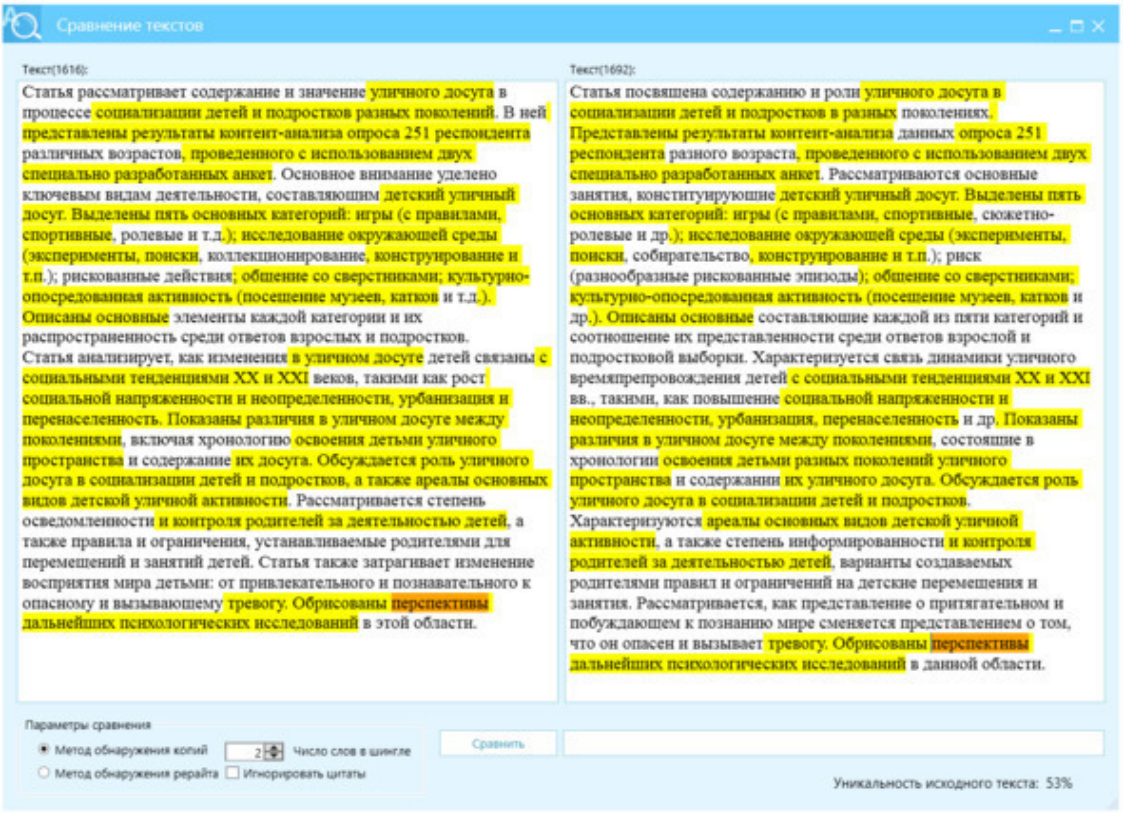


Рисунок 17 – Сверка текстов по методу Ш2, контекст см. в правом окне

Как видно из рис. 17, Copilot сделал классический поверхностный рерайт, с отличием текстов по показателю Ш2 = 53, что может оказаться недостаточным для систем антиплагиат. Причина такого ответа ИИ – примитивный запрос, он не конкретизирует сущности «глубокого рерайта».

Gemini. Google

Gemini разработан с нуля для работы с текстом, изображениями, аудио, видео и кодом. Gemini Pro 1.5, одна из версий модели, способна обрабатывать вдвое больше данных по сравнению с предыдущими версиями. Модель используется для различных целей, включая создание контента, планирование, обучение и многое другое. Google также представила облегченную версию Gemini под названием 1.5 Flash, оптимизированную для чат-приложений, видео и подписей к изображениям.

Ниже, рис. 18 и 19 показан пример загрузки в интерфейс примитивного запроса.

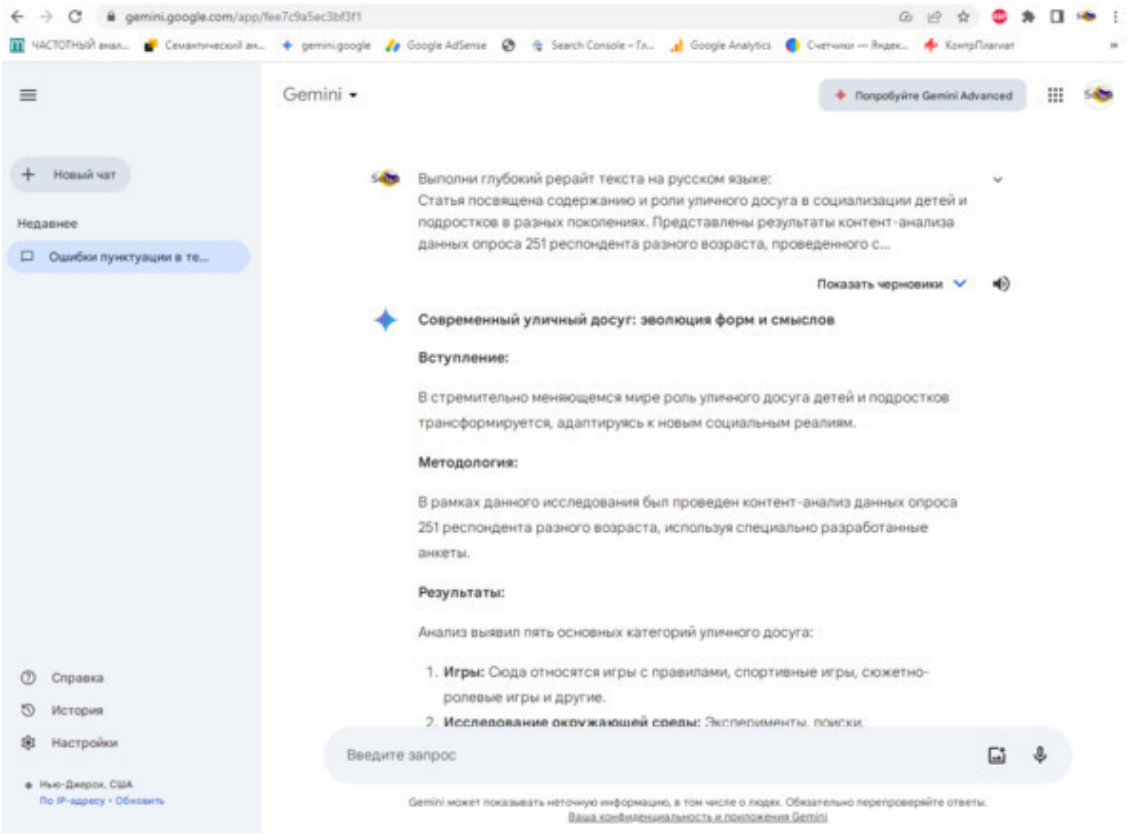


Рисунок 18 – Загрузка инструкции и текста в интерфейс gemini. google

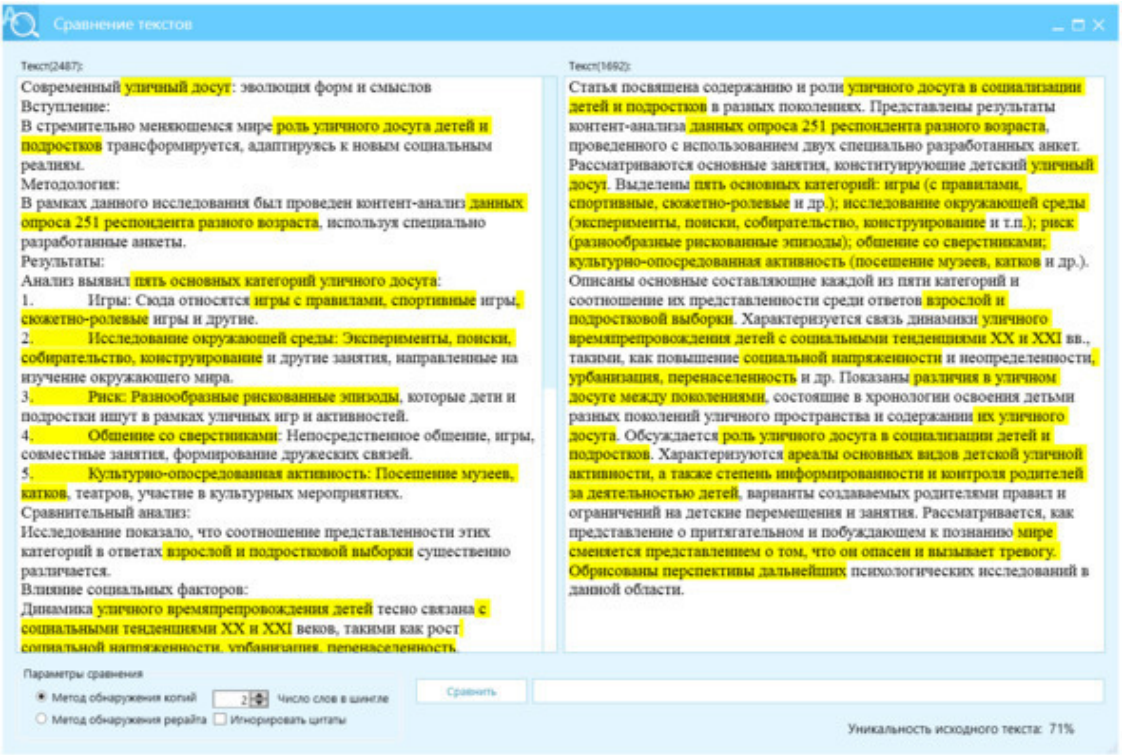


Рисунок 19 – Результат сверки текста рерайта gemini. google

Gemini. Google внес в текст свой незабываемый колорит, создав массу маркеров генеративности в виде нумерованных и маркированных списков, тавтологии (Игры: Сюда относятся

игры …), рис. 19. Текст Gemini увеличился в объеме почти на 1000 знаков, и его редактура займет достаточно много времени, процент уникальности по показателю Ш2 = 71%, что приравнивает эту нейросеть к онлайн-переводчикам, которые создают гораздо меньше текстовых проблем.

После вставки сгенерированного текста в работу необходимо, выделив весь текст, указывать правописание (меню: рецензирование – язык – язык проверки правописания – русский).

Производительность – примерно одинаков для любой нейросети – 100 тыс. знаков в минуту.

Р Использование GPT с применением параметров

Мы уже отмечали, что пользователи GPT предпочитают использовать примитивные запросы, однако есть правило, каков вопрос – таков ответ. Для многих вопросы оптимизации рерайта не праздный вопрос, а вопрос выживания или заработка.

Из того, что на слуху известно, что «температура» творит чудеса, есть еще ряд параметров, но GPT про них не рассказывает, а информация в интернет крайне неполная. Ниже мы приводим параметры, которые вы можете ввести в свою инструкцию промпт и получить выдающиеся результаты генерации или перефразирования.

Ниже обобщены параметры, которые могут использоваться для составления инструкций-промптов.

Adaptive Beam Search

Параметр «adaptive_beam_search» используется в алгоритмах поиска лучей (beam search) для динамического адаптирования процесса генерации текста. В отличие от статического beam search, который использует фиксированное количество лучей, адаптивный beam search может изменять количество лучей на основе качества текущего генерационного состояния.

Этот параметр позволяет алгоритму более гибко управлять поиском, улучшая как разнообразие, так и когерентность текста. Он адаптируется к процессу генерации, что помогает избежать избыточного повторения и поддерживать высокое качество.

Значение параметра может быть булевым (True/False) или числовым, указывающим степень адаптации (например, процент от общего количества лучей). Например, «adaptive_beam_search=True» включает адаптивный поиск лучей.

Bad Words

Параметр «bad_words» представляет собой список нежелательных слов или фраз, которые должны быть исключены из генерируемого текста. Он используется для фильтрации и предотвращения появления неуместных или неприемлемых выражений в результатах генерации.

Этот параметр помогает контролировать содержание и тональность текста, предотвращая включение слов или фраз, которые могут быть оскорбительными или неуместными.

Значения представляют собой список строк (например, « [«плохое_слово1», «нежелательное_слово2»]»). Параметр может быть пустым, если нет необходимости в фильтрации.

Coherence Threshold

Параметр «coherence_threshold» устанавливает порог для оценки когерентности текста, который генерируется моделью. Он помогает определить, насколько логично и последовательно текст соответствует контексту.

Позволяет настроить уровень когерентности текста, обеспечивая, чтобы выходные данные не содержали нелогичных или несогласованных частей. Полезен для поддержания высокого качества и понимания текста.

Значение может быть числовым, например, от 0 до 1, где 1 указывает на высокий уровень когерентности. Например, «coherence_threshold=0.8» означает, что текст должен соответствовать когерентности на уровне 80%.

Cohesion

Параметр «cohesion» отвечает за поддержание логической связности и плавности текста. Он управляет тем, насколько хорошо предложения и идеи соединены друг с другом в рамках текста.

Помогает обеспечить, чтобы текст не только был грамматически правильным, но и имел внутреннюю согласованность, что важно для естественного и понятного изложения информации.

Значения могут варьироваться от числовых (например, от 0 до 1) до категориальных (например, «низкая», «средняя», «высокая»). Например, «cohesion=0.7» может означать средний уровень связности.

Context Window

Параметр «context_window» определяет размер контекстного окна, который модель использует для анализа текста. Это количество слов или предложений, которые модель учитывает для генерации следующего слова или предложения.

Более широкий контекст позволяет модели захватывать более сложные зависимости и связи, что улучшает качество генерации текста, но требует больше вычислительных ресурсов.

Значение может быть числовым, указывающим количество слов или предложений в контексте (например, «context_window=50»). Большие значения дают модели больше информации, но могут замедлить процесс генерации.

Contextual Embedding Size

Параметр «contextual_embedding_size» определяет размер векторных представлений (эмбеддингов) контекста, который модель использует для анализа и генерации текста. Эмбеддинги представляют слова или фразы в виде многомерных векторов, которые учитывают контекстуальные связи.

Больший размер эмбеддингов позволяет модели захватывать более сложные и тонкие смысловые отношения между словами, что может улучшить качество генерации текста и понимания контекста. Однако, увеличение размера требует больше вычислительных ресурсов и памяти.

Значение может быть числовым, определяющим размер векторного пространства, например, «contextual_embedding_size=256» или «contextual_embedding_size=512». Размер обычно выбирается в диапазоне от 100 до 1024, в зависимости от доступных вычислительных ресурсов и требуемой точности.

Diversity Penalty

Параметр «diversity_penalty» применяется для управления разнообразием генерируемого текста. Он штрафует модель за избыточное использование одинаковых слов или фраз, способствуя созданию более разнообразных выходных данных.

Помогает предотвратить избыточное повторение слов и фраз, что делает текст более интересным и менее однообразным. Это важно для создания содержательных и разнообразных результатов.

Значение может быть числовым, где более высокие значения (например, «diversity_penalty=1.5») увеличивают штраф за повторение, а более низкие значения (например, «diversity_penalty=0.5») уменьшают его. Может также быть в диапазоне от 0 до 2.

Diversity Temperature

Параметр «diversity_temperature» регулирует уровень разнообразия в тексте, изменяя распределение вероятностей предсказанных слов. Он работает в связке с температурой для создания текстов с заданным уровнем креативности и непредсказуемости.

Более высокая температура (например, «diversity_temperature=1.2») делает распределение более равномерным, увеличивая разнообразие и креативность текста. Низкая температура (например, «diversity_temperature=0.7») делает распределение более сосредоточенным, снижая разнообразие.

Значение может быть числовым, например, от 0.5 до 2.0. «diversity_temperature=1.0» является стандартным значением, при котором модель генерирует текст с умеренным уровнем разнообразия.

Early Stopping

Параметр «early_stopping» управляет тем, когда процесс генерации текста должен завершиться, если модель достигает определенного состояния. Это предотвращает генерацию избыточно длинных текстов и помогает контролировать длину выхода.

Позволяет модели прекратить генерацию, когда достигнут определенный критерий, такой как достижение заданной длины текста или начало повторения. Это улучшает качество и релевантность текста, предотвращая его излишнюю длину.

Значение может быть булевым («True»/«False»), где «True» включает раннюю остановку, а «False» отключает. Также могут быть установлены дополнительные параметры, такие как количество токенов до остановки.

Encoder No Repeat Ngram Size

Параметр «encoder_no_repeat_ngram_size» применяется для предотвращения повторения определенных n-грамм в тексте, создаваемом моделями с энкодером-декодером. Это помогает избежать избыточного повторения последовательностей слов.

Устанавливает размер n-грамм, повторение которых в тексте будет запрещено. Это важно для поддержания разнообразия и избегания избыточного повторения в тексте, особенно при длительных генерациях.

Значение может быть числовым, указывающим размер n-грамм, например, «encoder_no_repeat_ngram_size=2» (запрещает повторение биграмм). Значения варьируются от 1 (без ограничения) до более высоких значений, таких как 3 или 4, в зависимости от требуемого уровня контроля над повторением.

Frequency Penalty

Параметр «frequency_penalty» регулирует штраф за частое использование одних и тех же слов в генерируемом тексте. Этот параметр помогает контролировать избыточное повторение слов и фраз, способствуя созданию более разнообразного текста.

При высоком значении этого параметра модель менее склонна к повторению часто встречающихся слов. Это способствует увеличению разнообразия и улучшению качества текста, особенно в длительных текстах.

Значение может быть числовым, в диапазоне от 0 до 2. Например, «frequency_penalty=0.5» обеспечивает умеренное снижение частоты повторяющихся слов, а «frequency_penalty=1.5» значительно увеличивает штраф.

Length Penalty

Параметр «length_penalty» контролирует, как длина генерируемого текста влияет на его вероятность. Этот параметр используется для управления длиной текста, обеспечивая баланс между слишком короткими и слишком длинными результатами.

Значение этого параметра позволяет модели избегать генерации слишком длинных или слишком коротких текстов. Помогает поддерживать оптимальную длину текста, что важно для соблюдения заданных требований.

Значение может быть числовым. Например, «length_penalty=1.0» означает нейтральное отношение к длине текста, «length_penalty> 1.0» стимулирует генерацию более длинных текстов, а «length_penalty <1.0» – более коротких.

Length Penalty Weight

Параметр «length_penalty_weight» управляет весом, который применяется к штрафу за длину текста. Этот параметр позволяет более точно настраивать влияние длины текста на вероятность его выбора.

Позволяет различать степень влияния длины текста на его вероятность. Чем выше значение, тем больше модель будет штрафовать за генерацию текста с несоответствующей длиной.

Значение может быть числовым, например, от 0.1 до 2.0. «length_penalty_weight=1.0» является стандартным значением, обеспечивающим нейтральное влияние длины, а значения выше 1.0 увеличивают штраф за превышение длины.

Max Length

Параметр «max_length» устанавливает максимальную длину текста, который может быть сгенерирован моделью. Этот параметр помогает ограничить размер выходных данных, чтобы текст не становился слишком длинным.

Обеспечивает контроль над длиной текста, предотвращая его чрезмерное удлинение. Полезен для соблюдения требований к длине текста или ограничений по памяти.

Значение представляет собой целое число, указывающее максимальное количество токенов в тексте. Например, «max_length=100» ограничивает текст 100 токенами.

Max Tokens

Параметр «max_tokens» определяет максимальное количество токенов, которые могут быть сгенерированы моделью. Это эквивалентно максимальной длине текста и контролирует объем выходных данных.

Позволяет ограничить длину текста, предотвращая его избыточное удлинение и управление ресурсами при генерации. Это важно для поддержания эффективности и качества.

Значение может быть числовым, указывающим количество токенов, например, «max_tokens=50» или «max_tokens=200», в зависимости от требований к длине текста.

Min Length

Параметр «min_length» устанавливает минимальную длину текста, который должен быть сгенерирован моделью. Этот параметр предотвращает генерацию слишком коротких текстов, обеспечивая минимально приемлемый объем информации.

Гарантирует, что текст не будет слишком кратким и обеспечит необходимую глубину или содержание. Полезен для генерации более содержательных и полноценных текстов.

Значение представляет собой целое число, указывающее минимальное количество токенов. Например, «min_length=20» гарантирует, что текст будет содержать не менее 20 токенов.

N-Gram Repetition Penalty

Параметр «n_gram_repetition_penalty» регулирует штраф за повторение определенных n-грамм в тексте, который генерируется моделью. Это помогает избежать избыточного повторения последовательностей слов, улучшая качество текста.

Этот параметр предназначен для контроля повторяемости фраз и словосочетаний. При высоком значении модель получает больший штраф за повторение одной и той же n-граммы, что способствует созданию более разнообразного текста.

Значение может быть числовым, например, от 0 до 2. Например, «n_gram_repetition_penalty=1.5» увеличивает штраф за повторение n-грамм, а «n_gram_repetition_penalty=0.8» снижает его. Значение 1.0 представляет стандартный уровень штрафа.

No Repeat Ngram Size

Параметр «no_repeat_ngram_size» устанавливает размер n-грамм, повторение которых в тексте запрещено. Этот параметр предотвращает генерацию текста, содержащего повторяющиеся последовательности слов.

Полезен для контроля за тем, чтобы текст не содержал избыточных повторений определенных фраз или словосочетаний, что делает текст более разнообразным и естественным.

Значение может быть целым числом, определяющим размер n-грамм. Например, «no_repeat_ngram_size=2» предотвращает повторение биграмм, а «no_repeat_ngram_size=3» – триграмм. Значение выбирается в зависимости от желаемого уровня разнообразия.

Num Beam Groups

Параметр «num_beam_groups» управляет количеством групп лучей (beam groups) в алгоритме поиска лучей (beam search). Это позволяет создавать несколько групп лучей, каждая из которых исследует различные пути генерации текста.

Позволяет улучшить разнообразие и качество текста, обеспечивая многогранный поиск и избегая избыточного сосредоточения на одном пути. Это способствует созданию более креативных и разнообразных текстов.

Значение может быть числовым, указывающим количество групп, например, «num_beam_groups=5». При значении 1 модель выполняет стандартный beam search, а значения выше 1 увеличивают разнообразие путем использования нескольких групп.

Num Beams

Параметр «num_beams» определяет количество лучей (beams) в алгоритме поиска лучей (beam search). Это влияет на качество и разнообразие текста, который генерируется моделью.

Большое количество лучей позволяет модели исследовать больше возможных последовательностей и выбрать наиболее вероятную, улучшая качество текста. Однако это также увеличивает вычислительные затраты.

Значение может быть числовым, например, «num_beams=5», что указывает на использование пяти лучей. Значения могут варьироваться от 1 (жадный поиск) до 10 или более, в зависимости от требуемого баланса между качеством и вычислительными ресурсами.

Presence Penalty

Параметр «presence_penalty» управляет штрафом за появление определенных слов или фраз в тексте. Он применяется для уменьшения частоты использования слов, которые уже появились в тексте.

Помогает предотвратить повторение и поддерживает разнообразие текста, обеспечивая более естественное и разнообразное изложение. Это важно для создания текстов без ненужных повторений.

Значение может быть числовым, например, от 0 до 2. Например, «presence_penalty=0.5» применяет умеренный штраф за повторение слов, а «presence_penalty=1.5» увеличивает штраф, уменьшая вероятность повторений.

Repetition Penalty

Параметр «repetition_penalty» используется для снижения вероятности повторения одинаковых слов или фраз в генерируемом тексте. Он применяется для повышения разнообразия и предотвращения избыточного повторения в результате работы модели.

Этот параметр настраивает степень наказания за повторение слов, что помогает избежать монотонности и однообразия в тексте. Чем выше значение, тем сильнее модель штрафует за повторения, что способствует более оригинальному контенту.

Значение может быть числовым, например, от 1.0 до 2.0. Значение «repetition_penalty=1.0» соответствует нейтральному штрафу, «repetition_penalty=1.5» увеличивает штраф за повторения, а значения выше 1.5 предоставляют значительный штраф.

Repetition Penalty Weight

Параметр «repetition_penalty_weight» регулирует степень штрафа за повторение слов или фраз в генерируемом тексте. Он позволяет более гибко настраивать, как сильно модель должна избегать повторений.

Позволяет управлять весом, который применяется к штрафу за повторение, чтобы добиться желаемого уровня разнообразия. Это помогает лучше контролировать качество и креативность текста, особенно в длинных генерациях.

Значение может быть числовым, например, от 0.1 до 2.0. Значение «repetition_penalty_weight=1.0» является стандартным, а значения выше 1.0 увеличивают штраф, в то время как значения ниже 1.0 уменьшают его.

Stop Sequences

Параметр «stop_sequences» определяет последовательности слов, при обнаружении которых генерация текста должна быть остановлена. Это позволяет предотвратить продолжение текста за пределы заданных границ или контекста.

Используется для контроля завершения текста и предотвращения появления нежелательных окончаний. Полезен для создания текстов, которые должны соответствовать определённым критериям или не выходить за рамки заданной темы.

Значение представляет собой список строк, например, «stop_sequences= [„Конец“, „Заключение“]», при обнаружении которых модель завершит генерацию. Могут быть указаны различные последовательности, в зависимости от требований к окончанию текста.

Temperature

Параметр «temperature» управляет степенью случайности в выборе следующего слова при генерации текста. Он изменяет распределение вероятностей, делая генерацию более креативной или более предсказуемой.

Высокие значения температуры (например, «temperature=1.2») делают распределение более равномерным, что увеличивает разнообразие и креативность текста. Низкие значения (например, «temperature=0.7») делают модель более детерминированной и менее разнообразной.

Значение может быть числовым, например, от 0.1 до 2.0. Значение «temperature=1.0» представляет стандартный уровень, а значения выше 1.0 повышают креативность, в то время как ниже 1.0 уменьшают её.

Temperature Decay

Параметр «temperature_decay» управляет изменением температуры в процессе генерации текста. Он позволяет динамически изменять уровень случайности, начиная с более высокого значения и уменьшая его по мере генерации.

Позволяет модели адаптироваться по мере генерации текста, обеспечивая высокий уровень креативности в начале и более предсказуемое завершение. Это помогает поддерживать баланс между разнообразием и когерентностью.

Значение может быть числовым, определяющим, как быстро температура уменьшается, например, «temperature_decay=0.9», что означает постепенное снижение температуры. Значения могут варьироваться от 0.1 до 1.0, в зависимости от желаемой динамики.

Top-K

Параметр «top_k» контролирует выбор следующего слова из фиксированного количества наиболее вероятных вариантов. Это метод ограничивает выбор слов до наиболее вероятных, что помогает улучшить качество генерации текста и избежать непредсказуемых результатов.

Параметр «top_k» задает, сколько лучших вариантов слов модель будет рассматривать при каждом шаге генерации. Это ограничивает выбор и помогает предотвратить появление низкокачественных слов или фраз.

Значение «top_k» является целым числом, определяющим количество слов в рассмотрении. Например, «top_k=50» означает, что будут рассматриваться 50 наиболее вероятных слов. Значения могут варьироваться от 1 (жадный поиск) до 100 или более, в зависимости от желаемого уровня разнообразия и качества.

Top-K Decay

Параметр «top_k_decay» управляет изменением значения «top_k» в процессе генерации текста. Этот параметр позволяет динамически изменять количество вариантов слов, которые рассматриваются на каждом шаге генерации.

Использование «top_k_decay» позволяет постепенно уменьшать количество рассматриваемых вариантов, что может улучшить стабильность и когерентность текста по мере его генерации. Это помогает достичь лучшего баланса между креативностью и точностью.

Значение «top_k_decay» может быть числовым, указывающим на скорость изменения «top_k». Например, «top_k_decay=0.95» означает, что «top_k» будет уменьшаться на 5% на каждом шаге. Значения могут варьироваться от 0.1 до 1.0, определяя, насколько быстро происходит уменьшение.

Top-P

Параметр «top_p» (также известный как «nucleus sampling») управляет выбором следующего слова из наиболее вероятных слов, сумма вероятностей которых достигает порогового значения «top_p». Это метод помогает сохранять разнообразие при генерации текста, фокусируясь на наиболее вероятных словах, которые составляют наиболее значимую часть распределения вероятностей.

Параметр «top_p» задает порог вероятности, до которого суммируются вероятности слов. Например, «top_p=0.9» означает, что будут рассматриваться слова, вероятность которых в сумме составляет 90% от всей вероятности. Это помогает избежать генерации текстов с низким качеством и повысить их разнообразие.

Значение «top_p» может быть числовым, например, от 0.1 до 1.0. Значение «top_p=0.9» обычно используется для сохранения хорошего баланса между разнообразием и когерентностью текста. Значения близкие к 1.0 приводят к более широкому выбору слов, а значения близкие к 0.1 – к более узкому выбору.

Top-P Decay

Параметр «top_p_decay» регулирует изменение значения «top_p» в процессе генерации текста. Этот параметр позволяет динамически изменять порог вероятности, который используется для выбора слов.

Использование «top_p_decay» позволяет постепенно корректировать уровень разнообразия текста, начиная с более высоких значений «top_p» и уменьшая его по мере генерации. Это может улучшить когерентность текста по мере его создания.

Значение «top_p_decay» может быть числовым, определяющим скорость изменения порога «top_p». Например, «top_p_decay=0.95» означает, что «top_p» будет уменьшаться на 5% на каждом шаге.

Рассмотрим на примерах промпты с параметрами, которые могут быть использованы для генерации и перефразирования текстов.

Пример 1, контроль длины и разнообразия, рис. 20.

Перефразируй следующий текст, используя следующие параметры:

– max_length: 120

– min_length: 80

– diversity_penalty: 0.7

– temperature: 0.9

– top_p: 0.85

Текст:

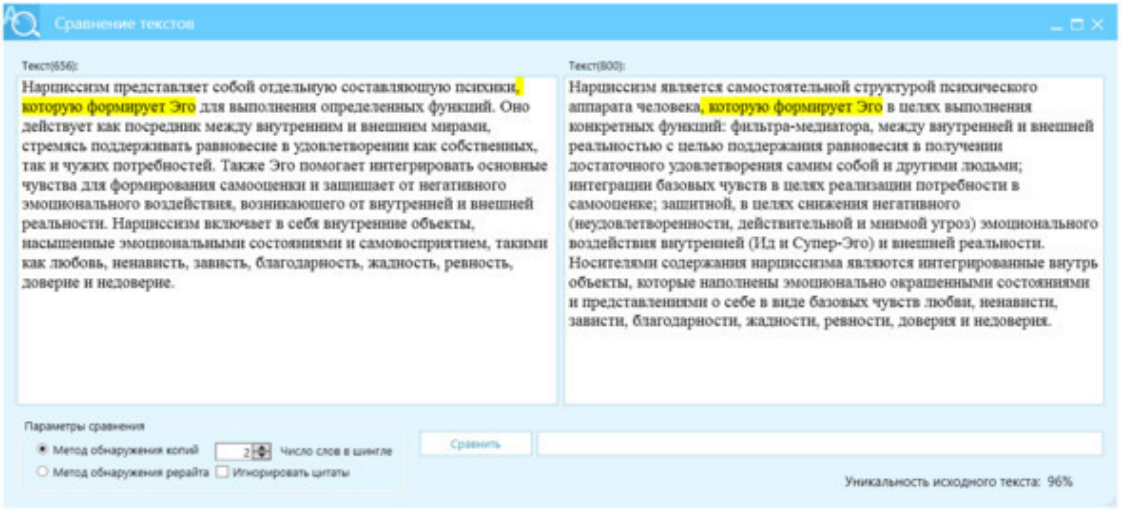


Рисунок 20 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=96%

Пример 2, фокус на когерентности и отсутствие повторений, рис. 21.

Перефразируй данный текст с учетом следующих параметров:

– coherence_threshold: 0.8

– no_repeat_ngram_size: 2

– frequency_penalty: 0.5

– presence_penalty: 0.3

– temperature: 0.7

Текст:

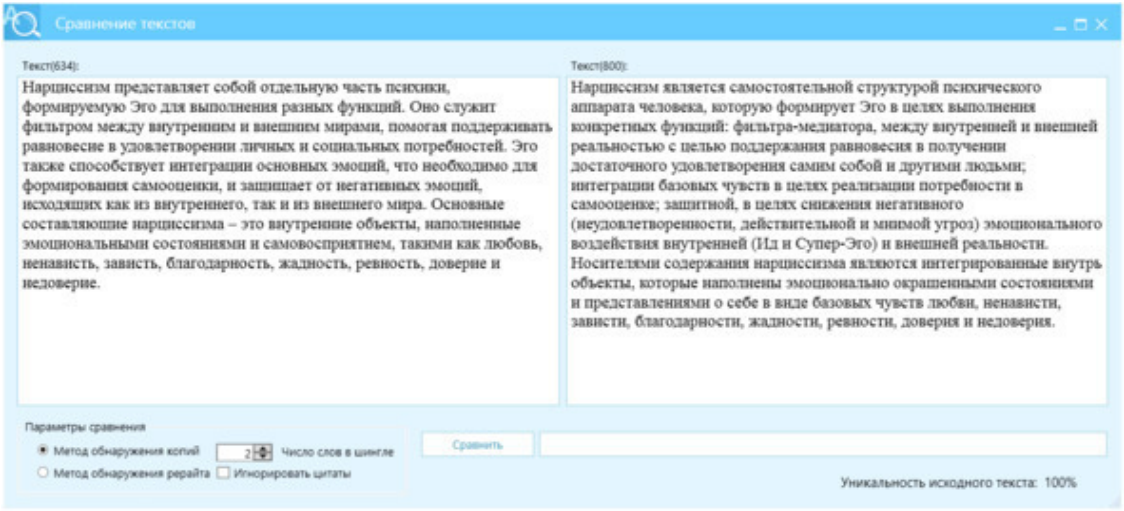


Рисунок 21 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=100%

Пример 3, контроль за разнообразием и последовательностью, рис. 22.
Перефразируй текст, используя следующие параметры:
– diversity_temperature: 0.8
– top_k: 50
– coherence_threshold: 0.9
– early_stopping: true
– repetition_penalty: 1.2
Текст:

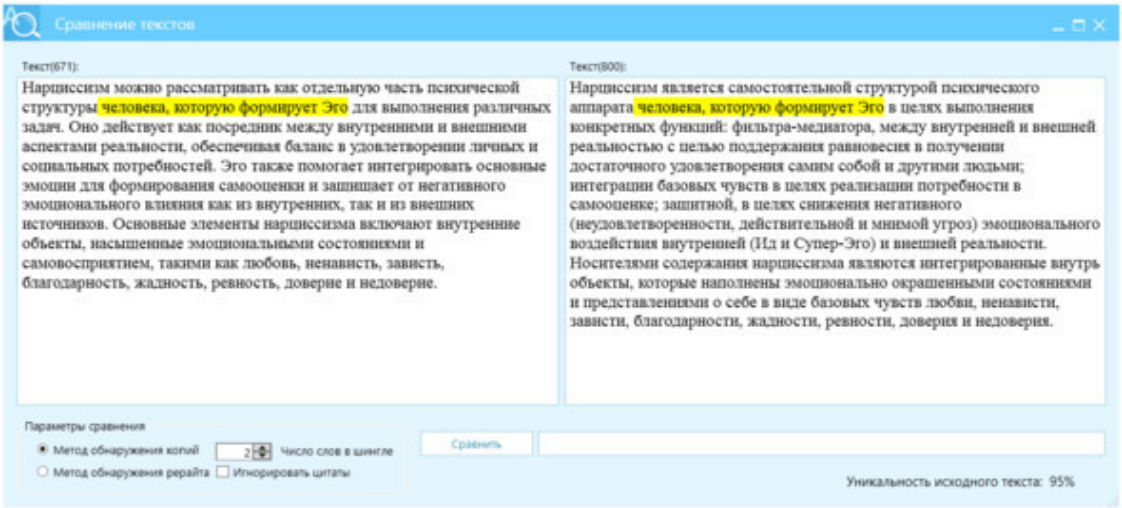


Рисунок 22 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=95%

Пример 4, фокус на разнообразие и контекст, рис. 23.

Перефразируй текст, используя следующие параметры:
– diversity_penalty: 0.6
– diversity_temperature: 0.9
– context_window: 15

– max_length: 120
– min_length: 80
– coherence_threshold: 0.75
– stop_sequences: [».», «;»]
Текст:

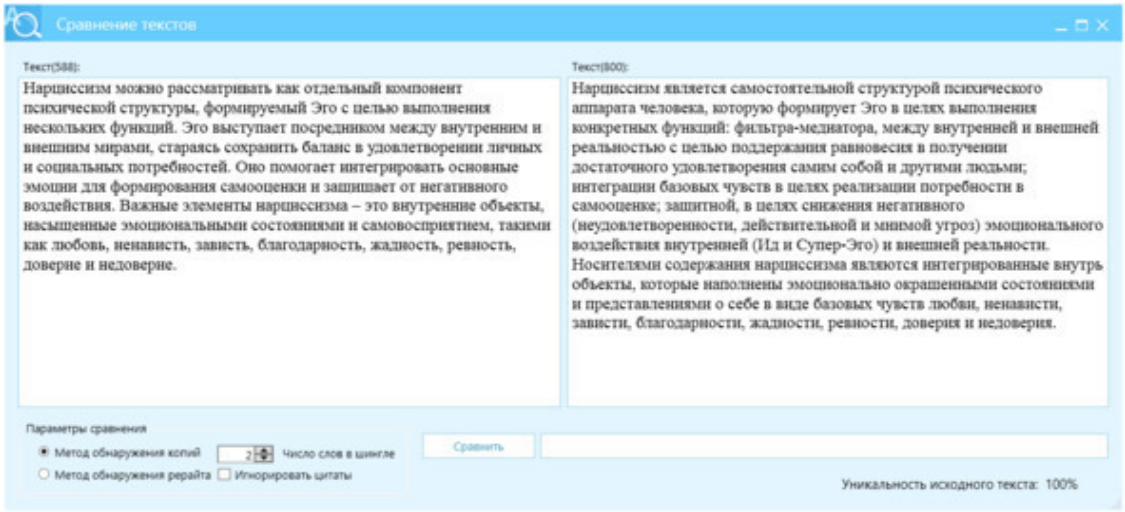


Рисунок 23 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=100%

Пример 5, увеличение когерентности и избегание длинных предложений, рис. 24.

Перефразируй текст с использованием следующих параметров:
– coherence_threshold: 0.9
– max_length: 110
– min_length: 70
– encoder_no_repeat_ngram_size: 2
– length_penalty: 0.8
– repetition_penalty: 1.0
– temperature: 0.6
Текст:

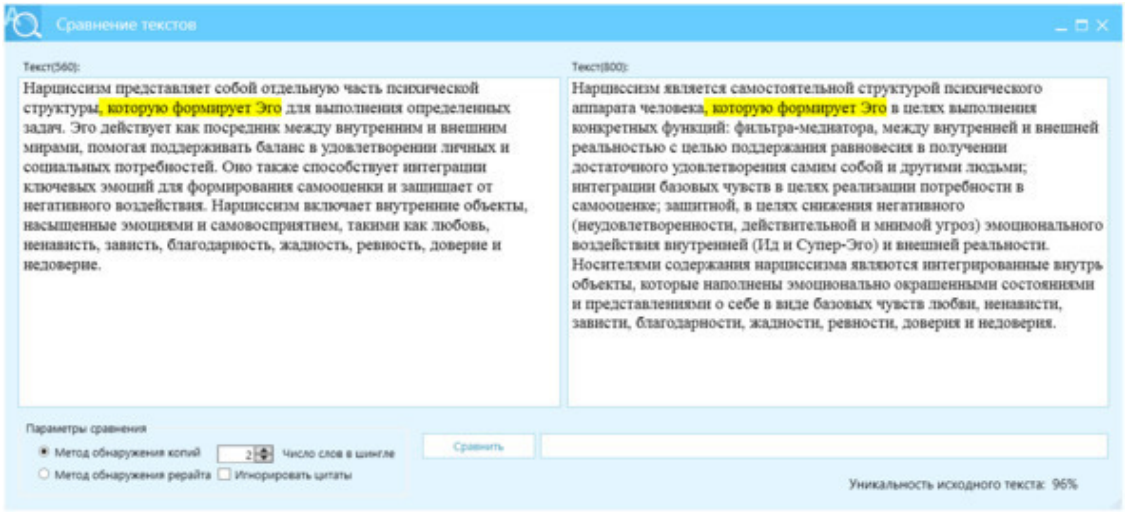


Рисунок 24 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=96%

Пример 6, управление длиной текста и температурой, рис. 25.

Перефразируй текст, используя следующие параметры:
– max_length: 150
– min_length: 100
– temperature: 0.85
– top_k: 30
– top_p: 0.8
– diversity_penalty: 0.4
– early_stopping: true
Текст:

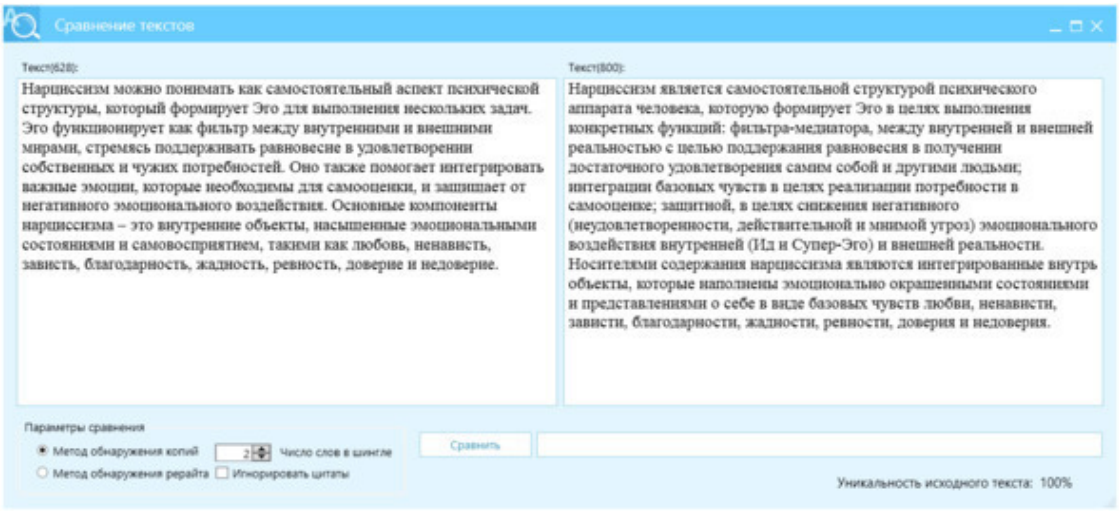


Рисунок 25 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=100%

Пример 7, управление сложностью и когерентностью, рис. 27.

Перефразируй данный текст, используя следующие параметры:
– coherence_threshold: 0.85
– contextual_embedding_size: 256
– num_beams: 5
– length_penalty: 1.0
– temperature: 0.7
– top_p: 0.85
– early_stopping: true
Текст:

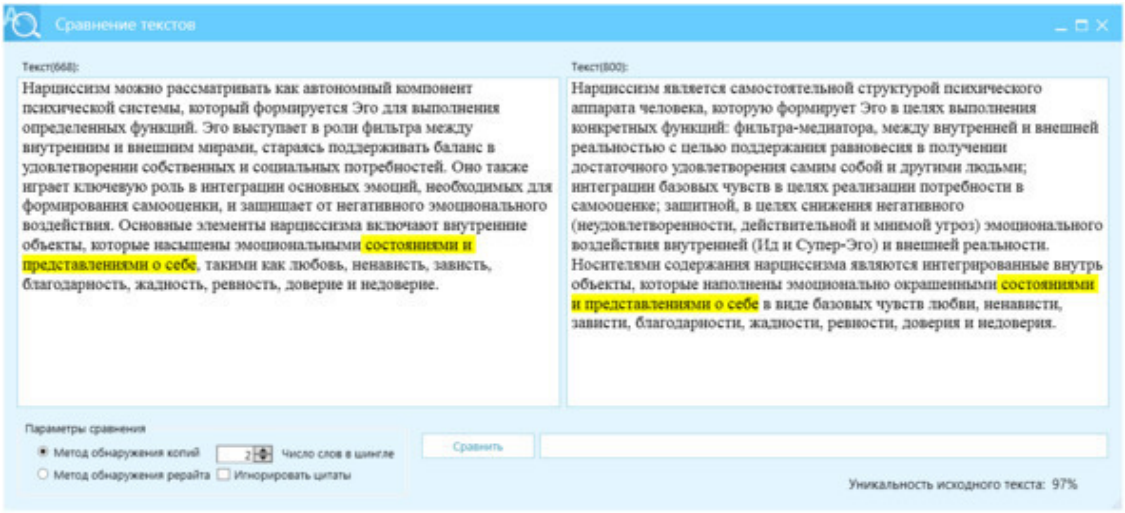


Рисунок 26 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=97%

Пример 8, управление новизной и разными стилями, рис. 27.

Перефразируй текст с использованием следующих параметров:
– diversity_temperature: 0.75
– top_k: 50
– repetition_penalty: 1.3
– max_length: 140
– min_length: 90
– bad_words: [«неудовлетворенности», «угроз»]
Текст:

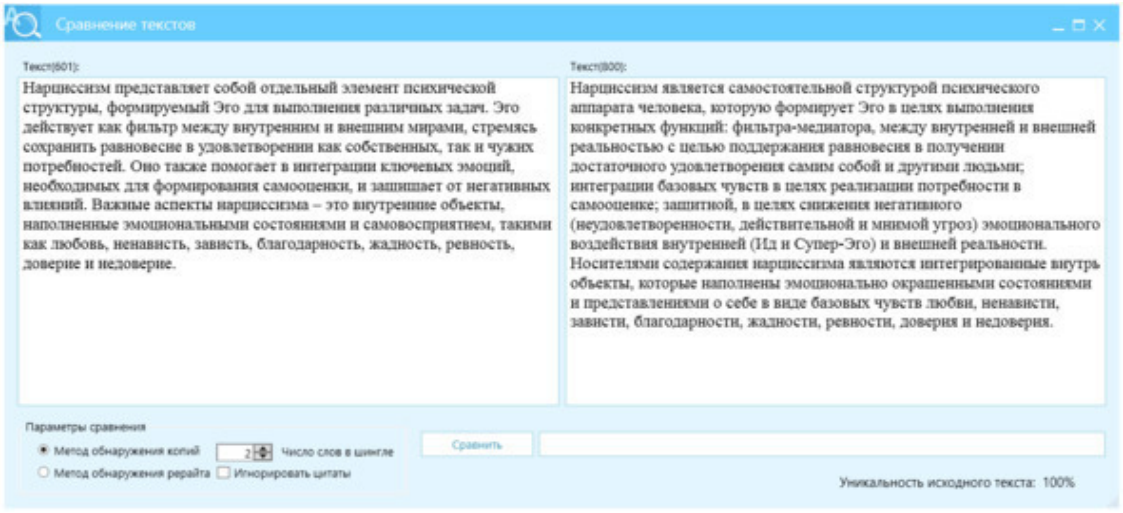


Рисунок 27 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=100%

Пример 9, фокус на длине и разнообразии, рис. 28.

Перефразируй текст, используя следующие параметры:
– max_length: 130

– min_length: 100
– diversity_penalty: 0.5
– top_k: 40
– temperature: 0.75
– no_repeat_ngram_size: 2
– early_stopping: true
Текст:

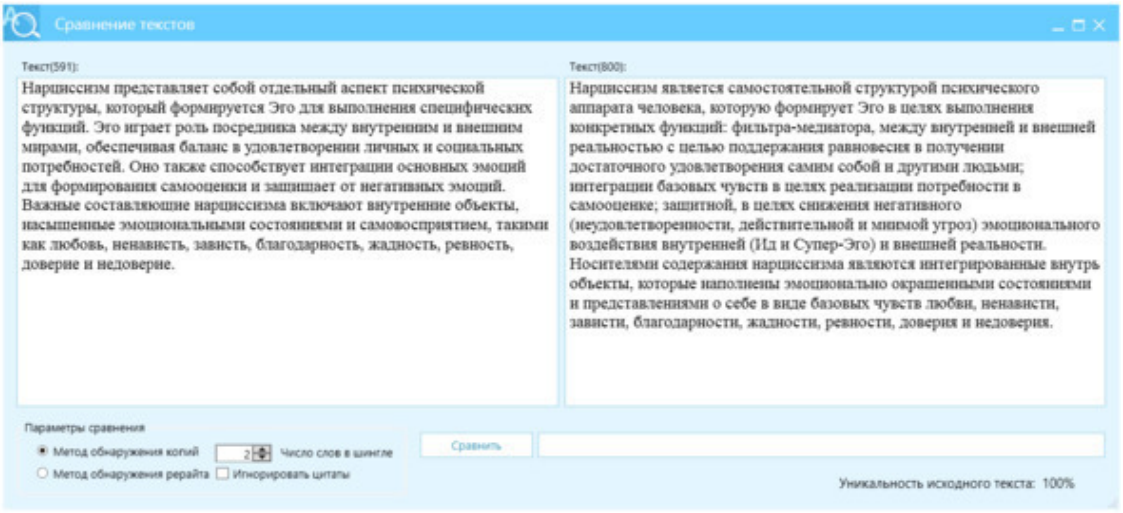


Рисунок 28 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=100%

Пример 10, управление стилем и частотой слов, рис. 29.

Перефразируй текст, используя следующие параметры:
– frequency_penalty: 0.7
– repetition_penalty_weight: 1.2
– top_p: 0.9
– max_length: 150
– min_length: 100
– diversity_penalty: 0.5
– early_stopping: true
Текст:

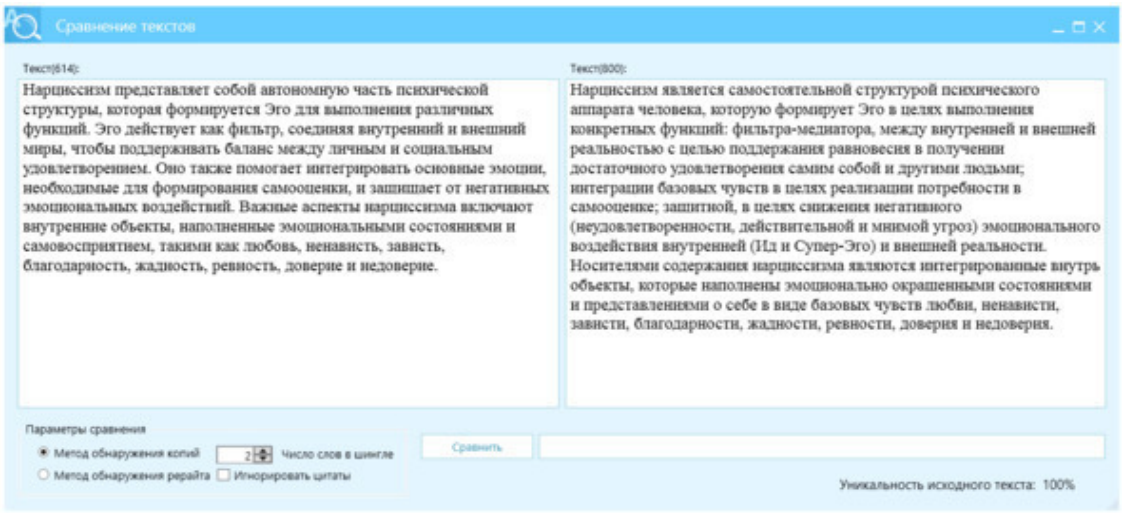


Рисунок 29 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=100%

Пример 11, акцент на оригинальность и краткость, рис. 30.

Перефразируй текст, используя следующие параметры:
– diversity_temperature: 0.9
– top_k: 20
– length_penalty: 0.6
– max_length: 100
– min_length: 70
– coherence_threshold: 0.8
– no_repeat_ngram_size: 1
– early_stopping: true
Текст:

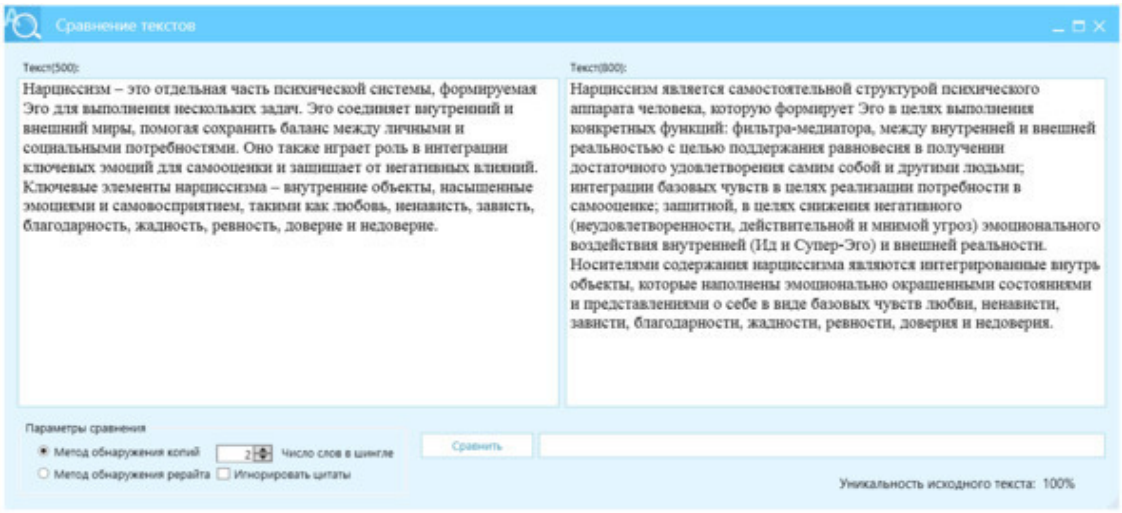


Рисунок 30 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=100%

Пример 12, поддержка краткости и специфики, рис. 31.

Перефразируй текст, используя следующие параметры:
– max_length: 120
– min_length: 85
– repetition_penalty: 1.1
– no_repeat_ngram_size: 3
– presence_penalty: 0.5
– temperature: 0.65
– top_p: 0.9
– early_stopping: true
Текст:

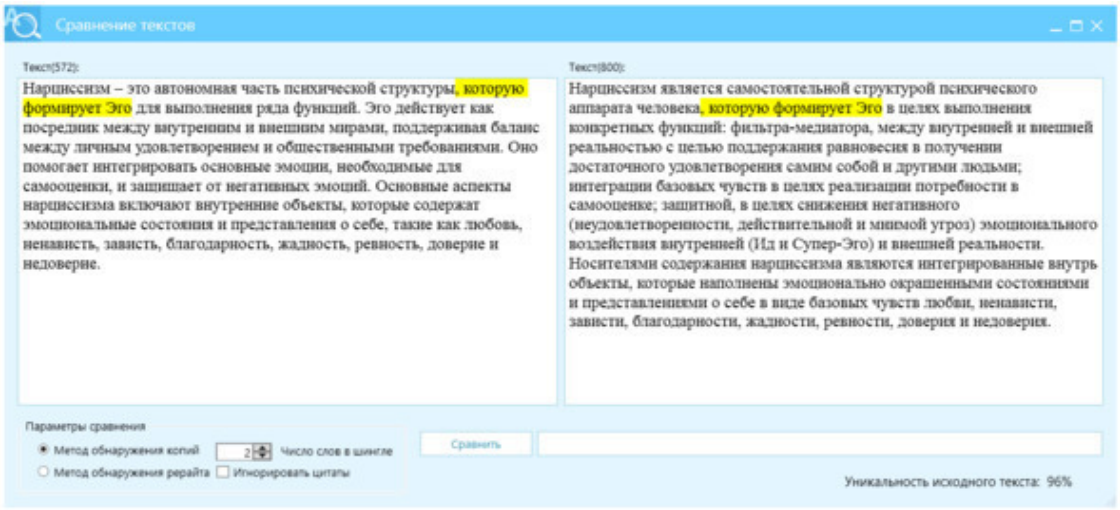


Рисунок 30.1 – Перефразированный текст, левое окно, источник – правое окно, отличие Ш2=96%

ЧР Синонимайзеры, синонимизация – ручной и с помощью программ

Метод синонимов, достаточно архаичен и не работает со дня его изобретения. Синомизацию эффективно использовать как метод доводки перевода или применения ИИ GPT. Синонимизацией можно заниматься как вручную, так и с использованием программного обеспечения. Сегодня все еще считается, что метод является одним из способов повышения оригинальности текста.

Автоматическая синонимизация, рис. 31, даже если вы используете профессиональные базы синонимов (Словари синонимов русского языка А. П. Евгеньевой, З. Е. Александровой, Н. Абрамова и т.д.), всегда была неприемлемой, так как результаты стабильно плохие. Синонимизация делает текст непонятным. Практически все системы антиплагиата распознают использование синонимов. С текстами, полученными с помощью синонимизации, можно бороться, например, с помощью переводчика Google. Однако доля ручного труда, связанного с просмотром текста и исправлением неточностей.

В интернете «гуляет» макрос, который позволяет выполнять синонимизацию локально, в WORD, используя базу синонимов MS Office. Особенность макроса в том, что он имеет настраиваемый параметр синониммизации, например можно задать требование – синонимизировать каждое третье слово текста. Результат, при проверке в антиплагиат ВУЗ положительный, но читабельность текста крайне низка.

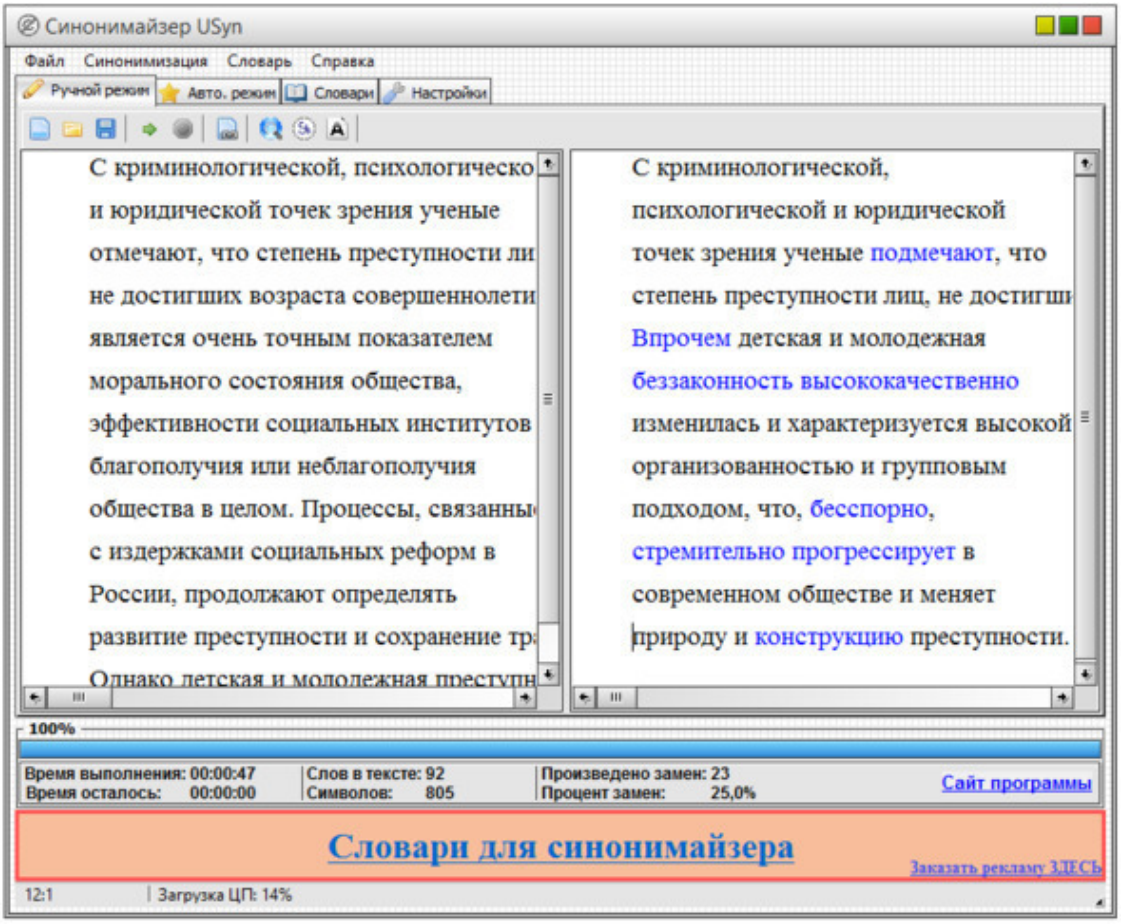


Рисунок 31 – Автоматическая локальная синонимизация, USyn, словарь синонимов Н. Абрамова

Явным недостатком авто синонимайза является то, что текст нельзя уникализировать до 100% оригинальности, в том числе и путём многократного прогона. После любой синонимизации текст становится практически непонятным, и его трудно исправить.

Напомним, что ручную синонимизацию мы рекомендуем использовать после автоматизированных методов рерайта, на этапе доводки отличия текста до нужных параметров, что позволяет быстро выполнить большой объём работы и достичь высокого показателя оригинальности.

Производительность, при норме впечатывания синонимов 10—20 слов на страницу, 1800 знаков, 150 зн./мин.

Добавление в текст «воды», вводных фраз и академических штампов

Метод используется для повышения его уникальности и читабельности.

«Вода» в тексте – это избыточная информация, которая не несет значимой смысловой нагрузки. Это могут быть общие фразы, повторения, длинные вводные конструкции и т. д. Как правило «вода» отсекается антиплагиатом, т.к. все это относится к стоп-словам.

Умеренное добавление «воды» увеличивает объем текста и изменяет структуру n-грамм, что может помочь обойти системы проверки на плагиат.

Академические штампы – это стандартные фразы и выражения, часто используемые в научных и академических текстах. Примеры включают «следует отметить», «в данном исследовании», «на основании вышеизложенного» и т. д. Данные конструкции также могут относится к стоп-словам и отсекаться АП ВУЗ до лемматизации текста.

Чрезмерная загрузка текста стоп-словами может перевести его в разряд генеративных, поэтому ниже мы даем примеры разнообразия, которое можно использовать в процессе наводнения текста.

– Введение и обзор литературы

«В данном исследовании рассматривается…»

«Настоящая работа посвящена изучению…»

«В последние годы наблюдается рост интереса к…»

«Следует отметить, что…»

– Методология

«Экспериментальное исследование было проведено для изучения…»

«В данном исследовании использовались методы…»

«Для анализа данных использовались следующие методы…»

– Результаты и обсуждение

«Результаты показывают, что…»

«На основании полученных данных можно сделать вывод, что…»

«Полученные результаты свидетельствуют о том, что…»

«Следует отметить, что…»

– Заключение

«Таким образом, можно сделать вывод, что…»

«В заключение следует отметить, что…»

«Дальнейшие исследования могут быть направлены на…»

«На основании вышеизложенного можно заключить, что…»

Эти штампы помогают структурировать текст и делают его более формальным и соответствующим академическим стандартам. Важно не злоупотреблять ими, чтобы текст не стал однообразным и генеративноподобным.

ЧР Удалить лишнее, в том числе и текст, который показан в отчете АП ВУЗ плагиатом

Удаление лишних слов и фраз, без учета проверки в АП ВУЗ, – не является эффективным способом повышения оригинальности текста. Для того чтобы удаление слов влияло на оригинальность, необходимо удалить примерно каждое третье-четвёртое слово. Это требует значительных усилий и времени, так как весь текст должен быть существенно отредактирован, в том числе и за счёт впечатывания новых слов.

Удаление неоригинальных блоков текста, согласно отчету АП ВУЗ может повлиять на оригинальность работы, метод работает, если текст перефразировался полностью, от «корки и до корки». Если перефразирования не было, после удаления фрагментов текста и проверки в антиплагиат, плагиатом могут быть отмечены места, которые раньше таковыми не были.

Наибольшую эффективность показывает скальпирующее удаление участков плагиата, при котором текст вычищается до достижения отличия по показателю Ш2 = 100%.

Р метод шингла, состоящего из двух слов, Ш2

Шингл (от английского слова «shingle», что означает «ячейка» или «кирпичик») представляет собой фрагмент канонизированного текста, состоящий из заданного количества слов (обычно от 3 до 8). Канонизированный текст – это текст, из которого удалены слова, не несущие смысловой нагрузки, такие как союзы, предлоги и знаки препинания.

Отличие между шинглами и биграммами заключается в длине. Биграммы – это последовательности из двух слов, тогда как шинглы могут содержать от 3 до 8 слов. Шинглы используются для определения уникальности контента на веб-сайтах. Поисковые системы используют

алгоритм шинглов для проверки текста на плагиат. Сайты с высокой уникальностью материалов ранжируются выше в результатах поиска.

Метод шинглов нашел свое применение в области копирайтинга, а также в анализе текстов для определения схожести и уникальности контента.

Как видно, наименьшее значение шингла = 3 словам, для получения уникального текста, который успешно пройдет проверку в АП ВУЗ необходимо добиться отличие текста рерайта от текста источника на уровне 100%.

Рассматриваемый нами шаг шингла = 2 словам, для успешного прохождения проверки в антиплагиат ВУЗ текст источника должен отличаться от полученного рерайта на показатель Ш2 = 80—95%. В случае, если текст высокочастотен на 100%.

Р Метод биграмм, отличие от метода шингла – Ш2

Биграммы – это последовательности из 2 слов, они не накладываются друг на друга как шинглы. Для успешного прохождения проверки в АП ВУЗ необходимо, при сверке по биграммам из 2 слов добиваться отличия текстов на уровне 100%.

Р КонтрПлагиат

В основе КонтрПлагиата академическая нейросеть, которая обрабатывает рутинные задачи (предложение синонимов, сверка n-грамм НКРЯ, отчеты сверки текстов индексом антиплагиат). КонтрПлагиат использует архитектуру трансформеров, в его основе также лежат рекуррентные нейронные сети (RNN). Рекуррентные нейронные сети эффективны для обработки последовательных данных, таких как текст. Они могут учитывать контекст предыдущих слов в предложении, что позволяет создавать более точные перефразирования, с учетом пересечения шинглов.

Говоря о глубоком рерайте рассмотрим результат КонтрПлагиата, рис. 32 и 33. Как видно из скриншота, красный, зачеркнутый текст выделяет слова и фразы, удаленные из текста, синим цветом помечен текст, который написан заново, остатки текста источника, в виде одиночных слов выделены черным шрифтом.

Результат проверки Ш2 показывает уникальность 97%, поэтому данный рерайт заслуженно можно отнести к глубокому перефразированию. Отличие КонтрПлагиата в его принудительным воздействием на текст, которые выходят из правил и норм русского языка, КонтрПлагиат исходит из требований антиплагиат ВУЗ.



Рисунок 32 – Процесс глубокого перефразирования – КонтрПлагиат

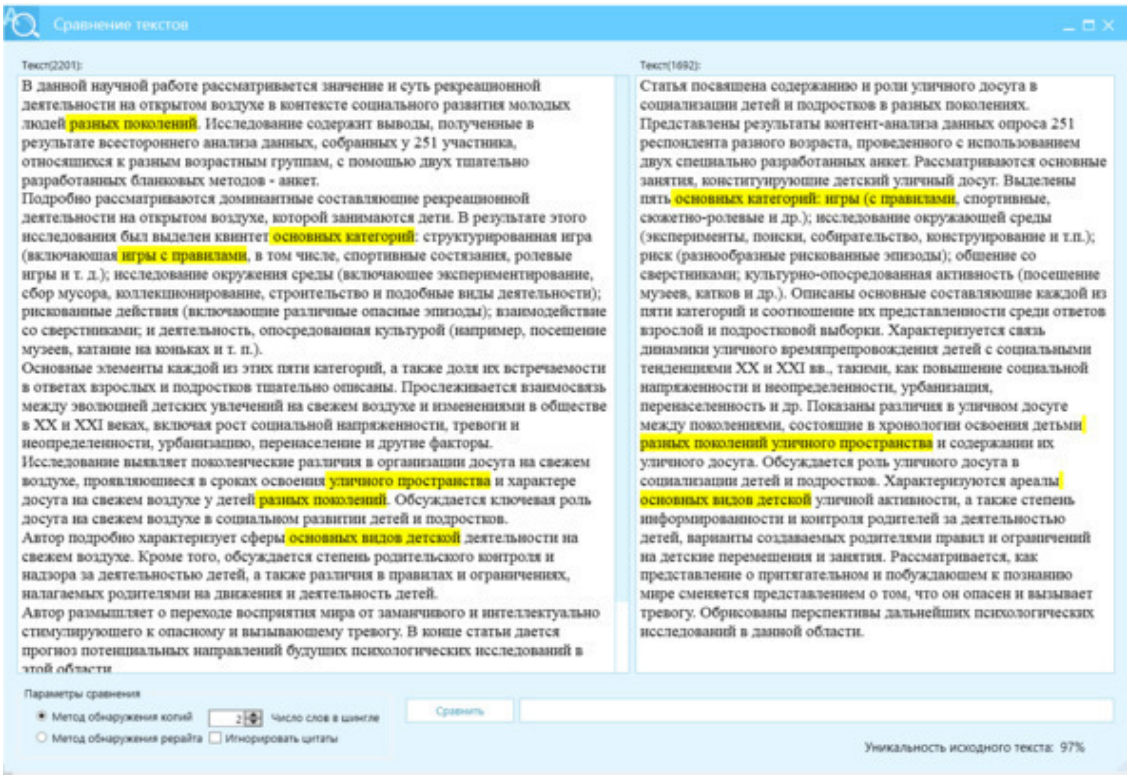


Рисунок 33 – Сверка текстов на отличие по методу Ш2, левое окно – текст КонтрПлагиата, правое окно – контекст, отличие текстов 97%

Производительность – заметно медленней переводчиков и нейросетей за счет большего числа процессов и повышенной уникальности – 1—10 тыс. знаков в минуту.

КонтрПлагиат эффективен для всех способов проверки на заимствования. Идея КонтрПлагиата – в том, чтобы после каждого второго-четвёртого слова вставить новое, авторское слово, разбивающее шинглы из 2 слов, также можно заменять каждое второе-четвёртое слово. Текст при этом не теряет смысловую нагрузку, Приложение 2, а системы антиплагиат не имеют

возможности зацепиться за последовательности хешей шинглов, пропуская текст как уникальный.

Р – Пересказ (подъём оригинальности методом изложения)

Пересказ – это трудоёмкий и длительный способ повышения оригинальности текста, который позволяет с высокой степенью достоверности гарантировать прохождение проверки на плагиат. Этот метод используется в случаях, когда требуется переписать работу «другими словами», и получить на выходе приемлемый, с точки зрения антиплагиат проверки результат.

Суть метода, необходимо прочитать абзац текста источника, все что удалось запомнить нужно напечатать «другими словами», т.е. запрещено дословное воспроизведение.

Пересказ, дабы убедиться в его эффективности, стоит сверять локально с источником, если показатель отличия текстов находится на уровне более 80%, то персказ удачен, наилучший показатель 90%, для его достижения текст необходимо подвергать постправке, принудительно избавляясь от фраз, привычных в обыденной практике, например, вместо «Гражданский кодекс РФ», можно написать «Гражданский, кодифицирующий акт».

Пересказ – очень долгий, трудозатратный способ переписывания, выполняется в два этапа, собственно пересказ, корректировка текста.

Производительность – переписать более 30 тыс. знаков в сутки крайне сложно, даже если наговаривать перефразирование в микрофон.

Несмотря на трудоёмкость, пересказ остаётся одним из надёжных способов повышения оригинальности текста.

Р Ссылочный аппарат текста источника

Новацией АП ВУЗа последних месяцев стала проблема рерайта статей, честно скаченных с elibrary.ru. Наши клиенты отмечают, что самый глубокий рерайт не дает эффекта, статьи остаются плагиатом. Виной всему ссылочный аппарат и фамилии авторов, если они употребляются в статье.

Ссылочный аппарат, содержащийся в текстах, может быть заключён в квадратные скобки – [45, С. 67—71]. Как правило, рерайтеры ссылки не трогают, а если работа написана копипастом, то не трогают точно. Ссылочный аппарат является маркером плагиата. Мы проводили эксперимент – после КонтрПлагиата текст имел отличие Ш2=100%, за счёт ссылочного аппарата текст показывал 100% плагиата, изменили цифры в квадратных скобках, плагиат пропал.

Как с этим можно бороться, взять из библиографического описания второго, третьего автора или заменить публикацию на похожую.

Р Список литературы – плагиат

Список литературы в некоторых работах доходит до 10% от общего объема, следовательно, если список литературы отмечен плагиатом, то минус 10% из вашей уникальности. Если ВУЗ в понятие уникальности включает оригинальность + цитирования, обидно вдвойне.

Многие считают плагиат литературы «глюком» в работе АП ВУЗ, но это не так, плагиат литературы возможен, если вы заимствуете литературу источника целиком, без изменений. Правки литературы в виде изменения дефисов, точек, запятых и т. д. не изменяют шинглы, поэтому они бесполезны.

Практически полезным является метод разбивки списка, путем вставки после каждых двух записей новой, оригинальной.

Было

Григорьев, А. А. Анализ динамики развития банковских карт в России / А. А. Григорьев, О. И. Михайлова // Современные вызовы и реалии экономического развития России материалы II Международной научно-практической конференции. – 2021. – №6.

Дегтерева, А. А. Формы безналичных расчетов с использованием пластиковых карт и новых банковских технологий / А. А. Дегтерева // Современные проблемы и перспективы развития банковского сектора материалы международной научно-практической конференции (заочной). – 2021. – №10.

Заборовская, А. Е. Специфика организации и современные тренды в безналичных расчетах: российская практика / А. Е. Заборовская, Е. А. Трофимова, З. К. Зоидов // Проблемы рыночной экономики. – 2021. – №4. – С. 112—132.

Исмаилов, И. Ш. Базельские стандарты банковской деятельности в историческом разрезе: предпосылки, проблемы внедрения и перспективы // Муниципальная академия. 2022. №3. С. 135—140.

Стало

Григорьев, А. А. Анализ динамики развития банковских карт в России / А. А. Григорьев, О. И. Михайлова // Современные вызовы и реалии экономического развития России материалы II Международной научно-практической конференции. – 2021. – №6.

Дегтерева, А. А. Формы безналичных расчетов с использованием пластиковых карт и новых банковских технологий / А. А. Дегтерева // Современные проблемы и перспективы развития банковского сектора материалы международной научно-практической конференции (заочной). – 2021. – №10.

Дюдикова, Е. И. Влияние электронных денег на денежное обращение // Современная наука: актуальные проблемы теории и практики. Серия: Экономика и право. – 2022. – №11. – С. 70—72.

Заборовская, А. Е. Специфика организации и современные тренды в безналичных расчетах: российская практика / А. Е. Заборовская, Е. А. Трофимова, З. К. Зоидов // Проблемы рыночной экономики. – 2021. – №4. – С. 112—132.

Исмаилов, И. Ш. Базельские стандарты банковской деятельности в историческом разрезе: предпосылки, проблемы внедрения и перспективы // Муниципальная академия. 2022. №3. С. 135—140.

Карякина, И. Е. Анализ современного состояния российского рынка платежных систем и направления его развития / И. Е. Карякина, Е. М. Тян // Экономика и бизнес: теория и практика. – 2019. – №4—3. – С. 41—49.

Р Метод пересчёта в табличных данных

Отдельно стоит рассказать о таблицах, которые являются плагиатом во всех работах, в экономических, в частности. Суммарно в экономических работах табличный материал может составлять 5—10%, это среднестатистический показатель.

Преобразовывать таблицы в рисунки с высоким разрешением можно, но в тексте работы их оставлять нельзя, поэтому рекомендуется такие таблицы убирать в приложения. В приложениях таблицы в виде рисунков разрешены. Картинки можно сделать с помощью программы ABBYY Screenshot Reader, функция – передать изображение в буфер.

Для повышения оригинальности и улучшения внешнего вида таблицы, можно преобразовать числа, например, перевести рубли в тысячи: 147 000 000 рублей станут 147 000 тыс. рублей или в миллионы – 147 млн. рублей. Это позволит не только немного увеличить оригинальность таблицы, но и даст однообразие, что улучшит читаемость таблицы.

Возможно пересчитать цифры в таблице, например, с коэффициентом 1,00095, тогда у вас будет не 147 млн. руб., а 147,14 млн. руб.

Р Замена повторяющихся (высокочастотных) слов и словосочетаний на синонимичные конструкции

Методика замены повторяющихся слов и словосочетаний на другие слова или выражения также может быть использована для повышения оригинальности текста. Метод работат на перефразированном тексте. На рис. 34 представлен частотный анализ, особенно эффективно заменять часто встречающиеся слова, как видно из рисунка, такие как «предприятия», «портативных» и т. д. Вместо этого можно использовать фразы вроде «коммерческие организации», «предпринимательские структуры», «субъекты коммерческо-предпринимательской деятельности», «субъекты экономической, рыночной деятельности», что добавит разнообразия в тексте и поднимет его уникальность.

Слово	Слова	Плотность, %	Зона текста + зона ссылок	Зона текста	Зона ссылок
колонка	колонок, колонки, колонка, колонку, колонке, колонками, колонкой, колонкам, колонках	2,97	173	173	0
портативный	портативных, портативные, портативной, портативную, портативная, портативными	1,9	111	111	0
использование	использования, использование, использовании, использованием	1,03	60	60	0

Рисунок 34 – Частотная характеристика текста

На рисунке 34 представлена частотная характеристика текста, как видно, слово «колонка» разнообразно употреблено 173 раза, на втором месте «портативный», на третьем «использование». Чтобы определить, какие слова чаще всего повторяются в работе, можно воспользоваться сервисом advego, где частные слова будут отражены в колонке. Недостаток сервиса в том, что он не отображает падежи, т.е. если у вас в тексте есть слова «предприятие, предприятием, предприятию» и т.д., Advego напишет лемматизированное слово «предприятие».

Высокочастные слова можно заменить в Word посредством команды «найти и заменить» на синонимы. Этот метод даст рост уникальности на 3—5 процентов. Применение метода занимает сравнительно небольшой промежуток времени, примерно меньше часа.

На рис. 35 приведен пример заспамленного текста, где встречаются словоформы от слова «банк».

На современном этапе развития банковских и платежных систем все большее значение приобретают операции с использованием пластиковых карт. Эти транзакции считаются одним из перспективных направлений развития сектора финансовых услуг. Тем не менее, проблема трактовки понятия «пластиковая карта» однозначно не разрешена.

Таким образом, банковской картой является инструмент безналичных расчетов, предназначенный для оплаты товаров, услуг и получения наличных денежных средств банках и банковских автоматах (банкоматах). Собственником банковской карты является кредитная организация, выпустившая ее (банк-эмитент), а держателем – клиент Банка. Каждая карта защищена персональным идентификационным номером.

«Электронные» карты в соответствии с правилами принимаются только в электронных устройствах (банкоматах, кассовых аппаратах, платежных терминалах).

Дистанционное банковское обслуживание физических лиц занимает центральное, особое место в деятельности банков, т.к. увеличивает поток клиентов, отсюда дополнительный доход и возможность разрабатывать новые банковские услуги и продукты.

Сегодня широкое распространение среди дистанционных технологий банковских услуг России получают разнообразные способы ДБО по международной технологии электронного банкинга «E- Banking».

Развитие технологий началось в 1990-х гг., когда российские банки стали внедрять системы «Банк клиент». Сейчас применяют около десятка таких технологий, различающихся способами передачи финансовой информации, техническими решениями и каналами связи.

Рисунок 35 – Пример заспамленного текста

При замене слов важно использовать принцип разнообразия, например: портативных – заменяем на носимых, портативные – переносные, портативной – носимой, портативную – небольшого размера, портативная – небольшая, портативными – носимыми. Банк – на организацию финансово-кредитного сектора, кредитную организацию, банковскую организацию и т. д. Если данный подход игнорировать, то однообразные замены по всему тексту могут привести его к генеративному-подобию, а зачищать скомпрометированный в антиплагиат документ со статусом – «подозрительный», значительно сложнее.

Р Использование аббревиатур, аббревиатурных расшифровок и собственных сокращений

Другой способ заключается в создании аббревиатур и сокращений и их последующей расшифровке. Часто в работах встречаются фразы, которые невозможно полностью избежать, такие как «финансовые ресурсы предприятия». В данном случае, можно один раз записать полную фразу, а далее указать, что в дальнейшем будет использоваться сокращение (далее по тексту – ФРП). Лучший результат дает придумывание собственных аббревиатур, например – ФинРП. Научное обоснование простое, под ФРП подразумевается множество расшифровок: Фонд развития промышленности; Фонд регионального развития; Федеральная реновационная программа; Фонд рыночных преобразований; Федеральная регистрационная палата; Фонд развития предпринимательства; Фонд развития проектов; Фонд ресурсного потенциала; Фонд региональных программ; Фонд развития патриотизма и т. д.

Таким образом, все аббревиатуры или высокочастотные фразы могут быть заменены, что позволит сократить объем текста и увеличить его оригинальность.

Существуют также случаи, когда сокращения могут быть расшифрованы. Например, если в исходном материале аббревиатура – ЦБ РФ, встречается множество раз, ее можно заменить выражением «Банк России», что также повысит оригинальность текста, главное правило ЦБ РФ – меняем на синонимичную конструкцию, Центральный Банк России на абревиатуру. Подобные изменения, если вы используете устоявшиеся подходы, мало сказываются на уве-

личении уникальности, ограничиваясь лишь небольшим повышением 1—2%, причем в связи с контекстом.

Аббревиатуры могут убить уникальность текста, например, в одной работе автор писал о государственном частном партнерстве, везде, где он употреблял эту триграмму, в скобках прописывал (ГЧП), таким образом он фиксировал по тексту два маркера, которые изничтожали уникальность. В этом случае рекомендуем – там, где в тексте написано государственно-частное партнерство заменять эту триграмму на аббревиатуру, там, где в тексте доноре написана абревиатура заменять ее на расшифровку. Кстати, триграмму «государственно-частное партнерство» желательно перефразировать: «государственное и частное экономическое партнерство», «партнерство государства и бизнеса», «экономическое взаимодействие частного и государственного секторов» и т. д.

Необходимо учитывать, что большинство описанных приемов приводит только к незначительному повышению оригинальности, на уровне 1—3 процентов. Для достижения более значимых результатов, эти методы следует применять в комплексе, что повышает затраты времени на их реализацию, но и дает больший процент уникальности.

ЧР Перестановка частей больших предложений и абзацев

Перестановка частей больших предложений и абзацев также является способом изменения текста. Здесь действует правило – от перестановки слагаемых сумма не изменяется, т.к. в результате перестановки все шинглы остаются неизменными. Хотя это позволяет улучшить или ухудшить читаемость и структуру текста, изменение оригинальности остается незначительным, составляя всего 1—2 процента. Такое редактирование требует значительных затрат времени, подобных рерайту.

Мы отмечаем этот метод частично работающим, т.к. в генеративных текстах, которые как правило содержат привычный порядок слов (расположение членов предложения по определённым нормам), который называют прямым. Перестановка частей предложения – менее привычный порядок слов (нарушенный порядок норм местонахождения членов предложения) – обратный (инверсия – от латинского «перестановка») – выводит текст из статуса генеративного.

Инверсию предложений рекомендуется применять с частотой 1—2 инверсии на 2000 знаков текста.

ЧР Конвертация текста в формулы, формул в картинки

Популярным методом является его превращение в формулы, специальные поля или картинки. Некоторые сервисы позволяют выполнять это автоматически.

В целом, этот метод показывает удовлетворительные результаты. При использовании автоматических инструментов можно достичь заметного повышения оригинальности всего за несколько минут, например, если текст содержит математический аппарат, перевод файла в DEEPL на китайский и обратно на русский преобразует все формулы в картинки, что несомненно даст уникальность. DEEPL дает возможность, при регистрации, делать три перевода файлом в месяц, объем файла до 100 тыс. знаков.

ЧР Добавление или конвертация текста в графику

Существует возможность превращения части текста в иллюстрации или схемы, что также увеличит его оригинальность. Заметим, что схему, сделанную из текста, лучше сразу преобразовать в картинку, т.к. многие схемы из-за своего текстового наполнения являются плагиатом. Однако, необходимо обосновывать использование таких элементов внутри текста. Данный метод требует затрат времени, аналогичного глубокому рерайту. Если делать преобразование текста или таблицы в картинки, то лучше эти элементы переносить в приложения, тогда это

легальный метод. Например, тесты или игровые методики можно дать в приложении картинкой, в виде красочной карточки, это дает плюс в карму вашей работы.

Включение разнообразных графических элементов, таких как фотографии объектов и предметов исследования, готовые схемы и графики, является одним из методов, способствующих созданию наглядности (картинности) в дипломной или курсовой работе. Этот прием позволяет заменить часть неоригинального текста уже готовыми изображениями, например, фотографиями предметов, используемых в разрабатываемых мероприятиях, либо графиками, основанными на табличных данных, представленных в работе. Если табличный материал – плагиат, можно его представить в виде гистограммы, от этого работа выиграет в визуальном плане, а вы избавитесь от плагиата.

С таким подходом работа приобретает картинность, наглядность, при этом требуется меньше правки, внесений изменений в сам текст. Добавление графических элементов – простая в реализации процедура, не требующая значительных усилий и временных затрат, и может применяться для замены списков перечислений.

ЧР Изменение заимствований на цитирование

Переход от заимствований к цитированию является еще одним методом, оказывающим незначительное воздействие на показатель оригинальности текста. Этот прием позволяет снизить объем заимствованного текста в работе.

Один из возможных способов реализации этого приема состоит в заключении заимствованного текста в кавычки, а затем указании ссылки в квадратных скобках, например, «тут честно заимствованный текст» [10, с. 64].

Отметим, что данный метод не всегда работает, т.к. антиплагиат выделяет цитированием фразы, хранящиеся в индексе цитирования белых источников.

Метод приобрел актуальности в связи с генеративными текстами, наполнение сгенерированного текста цитатами, в соотношении = 1 цитата на 1000—2000 знаков помогает уйти от «подозрительного документа».

Дан генеративный текст:

Эти штампы помогают структурировать текст и делают его более формальным и соответствующим академическим стандартам. Однако важно не злоупотреблять ими, чтобы текст не стал однообразным и менее читабельным

Пример оформления цитирования:

Эти штампы помогают структурировать текст и делают его более формальным и соответствующим академическим стандартам, как отмечают практики, «важно не злоупотреблять ими, чтобы текст не стал однообразным и менее читабельным» [16, С. 21].

ЧР Добавление новых цитат

Разные метрики, используемые антиплагиатными системами вузов для оценки текста, включают показатели заимствования (неправомерно использованный текст), цитирования (правомерно использованный текст) и оригинальности. Применение данного метода помогает уменьшить объем заимствований, однако не способствует увеличению объема оригинального текста.

Достаточно эффективным методом для расширения объема работы, избежание статуса подозрительного документа в генеративных текстах и повышения ее качества является включение новых цитат, например, взятых из нормативно-правовых актов. Эти цитаты помогут увеличить количество цитирований в тексте.

ЧР Выводы и рассуждения после рисунков, таблиц, вконец параграфов и глав работы

Добавление выводов и рассуждений к разным элементам работы, таким как рисунки, таблицы, параграфы и главы, является очевидным способом повышения оригинальности. Однако это также является трудозатратным способом. Несколько уменьшить трудозатраты может помочь GPT. Берем текст объемом n знаков и вместе с инструкцией: обобщи текст, напиши вывод объемом 500 знаков, отправляем его в GPT, получаем 1—2 абзаца выводов. Такой текст антиплагиат не будет детектировать как ИИ, т.к. в основе лежит обычный текст, написанный человеком, следовательно вы не получите статус: «Внимание, документ подозрительный: в документе присутствует сгенерированный текст».

При горячем желании заставить GPT написать вывод с учетом содержания параграфа, его можно перевести на китайский язык, т.к. он уменьшает количество знаков в 4—5 раз. Например, 10 листов = 18 тыс. знаков, после перевода = 3 600 знаков, что вписывается в лимиты GPT. Инструкция будет иметь следующий вид:

обобщи текст, напиши вывод объемом 500 знаков, на русском языке.

В университете могут посоветовать вам добавить дополнительные выводы к рисункам, таблицам, а также к параграфам и главам, а также использовать вводные фразы типа «в этом параграфе рассмотрим…» или «выводы данной таблицы позволяют сделать следующие выводы…». При элементарном умении использования GPT исполнение требований научного руководителя не представляет сложностей, т. к. GPT прекрасно анализирует рисунки.

Как показывает практика, наличие выводов, «своими словами», в том числе и промежуточных, ситуативно влияет на оригинальность текста. Может возникнуть ситуация, когда у вас было 2 абзаца неоригинального текста, а между ними вы добавили один абзац оригинального текста. Однако, после проверки антиплагиатом, вы можете получить 3 абзаца плагиата. Такие ситуации могут возникать и с более объемными текстами.

Возможна и другая ситуация, вы выполнили глубокое перефразирование текста, с показателем Ш2 – более 80%, при проверке этот текст показал плагиат. Вставка абзаца с текстом промежуточного вывода (примерно 2—5 предложений) выступает в роли текста-перебивки структуры изложения, что всегда дает уникальность.

Выводы и рассуждения рекомендуется добавлять, всегда, так как они влияют на восприятие вашей работы руководителем и комиссией при защите. Однако не стоит полагаться на такие доработки как на способ кардинального наращивания оригинальности, метод работает, если текст уже перефразирован или содержит генеративность, которую вы дробите выводом, написанным руками.

ЧР Проверка текста в антиплагиат. ВУЗ

По нашему мнению, Антиплагиат ВУЗ иногда ведет себя как игровая система, ваш отчет о проверке может переспать ночь и изменить показания, рис. 36 и 37.

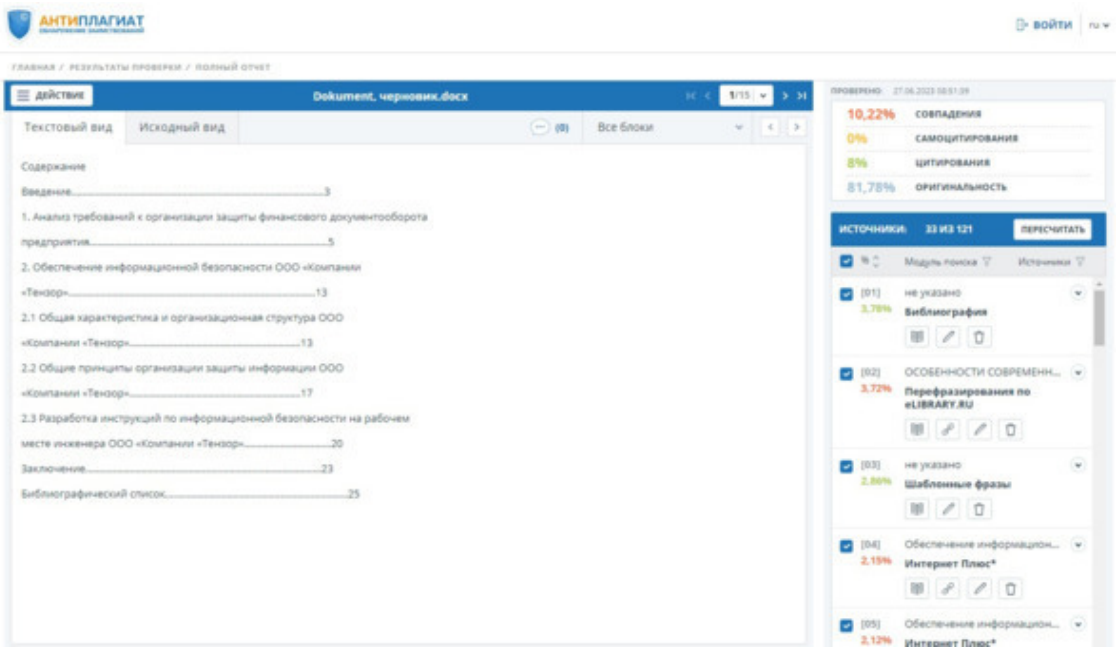


Рисунок 36 – Вечерняя проверка в АП ВУЗ

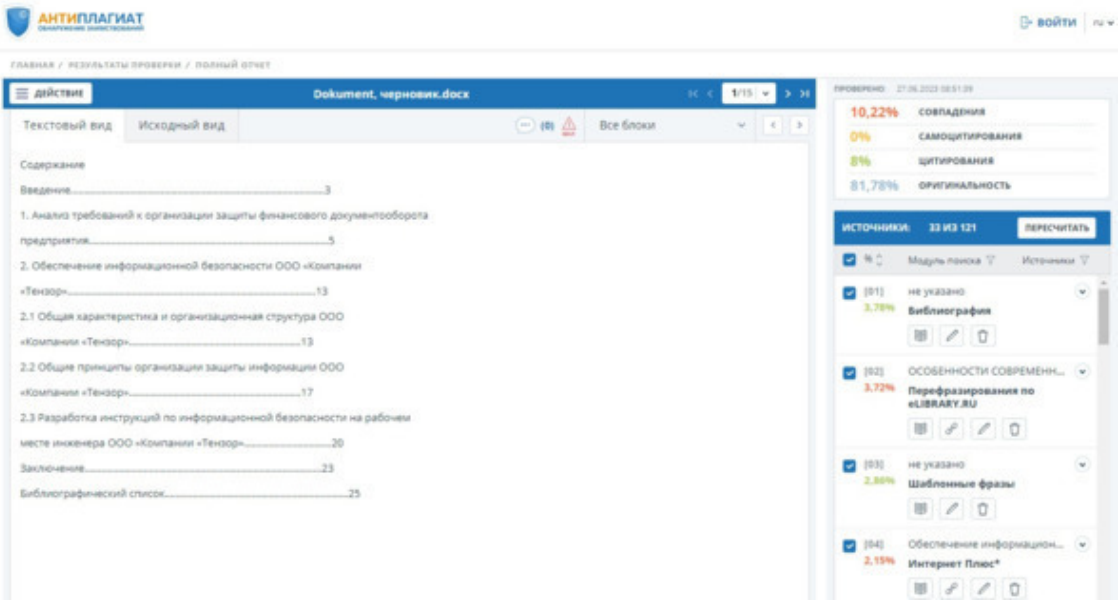


Рисунок 37 – Вечерняя проверка изменила свое мнение с утра

В этой связи мы относим проверку в антиплагиат ВУЗ к частично работающему методу. Зафиксированы случаи пересчёта процентов в отчетах, наиболее часто текст приобретает статус подозрительного спустя некоторое время, от нескольких минут, до нескольких часов.

Наш совет, если результат проверки удовлетворителен, брать справку о проверке – немедленно.

Р Добавление уникального контента

Суть этого граального подхода заключается в том, что вы вводите оригинальный текст, соответствующий по смыслу различным разделам работы, и вставляете его, например через каждый абзац не оригинального текста.

Метод кропотлив, и здесь можно применить возможности генеративных моделей, по состоянию на сегодня, такой текст не детектируется как подозрительный.

Как применить. Используйте инструкцию-промпт:

– Дан контекст, продолжи основную мысль контекста, написав 1000 знаков текста на русском языке.

Контекст:

ЧР Добавление контента, который не проиндексировал АП ВУЗ

Добавление непроиндексированного контента является способом, который схож с генерацией. Проблема метода в том, что уникальный контекст сложно найти, сегодня даже GPT уже не генерирует уникальные тексты, например известны случаи, текст имеет уникальность 10% и при этом он генеративен на 100%.

Тем не менее, при наличии непроиндексированного контента можно значительно повысить уровень оригинальности работы, просто заменив один текст другим, и на это не потребуется много времени. Однако поиск такого контента может занять не один день, а возможно и несколько недель. Поэтому рекомендуется не тратить время на этот метод.

Где брать контент, который не проиндексирован? Антиплагиат, обладая похожестью с поисковыми машинами имеет список приоритетных для индексации ресурсов, их много, т.к. сама база антиплагиата содержит миллиарды единиц хранения. Такой контекст может сохраниться на ресурсах, требующих авторизации, на сайтах существующих меньше года. Учитывайте, что подавляющее число сайтов сегодня стремиться тексты генерировать, получая оригинальный текст вы можете получить генеративные тексты.

ЧР Материалы зарубежных источников на иностранном языке

Использование материалов зарубежных источников на иностранном языке является методом, очень похожим на предыдущий, за исключением того, что эти материалы в открытом доступе.

Здесь имеется в виду следующее. Возможно, на сайтах, принадлежащих университетам в других странах, вы сможете найти статьи, учебники и другие материалы для темы своей работы. И поскольку эти источники не индексируются антиплагиатом, или индексируются на языке оригинала, их можно использовать.

Как и в предыдущем случае, необходимо убедиться в отсутствии индексации данных источников. Использование этого метода позволяет значительно увеличить уровень оригинальности работы, заменив один текст другим.

Стоит помнить, что при переводе текстов онлайн переводчиками возникает необходимость в дополнительной редакции и проверке полученного материала на оригинальность. Учитывайте, что АП ВУЗ содержит модули переводных заимствований, поэтому, если вы нашли текст на английском языке, переведите его на венгерский (финский и т.д.) а затем на русский.

ЧР Заказ платного повышения оригинальности, как правило без гарантий

Поищите на просторах Рунета – сервисы повышения уникальности, которые я отнесу в черный список пишут, что это единственный легальный метод повышения уникальности.

Для тех, кто не желает или не может самостоятельно заниматься повышением оригинальности текста, существуют специалисты, которые могут выполнить эту задачу. Однако, перед тем как принять решение о заказе услуги, стоит внимательно взвесить все за и против. Возможно, вам будет выгоднее потратить свое время на самостоятельное повышение оригинальности, чем оплачивать услуги специалиста. При выборе исполнителя важно обратить внимание на наличие гарантий. Множество ресурсов в интернете предлагают услуги по повышению оригинальности, однако не все из них обеспечивают гарантии, которые обещают. Рекоменду-

ется обращаться к проверенным и надежным источникам, чтобы избежать возможных неприятностей.

Что может положительно характеризовать исполнителя? Сертификат, рис. 38, хоть это и обычная «бумажка», обращайте внимание, когда и кем он выдан и за что. Если сертификат свеженький, то исполнитель без опыта, если организация, выдавшая сертификат отсутствует и ничем не примечательна, то вес сертификата стремится к нулю.



Рисунок 38 – Сертификат Серчьинформ (Компания «СёрчИнформ» – ведущий российский разработчик средств информационной безопасности. Резидент Инновационного центра «Сколково», входит в АРПП «Отечественный софт» и НП «Руссофт»

Второе – отзывы, они бывают разные, сгенерировать кучу положительной ванили не составляет труда, благо есть GPT, поэтому можно как генерировать, так и перефразировать уже имеющиеся отзывы на других сайтах. Стоимость отзывов на заказ 20—50 рублей, поэтому даже сотня отзывов в ВК – не гарантия профессионализма. Достоверность внушают отзывы в виде скриншотов переписки, желательно чтобы таких скриншотов было много, например несколько сотен. Ванильные отзывы, например, я погибала, меня буквально выгоняли из ВУЗа и не допускали до защиты, но тут я встретила ЕГО – сервис который сделал все за пару минут и я заплатила 100 руб. – такие отзывы проплачены или придуманы теми, кто оказывает услуги. Вдумайтесь сами, у заказчика дедлайн, он на нервах, нужно сдавать, а антиплагиат не пропускает, заказал, получил услугу – пошел проталкивать работу дальше. Реалистичны отзывы, где клиенты говорят спасибо, большое спасибо. Серьезно, при покупке в магазине нужного товара вы ведь не пишите ванильные отзывы, заплатили, забрали, сказали спасибо и на этом все.

Гарантийное обслуживание еще один момент, на который нужно обратить пристальное внимание, мы уже писали, что отчет, полежав ночь может измениться до неузнаваемости. Второе – антиплагиат показания могут отличаться в разных ВУЗах на 1—20%. Что делать клиенту, если он, при проверке в своем вузе не добрал уникальность, а исполнитель сдал работу с отчетом, в котором все нормально? Вот для всех этих случаев и нужно гарантийное обслуживание – бесплатно и быстро, например в течении суток или в течении нескольких часов – а не в течении недели.

ЧР Повышение оригинальности с помощью генеративных моделей ИИ

Данная тема будет рассмотрена ниже, отметим, что генерация текстов подразумевает разумное изложение текста, а антиплагиат требует, с учетом конкуренции в базе, аномального

изложения, ваш текст не должен быть похож на тысячи текстов на сходную тему. Поэтому генеративные модели не добирают ок. 30% отличия, фактически сравнялись с программами переводчиками. В случае примитивной генерации, благодаря примитивному запросу-инструкции можно прибавить несколько процентов, но получить генеративный текст, что чревато проблемами.

Однако, несмотря на все преимущества и потенциал данного подхода, его применение в настоящее время ограничено. Нейронные сети, используемые в моделях искусственного интеллекта, пока что не обладают полным пониманием сущности рерайта именно для антиплагиата, который ВУЗ, что может привести к возникновению проблем с оригинальностью текстов, она выражается в недоборе отличия текстов и он существенный, как правило удается достичь показателя Ш2 – 40%, вместо требуемых 80%.

Copilot, это попытка Microsoft укрепить свою репутацию, типа, мы умеем. Одной из простейших задач для всех GPT-образных является суммаризация (реферирование) и перефразирование текста, посмотрим, как с этой задачей справляется Copilot.

Простейший промпт (инструкция): Выполни глубокий рерайт текста на русском языке, рис. 39.

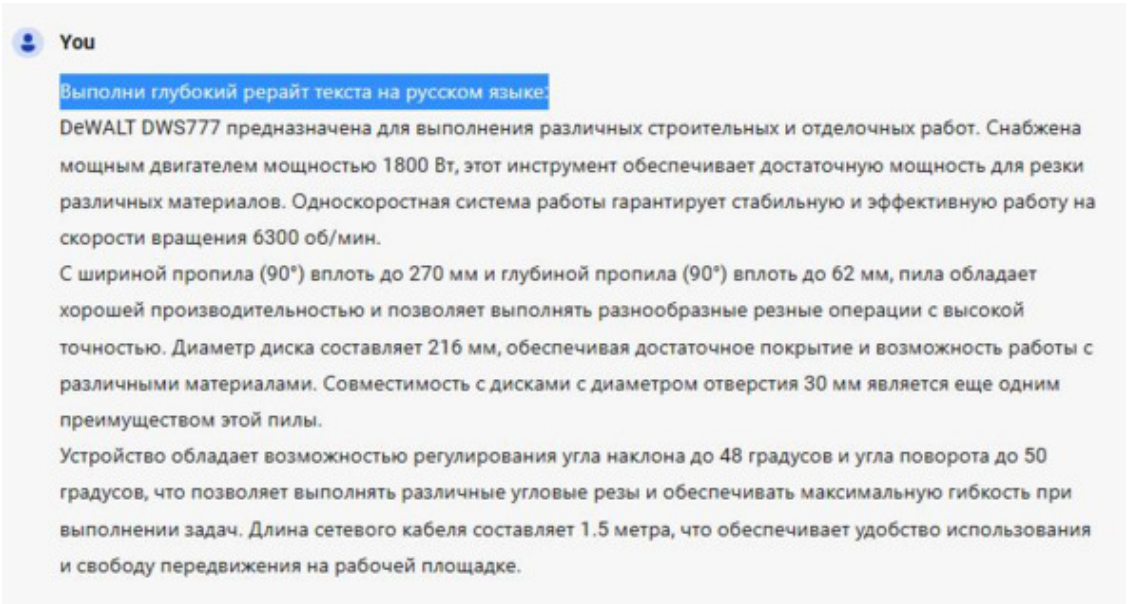


Рисунок 39 – Загрузка в Copilot текста для рерайта

На рис. 40 представлен скриншот перефразированного текста, а на рис. 41 сверка с источником, где в левом окне, текст после рерайта, в правом, источник. Желтым выделено то, что не изменилось.

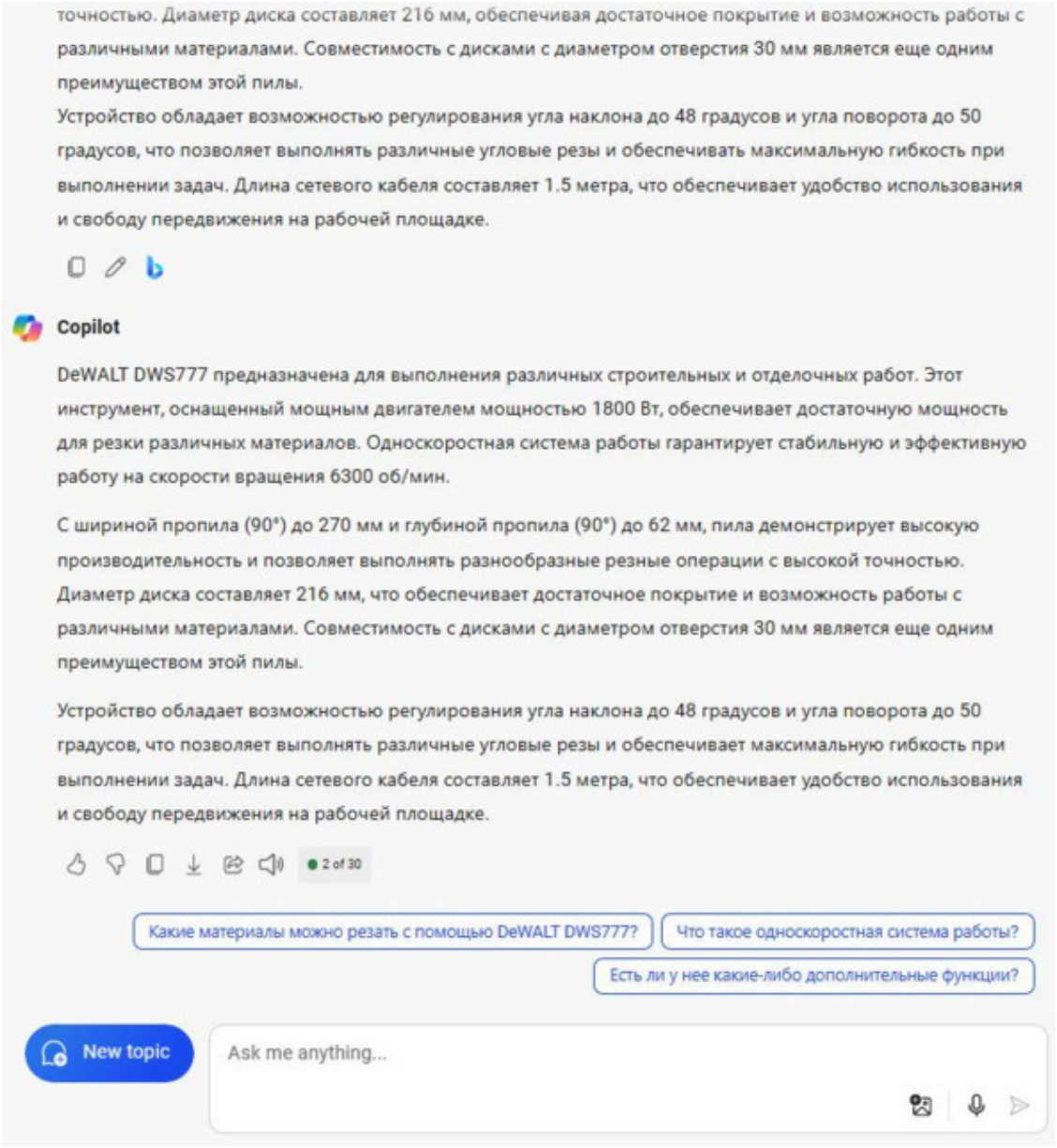


Рисунок 40 – Перефразирование от Copilot

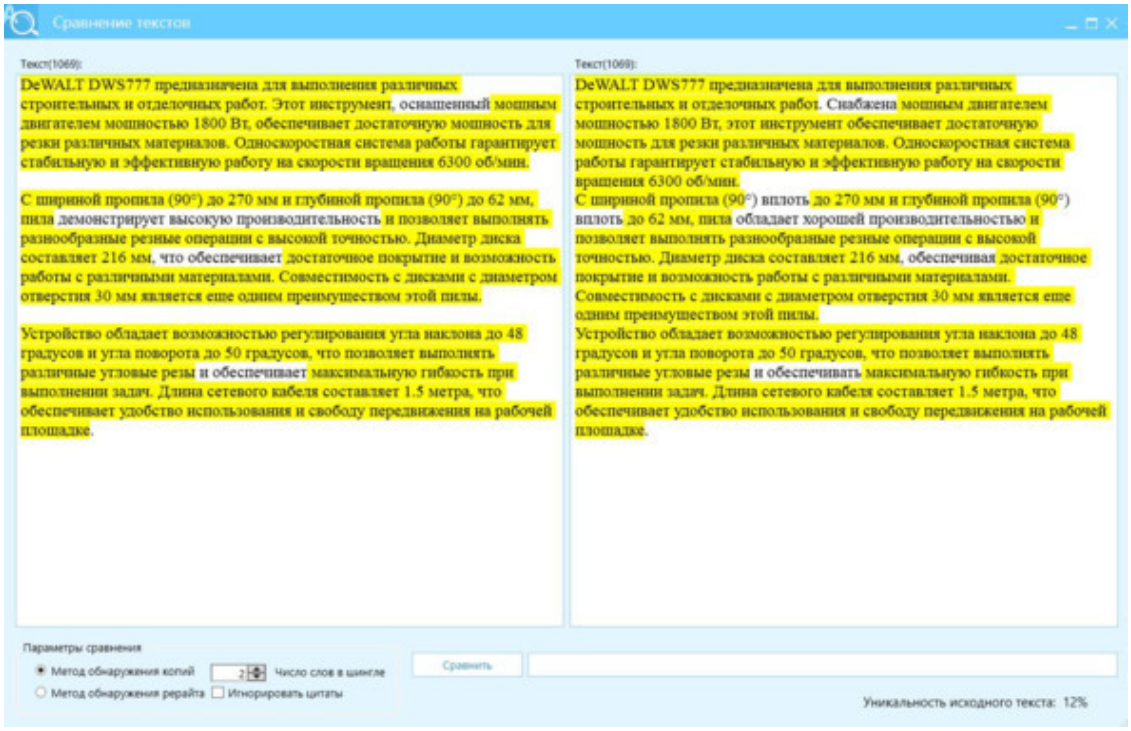


Рисунок 41 – Сверка текстов на отличие

Как видно из рисунка 3, Copilot обеспечил отличие текстов 41%, это примерно в 2—3 раза хуже, если бы мы переводили текст на английский язык и затем на русский. Результат эксперимента позволяет сделать вывод, что Copilot слабо приспособлен для рерайта текстов, т.к. для успешного прохождения антиплагиат ВУЗ, требуется отличие текстов на уровне 90—95%.

Конец ознакомительного фрагмента.

Текст предоставлен ООО «Литрес».

Прочитайте эту книгу целиком, купив полную легальную версию на Литрес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.