

Волшебный ИИ для страны ОЗ

Юрий Викторович Чехович, к.ф.-м.н.

Исполнительный директор компании Антиплагиат

Обнаружение заимствований строится на конфликте



Детекторы плагиата...

...в ожиданиях
пользователей



Детекторы плагиата...

...в ожиданиях
пользователей



...в предложении
производителей

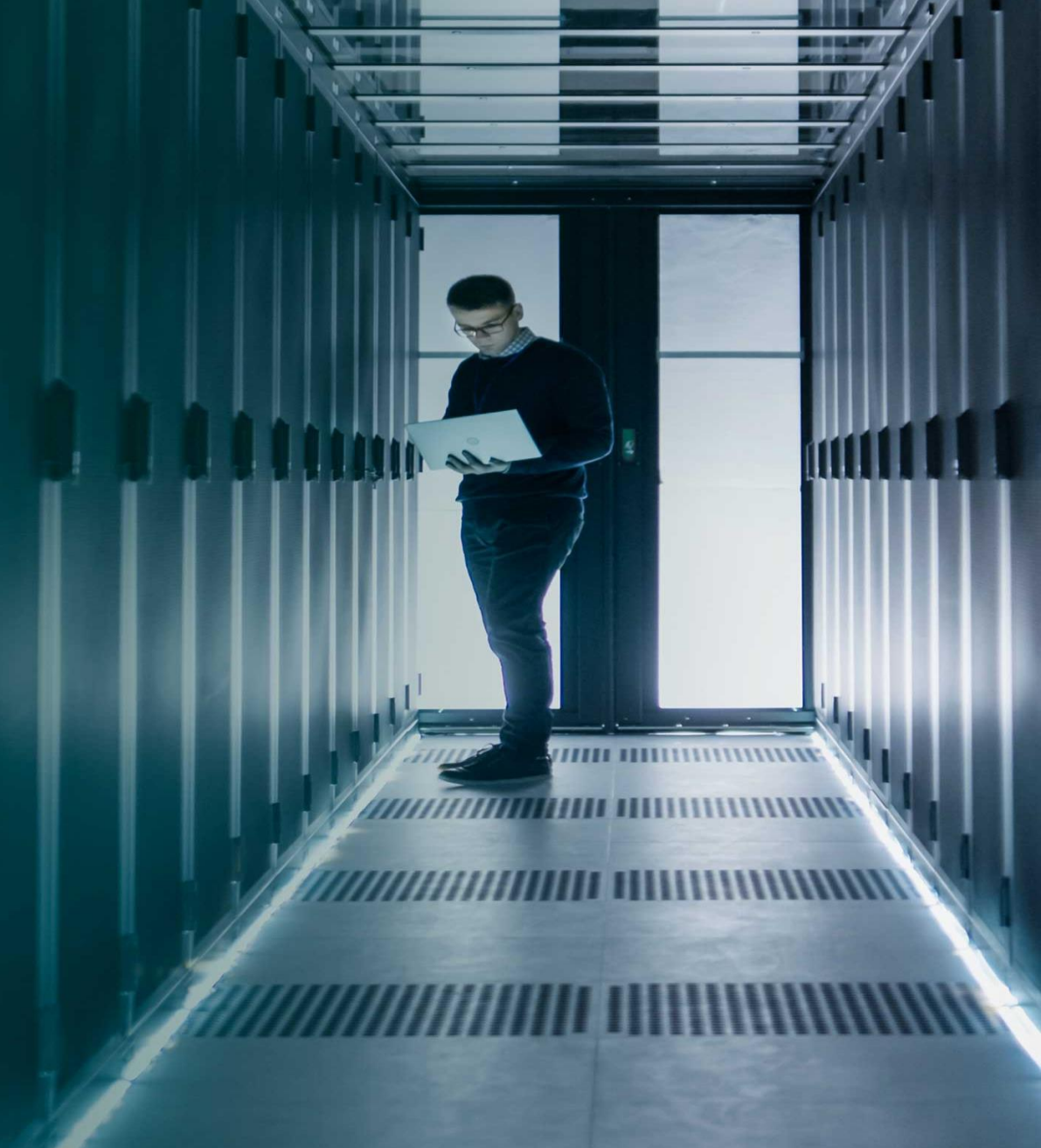


**Можно ли создать инструмент,
соответствующий представлениям пользователей?**

**Еще несколько лет назад
это казалось фантастикой**



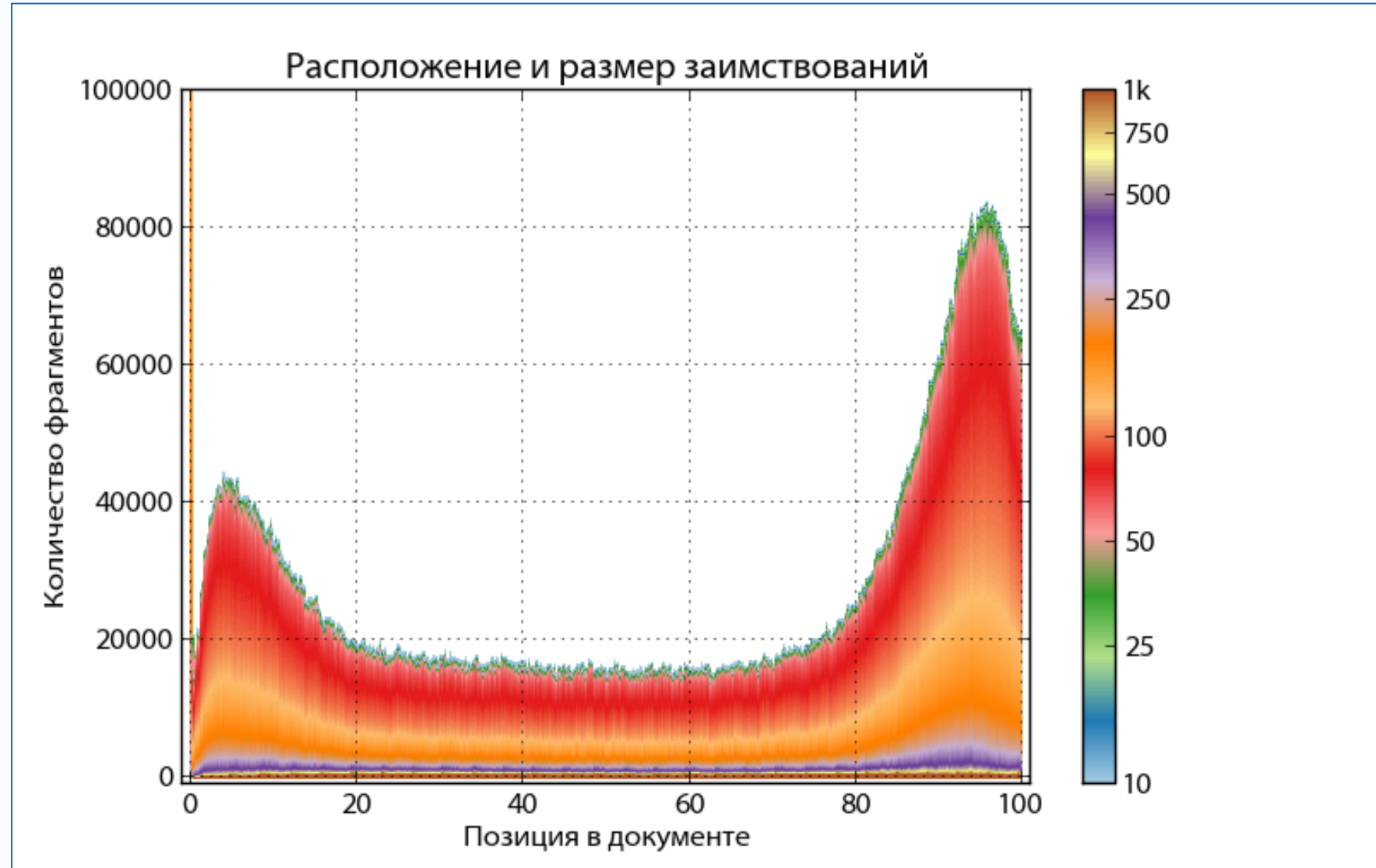
Галерея задач ИИ в обнаружении заимствований



Структура документа

Проблема:

совпадения распределены
неравномерно и имеют
разную значимость



Структура документа

Решение:

Автоматически выделять структурные части проверяемого документа с возможностью выбора

Скоро:

Выделение в основной части документа элементов структуры IMRAD

ЗАГРУЗКА ДОКУМЕНТОВ ?

Документы | **Параметры проверки**

Проверить по: ? ▼ **Интервал поиска:** ? с по

Проверить с учётом структуры: ?

Основная часть Титульный лист Содержание Библиография Приложение

Использовать распознавание текста (OCR) ?

Искать с учетом редактирования ?

? Для проверки выбраны все разделы документа. Чтобы исключить разделы из проверки, перейдите на вкладку "Параметры проверки". ×

Метаданные

Проблема:
необходимо среди совпадений
выявлять случаи
самоцитирования

Труды ИСР РАН, том 34, вып. 1, 2022 г. / Труды ИСР РАН/Proc. IP RAS, vol. 34, issue 1, 2022

DOI: 10.15514/ISPRAS-2022-34(1)-11

О комбинированном алгоритме обнаружения заимствований в текстовых документах

¹К.Ф. Сафин, ORCID: 0000-0001-9891-5513 – kamil.safin@phystech.edu
²Ю.В. Чебоксар, ORCID: 0000-0002-5204-5484 – chebovich@asp-team.ru
³Московский физико-технический институт,
141701, Россия, Московская область, г. Дельтаградский, Институтский переулок, д.9
⁴Федеральный исследовательский центр "Информатика и управление" РАН
119333, Россия, Москва, Выхомский, д.44, корп.2
⁵Коллегия "Антиплагиат"
121205, Россия, Москва, пер. Соколовский инженерного центра, Кольцовый б-р, д. 42, стр. 1

Аннотация. Поиск заимствований в текстовом документе по отношению к обширной коллекции потенциальных источников является вычислительно тяжелой задачей. При этом существуют как внешние, так и внутренние методы поиска заимствований, которые не используют внешней информации, а анализируют исключительно предоставленный документ. Эти методы не отличаются точностью, но обеспечивают довольно высокую производительность. В работе предлагается комбинированный подход к обнаружению текстовых заимствований, основанный на использовании внутренних методов для выявления высокооригинальных документов, проверка которых по внешней коллекции не требуется. Предлагаемый алгоритм позволяет разгрузить систему поиска заимствований по внешней коллекции, отфильтровав документы с высокой степенью оригинальности. В работе предлагается алгоритм поиска внутренних заимствований, основанный на результатах вычислительных экспериментов.

Ключевые слова: обработка естественного языка, обнаружение заимствований, внутренние заимствования, поиск выборов в статистике, антиплагиат

Для цитирования: Сафин К.Ф., Чебоксар Ю.В. О комбинированном алгоритме обнаружения заимствований в текстовых документах. Труды ИСР РАН, том 34, вып. 1, 2022 г., стр. 11-16. DOI: 10.15514/ISPRAS-2022-34(1)-11

Combined method for plagiarism detection in text documents

¹K.F. Safin, ORCID: 0000-0001-9891-5513 – kamil.safin@phystech.edu
²Y.V. Chebovich, ORCID: 0000-0002-5204-5484 – chebovich@asp-team.ru
³Moscow Institute of Physics and Technology,
9 Intitubnky per., Dolgoprudny, Moscow Region, 141701, Russia
⁴Federal Research Center "Computer Science and Control" RAS
44 Fiviolova St., Moscow, 119333, Russia
⁵AntiPlagiat Company
42 Vol'shoy Blvd., Moscow, 121205, Russia

Abstract. There are two global approaches to the problem of searching plagiarism in the text: external and internal search. The first approach implies search through an external collection of documents that could have been used for text reuse. The second approach, on the contrary, does not use any external data, but analyzes the text by itself. It is proposed to combine these two approaches to speed up the search for plagiarism. With a large flow of documents that need to be checked, the outer corpus search system processes each document and finds plagiarized blocks in each document, if there are any. However, internal search could be used to determine the fact of plagiarism. Thus, it is possible to reduce the number of documents for the expensive

151

Статья

Московский физико-технический институт
(национальный исследовательский университет)

На правах рукописи

Сафин Камилъ Файтович

Комбинированные методы выявления заимствований в текстовых документах

Специальность 05.13.17
«Информатика, включая информатику»
Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
к.ф.н.
Чибриков Юрий Владимирович

Москва, 2022

Диплом

Московский физико-технический институт
(национальный исследовательский университет)

На правах рукописи

Сафин Камилъ Файтович

Комбинированные методы выявления заимствований в текстовых документах

Специальность 05.13.17
«Информатика, включая информатику»
Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
к.ф.н.
Чибриков Юрий Владимирович

Москва, 2022

Диссертация

Труды ИСР РАН, том 34, вып. 1, 2022 г. / Труды ИСР РАН/Proc. IP RAS, vol. 34, issue 1, 2022

DOI: 10.15514/ISPRAS-2022-34(1)-11

О комбинированном алгоритме обнаружения заимствований в текстовых документах

¹К.Ф. Сафин, ORCID: 0000-0001-9891-5513 – kamil.safin@phystech.edu
²Ю.В. Чебоксар, ORCID: 0000-0002-5204-5484 – chebovich@asp-team.ru
³Московский физико-технический институт,
141701, Россия, Московская область, г. Дельтаградский, Институтский переулок, д.9
⁴Федеральный исследовательский центр "Информатика и управление" РАН
119333, Россия, Москва, Выхомский, д.44, корп.2
⁵Коллегия "Антиплагиат"
121205, Россия, Москва, пер. Соколовский инженерного центра, Кольцовый б-р, д. 42, стр. 1

Аннотация. Поиск заимствований в текстовом документе по отношению к обширной коллекции потенциальных источников является вычислительно тяжелой задачей. При этом существуют как внешние, так и внутренние методы поиска заимствований, которые не используют внешней информации, а анализируют исключительно предоставленный документ. Эти методы не отличаются точностью, но обеспечивают довольно высокую производительность. В работе предлагается комбинированный подход к обнаружению текстовых заимствований, основанный на использовании внутренних методов для выявления высокооригинальных документов, проверка которых по внешней коллекции не требуется. Предлагаемый алгоритм позволяет разгрузить систему поиска заимствований по внешней коллекции, отфильтровав документы с высокой степенью оригинальности. В работе предлагается алгоритм поиска внутренних заимствований, основанный на результатах вычислительных экспериментов.

Ключевые слова: обработка естественного языка, обнаружение заимствований, внутренние заимствования, поиск выборов в статистике, антиплагиат

Для цитирования: Сафин К.Ф., Чебоксар Ю.В. О комбинированном алгоритме обнаружения заимствований в текстовых документах. Труды ИСР РАН, том 34, вып. 1, 2022 г., стр. 11-16. DOI: 10.15514/ISPRAS-2022-34(1)-11

Combined method for plagiarism detection in text documents

¹K.F. Safin, ORCID: 0000-0001-9891-5513 – kamil.safin@phystech.edu
²Y.V. Chebovich, ORCID: 0000-0002-5204-5484 – chebovich@asp-team.ru
³Moscow Institute of Physics and Technology,
9 Intitubnky per., Dolgoprudny, Moscow Region, 141701, Russia
⁴Federal Research Center "Computer Science and Control" RAS
44 Fiviolova St., Moscow, 119333, Russia
⁵AntiPlagiat Company
42 Vol'shoy Blvd., Moscow, 121205, Russia

Abstract. There are two global approaches to the problem of searching plagiarism in the text: external and internal search. The first approach implies search through an external collection of documents that could have been used for text reuse. The second approach, on the contrary, does not use any external data, but analyzes the text by itself. It is proposed to combine these two approaches to speed up the search for plagiarism. With a large flow of documents that need to be checked, the outer corpus search system processes each document and finds plagiarized blocks in each document, if there are any. However, internal search could be used to determine the fact of plagiarism. Thus, it is possible to reduce the number of documents for the expensive

151

Другие статьи

Метаданные

Решение:

Автоматически определять авторов работы и классифицировать источники

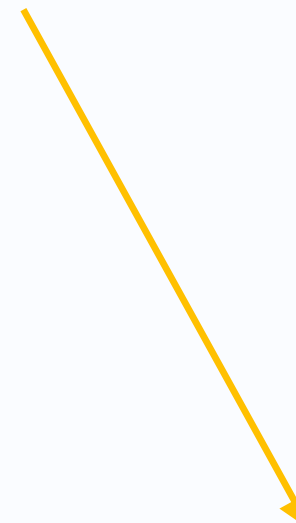
Авторы документа ?	Чехович	Юрий Викторович
Имя исходного файла	Том.pdf	
Название документа	Том	
Тип документа	Не указано	

[РЕДАКТИРОВАТЬ СВОЙСТВА](#)

Заимствованные изображения



**Проблема:
плагиат изображений**



**Проблема:
Фальсификация результатов**


Заимствованные изображения


 Свойства документа

 Структура документа

 Поиск по изображениям **NEW**

 Текстовые метрики **NEW**

 Параметры проверки

 Статистика по документу

Система может обнаруживать заимствования по изображениям в документах. Алгоритм помогает бороться с заимствованиями изображений из других документов даже с учетом их видоизменений.

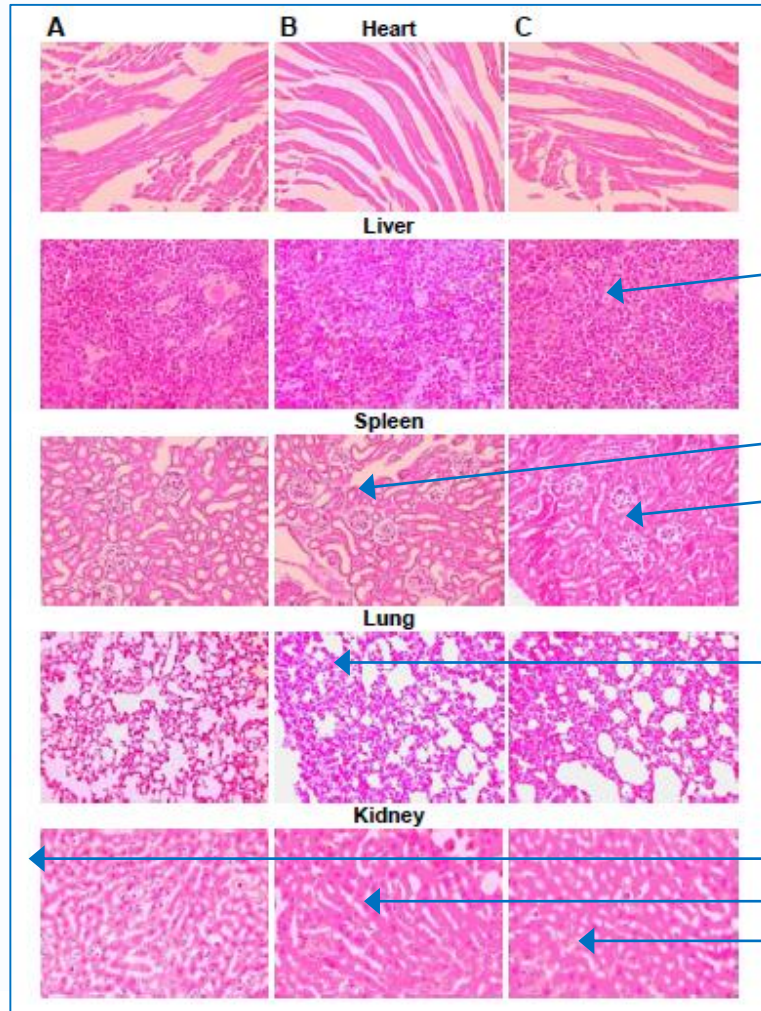
[ИСКАТЬ ЗАИМСТВОВАНИЯ](#)

Решение:

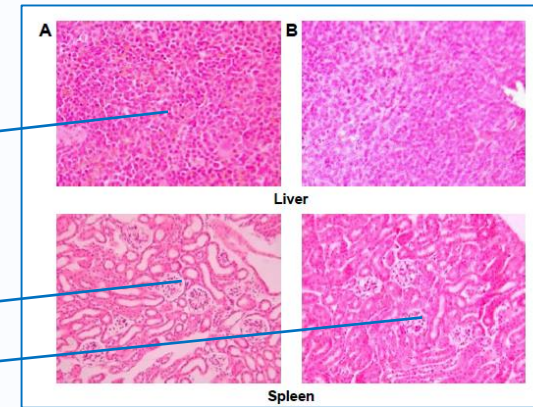
Поиск заимствованных изображений в системе Антиплагиат

Заимствованные изображения

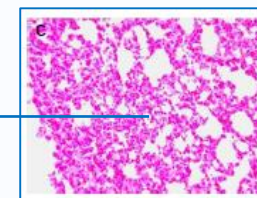
Zhou X, Tao H, Shi KH. Development of a nanoliposomal formulation of erlotinib for lung cancer and in vitro/in vivo antitumoral evaluation. *Drug Des Devel Ther.* 2018;12:1-8 <https://doi.org/10.2147/DDDT.S146925>



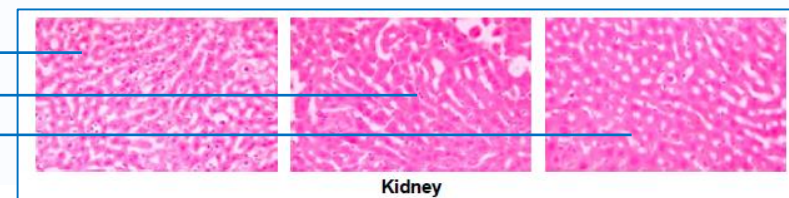
Плагиат изображений



Li S, Wang X. In vitro and in vivo evaluation of novel NGR-modified liposomes containing brucine. *Int J Nanomedicine.* 2017;12:5797-5804 <https://doi.org/10.2147/IJN.S136378>



Li D, Gong L. Preparation of novel pirfenidone microspheres for lung-targeted delivery: in vitro and in vivo study. *Drug Des Devel Ther.* 2016;10:2815-2821 <https://doi.org/10.2147/DDDT.S113670>



Chen J, Jiang H, Wu Y, Li Y, Gao Y. A novel glycyrrhetic acid-modified oxaliplatin liposome for liver-targeting and in vitro/vivo evaluation. *Drug Des Devel Ther.* 2015;9:2265-2275 <https://doi.org/10.2147/DDDT.S81722>

Заимствованные изображения

Фальсификация результатов

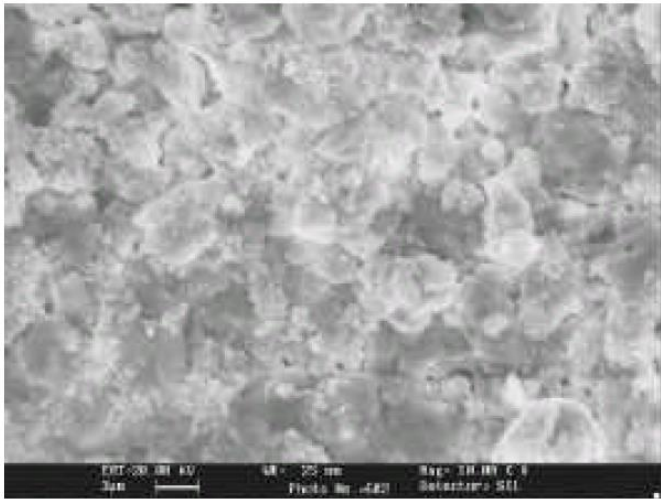


Fig. 5c. SEM image of MgSA6 composite.

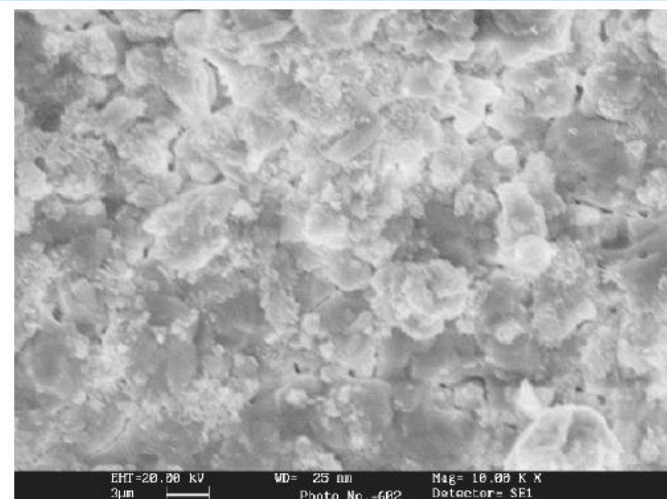
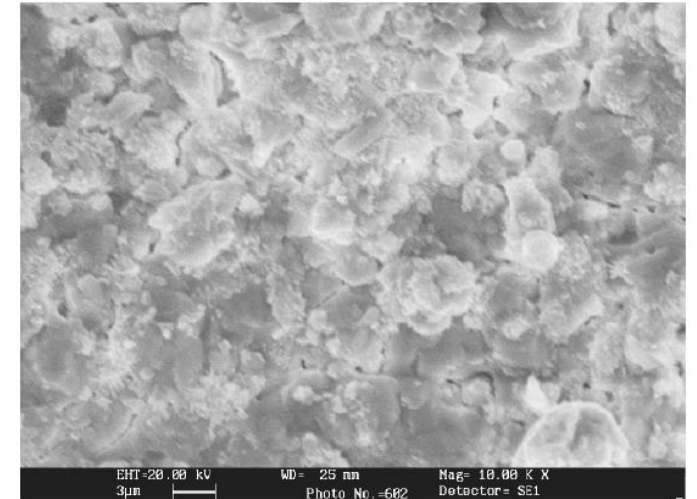


Fig. 5c. SEM image of ZnSA6 composite.



(c)

Fig. 3. SEM image of a) BaSA1, b) BaSA5 and c) SA composite.

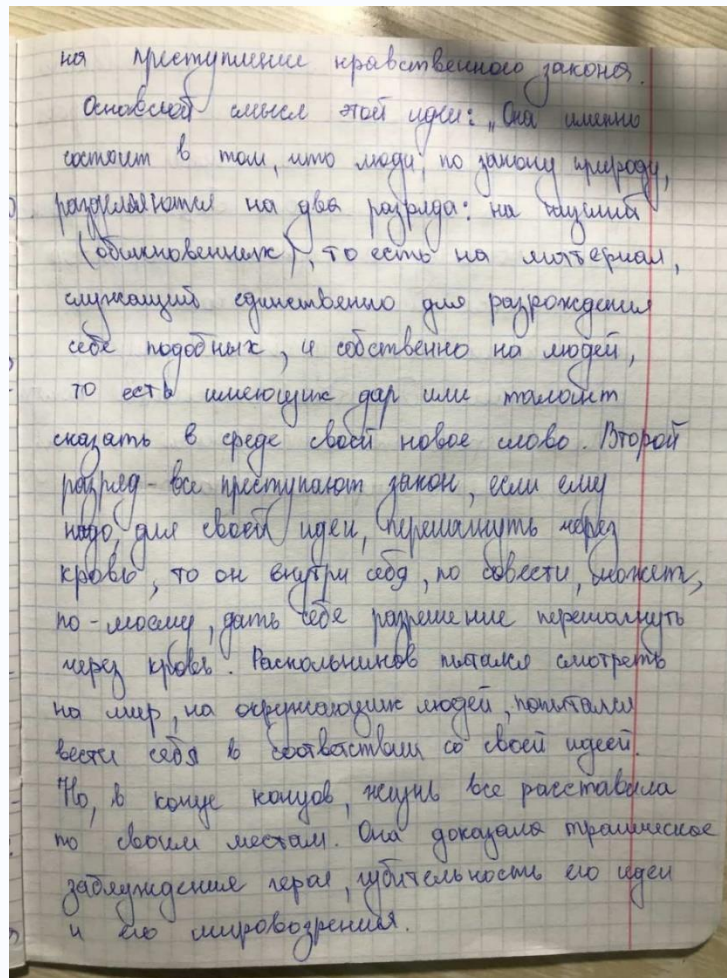
J. JUDITH VIJAYA, L. JOHN KENNEDY, G. SEKARAN, K.S. NAGARAJA
Methanol Sensing Behavior of Strontium(II) Added MgAl₂O₄ Composites Through Solid-State Electrical Conductivity Measurements
Sensors & Transducers Journal, Vol.74, Issue 12, December 2006, pp.864-873

J. JUDITH VIJAYA, L. JOHN KENNEDY, G. SEKARAN, K.S. NAGARAJA
Synthesis, Characterization and Acetone Sensing Properties of Novel Strontium(II)-added ZnAl₂O₄ Composites
Sensors & Transducers Journal, Vol.76, Issue 2, February 2007, pp.1008-1017

B. Jeyaraj, L. John Kennedy, G. Sekaran and J. Judith Vijaya
Benzene and Toluene Vapor Sensing Properties of Sr(II)-added Barium Aluminate Spinel Composites
Sensors & Transducers Journal, Vol. 96, Issue 9, September 2008, pp. 68-80

Рукописные работы

Проблема:
необходимо проверять на
наличие заимствований
рукописные сочинения
школьников

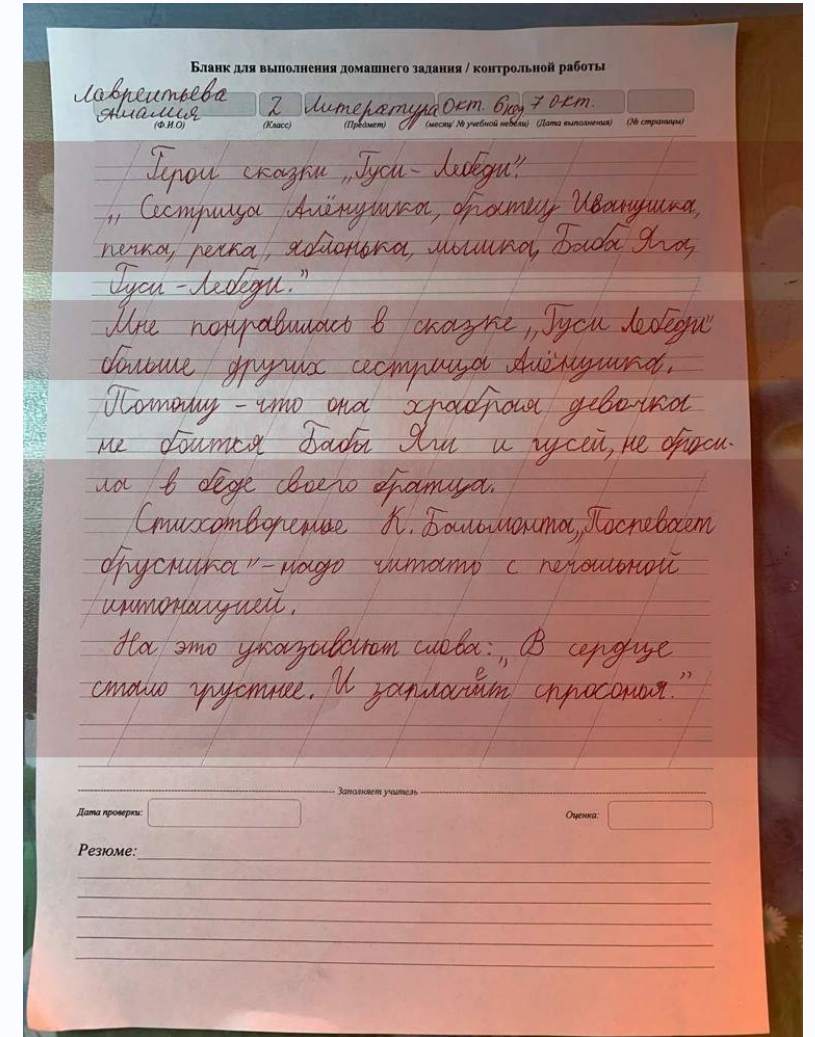
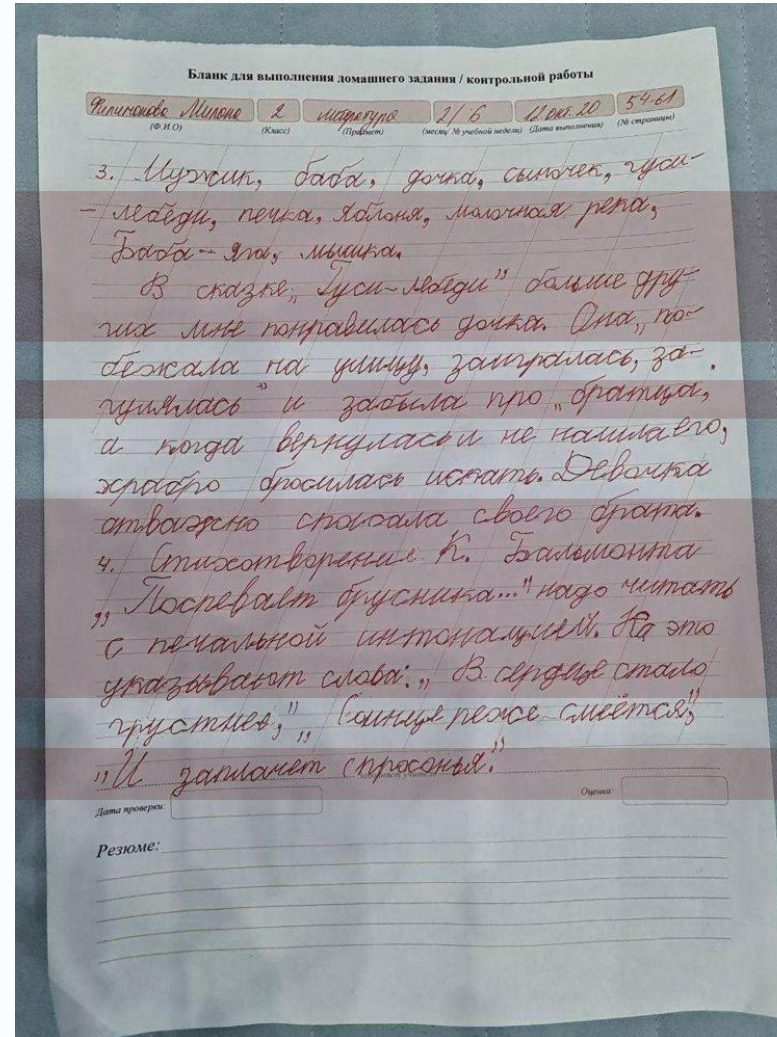


на преступлении нравственного закона.
основной смыслом этой идеи: Она именно
состоит в том, что люди, по закону природы,
разделяются на два разряда; на лучших
(добродетельных), то есть на материал,
служащий единственно для разрождения
себя подобных, и собственно на людей, го
есть имеющих дар или талант сказать в
среде своей новое слово. Второй разряд
все преступают закон, если ему надо, для
своей идеи, перешагнуть через кровь, то он
внутри себя, по совести, может, по-мощи,
дать себе разрешение перешагнуть через
кровь. Раскольники пытаются смотреть на
мир, на окружающих людей, пытаются
вести себя в соответствии со своей идеей. Но,
в конце концов, жизнь все расставала по
своим местам. Она доказала трагическое
заблуждение пера, убийственность его идеи
и ко смывозрения.

Рукописные работы

Решение:

Сервис по проверке рукописных документов на русском языке



Обнаружение переводных заимствований

Проблема:

существенно выросли риски использования неправомерных переводных заимствований

- Системы машинного перевода работают все лучше и лучше
- Можно легко найти подходящий текст и перевести его
- Переводные заимствования сложно обнаруживать – обратный перевод не помогает



Источник изображения:
https://pikabu.ru/story/make_iskusstvo_perevoda_great_again_4706417/author

Обнаружение переводных заимствований

Решение:

Постоянно совершенствуемый сервис по обнаружению переводных заимствований (2017 - 2022)

<https://habr.com/company/antiplagiat/blog/354142/>

- 2017 – русско-английский
- 2018 – англо-русский, казахско-русский, казахско-английский
- 2019 – киргизско-русский, киргизско-английский
- **2022 – 100 ведущих мировых языков в любых сочетаниях**



Выявление искусственного текста





Кадр из фильма Мэла Гибсона «Апокалипсис», 2006
Touchstone Pictures



Кадр из фильма Мэла Гибсона «Апокалипсис», 2006
Touchstone Pictures

Важный фактор: генерируется оригинальный текст

ГЛАВНАЯ / КАБИНЕТ

Кабинет

ПОИСК ПО НАЗВАНИЯМ ДОКУМЕНТОВ

УДАЛЕННЫЕ ДОКУМЕНТЫ 1/1

ПЕРЕПРОВЕРИТЬ ПЕРЕМЕСТИТЬ УДАЛИТЬ ИСТОРИЯ ОТЧЕТОВ В ИЗ ИНДЕКСА

Название	Дата загрузки	Оригинальность	
<input type="checkbox"/> <small>DOCX</small> sample-1	16 Мая 2023 13:35	97,62%	<input type="button" value="ПОСМОТРЕТЬ РЕЗУЛЬТАТЫ"/>
<input type="checkbox"/> <small>DOCX</small> sample-2	16 Мая 2023 13:35	98,16%	<input type="button" value="ПОСМОТРЕТЬ РЕЗУЛЬТАТЫ"/>

ПАПКИ: Все документы

- Корневая папка 46
- Кейс Совета по этике 4
- Яненко 12
- Кейсы GPT 2**

ПРОВЕРЕНО: 16.05.2023 10:42:10

1,2%	СОВПАДЕНИЯ
0%	САМОЦИТИРОВАНИЯ
1,17%	ЦИТИРОВАНИЯ
97,63%	ОРИГИНАЛЬНОСТЬ

ПРОВЕРЕНО: 16.05.2023 10:40:27

0,83%	СОВПАДЕНИЯ
0%	САМОЦИТИРОВАНИЯ
1,01%	ЦИТИРОВАНИЯ
98,16%	ОРИГИНАЛЬНОСТЬ

Детектор искусственных текстов в системе Антиплагиат

АНТИПЛАГИАТ
ОБНАРУЖЕНИЕ ЗАИМСТВОВАНИЙ

Частным пользователям Корпоративным клиентам Обучение

РЕГИСТРАЦИЯ ВОЙТИ

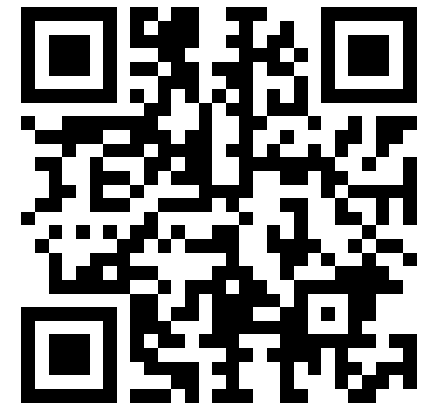
Реестр отечественного ПО Sk

Главная > Новости > Антиплагиат будет находить тексты, созданные ИИ

6 февраля 2023

Антиплагиат будет находить тексты, созданные ИИ

Недавние новости всколыхнули общественность – искусственный интеллект среди нас! В зимнюю сессию 2023 года студент РГУ использовал Chat GPT для создания текста своей дипломной работы. Все бы ничего, но Chat GPT – это специальный софт с применением искусственного интеллекта. Особенность его работы в том, что это не просто программа, а специально созданное программное обеспечение для создания текстов, которые, как кажется на первый взгляд, были написаны человеком, а оригинальность их составляет почти 100%.



<https://www.antiplagiat.ru/news/ai>



Но, черт возьми, как?!



Элементарно!

Выявление искусственных текстов

Кратко:

Машинно-сгенерированные тексты выявляет специальная нейросеть, обученная различать естественные и искусственные тексты

Выявление искусственных текстов

Подробнее:

За кулисами интеллекта ChatGPT: рассказ о том, как определяют тексты, созданные ИИ

🔔 Средний ⌚ 10 мин 👁 15K

Блог компании «Антиплагиат», Программирование*, Алгоритмы*, Машинное обучение*, Natural Language Processing*

Кейс



<https://habr.com/ru/companies/antiplagiat/articles/728112/>

19.10.2023

Детектор искусственных текстов

ГЛАВНАЯ / КАБИНЕТ

Кабинет

[ПРОВЕРИТЬ ДОКУМЕНТ](#)

[ПРОВЕРИТЬ ТЕКСТ](#)

ПАПКИ [Все документы](#)

- Корневая папка 46
- Кейс Совета по этике 4
- Яненко 12
- Кейсы GPT 2**

Поиск по названиям документов

[УДАЛЕННЫЕ ДОКУМЕНТЫ](#) 1/1

[ПЕРЕПРОВЕРИТЬ](#) [ПЕРЕМЕСТИТЬ](#) [УДАЛИТЬ](#) [ИСТОРИЯ ОТЧЕТОВ](#) [В](#) [ИЗ](#) ИНДЕКСА

<input type="checkbox"/>	Название ↕	Дата загрузки ↕	Оригинальность	
<input type="checkbox"/>	DOCX sample-1	16 Мая 2023 13:35	97,62%	ПОСМОТРЕТЬ РЕЗУЛЬТАТЫ
<input type="checkbox"/>	DOCX sample-2	16 Мая 2023 13:35	98,16%	ПОСМОТРЕТЬ РЕЗУЛЬТАТЫ

Детектор искусственных текстов в системе Антиплагиат

АНТИПЛАГИАТ
ОБНАРУЖЕНИЕ ЗАИМСТВОВАНИЙ

Sk Skolkovo
Реестр отечественного ПО

КРАТКОЕ НАИМЕНОВАНИЕ

МЕНЮ ru

ГЛАВНАЯ / КАБИНЕТ / РЕЗУЛЬТАТЫ ПРОВЕРКИ / ПОДОЗРИТЕЛЬНЫЙ ДОКУМЕНТ

Подозрительный документ

1/84

ОТМЕТКА ВКЛЮЧЕНА ВКЛ

На этой странице отображены возможные попытки маскировки заимствований. Будьте внимательны при работе с документом

В ДОКУМЕНТЕ НА СТРАНИЦЕ

В документе обнаружена маскировка заимствований из следующих групп:

- ВСТАВКА** ⓘ В данную группу входят фрагменты, которые, предположительно, созданы с помощью искусственного интеллекта
- СГЕНЕРИРОВАННЫЙ ТЕКСТ** ⓘ ▾

Детектор искусственных текстов в системе Антиплагиат

ГЛАВНАЯ / КАБИНЕТ / РЕЗУЛЬТАТЫ ПРОВЕРКИ

⚠ Внимание, документ подозрительный: в документе присутствует сгенерированный текст. [Подробнее](#)

Оригинальность 54,2% Совпадения 37,92% Цитирования

[ПОЛНЫЙ ОТЧЕТ](#) [КРАТКИЙ ОТЧЕТ](#) [ИСТОРИЯ ОТЧЕТОВ](#) [РАСПЕЧАТАТЬ](#)

В ДОКУМЕНТЕ НА СТРАНИЦЕ

В документе обнаружена маскировка заимствований из следующих групп:

СГЕНЕРИРОВАННЫЙ ТЕКСТ ⓘ ^

Искусственный текст (beta)

На страницах:

[3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#)
[11](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#)
[20](#) [21](#) [22](#) [23](#) [24](#) [25](#) [26](#)
[27](#) [44](#) [45](#) [46](#) [47](#) [48](#)

[Скрыть номера страниц](#)

Детектор искусственных текстов в системе Антиплагиат

↑ НАВЕРХ

5/50

В ДОКУМЕНТЕ НА СТРАНИЦЕ

ПОКАЗАТЬ ВСЕ СКРЫТЬ ВСЕ

СГЕНЕРИРОВАННЫЙ ТЕКСТ ⓘ

Искусственный текст (beta)

Например, на основе финансового анализа можно определить наиболее эффективные способы использования свободных денежных средств компании, принять решение о необходимости привлечения дополнительных финансовых ресурсов, выбрать наиболее оптимальный тип кредитования или выявить причины снижения прибыли компании. [6].

Важно отметить, что финансовый анализ не является единственным инструментом для оценки эффективности деятельности компании. Кроме того, необходимо учитывать многие другие факторы, такие как рыночная конъюнктура, конкуренция, риски и т.д. Однако финансовый анализ является важным компонентом в общей системе управления финансами компании.

Таким образом, сущность и значение финансового анализа заключается в оценке финансового состояния компании, выявлении проблемных зон и определении направлений для дальнейшего развития. Основными задачами финансового анализа являются анализ финансовой устойчивости, рентабельности, ликвидности, дебиторской и кредиторской задолженности, управления активами и капиталом, а также анализ инвестиционной деятельности.

Для проведения финансового анализа необходимо иметь знания и умения в области финансовой отчетности и финансовых показателей, а также опыт в их интерпретации и применении. [7].

Для проведения финансового анализа используются различные методы и инструменты, такие как горизонтальный и вертикальный анализ финансовой отчетности, показатели анализа рентабельности, ликвидности и финансовой устойчивости, денежный поток, анализ дебиторской и кредиторской задолженности, а также сравнение финансовых показателей с аналогичными компаниями на рынке.

Однако, при использовании финансового анализа необходимо учитывать его ограничения. Во-первых, финансовый анализ основывается на финансовой отчетности, которая может содержать ошибки, неполные данные или быть подвержена манипуляциям со стороны руководства компании. Во-вторых, финансовый анализ не учитывает факторы, связанные с внешней средой

Антиплагиат отметил документ как подозрительный. Что делать?



Что делать с подозрительными документами?

- Проблема в том, что нельзя предоставить «доказательства» того, что текст искусственный:
 - Базы данных искусственных текстов не существует
 - Ни человек, ни алгоритм не смогут гарантировать, что какой-то текст написан алгоритмом
 - ChatGPT «склонен присваивать» авторство

Рекомендации

Лучше с самого начала построить работу так, чтобы ИИ не требовался или использовался этично, чем выявлять использование ИИ после сдачи и не знать, что с этим делать.

Легко сказать...

И все-таки рекомендации

На уровне организации: признать возможность использования генеративных инструментов и постараться регламентировать их применение

И все-таки рекомендации

На уровне преподавателей и методистов:

- Контролировать прогресс в подготовке работы шаг за шагом
- Проверять знаком ли автор со своим текстом, обсуждать работу
- Включать работу с генераторами искусственного текста в учебный процесс
- Проверять факты, приведенные в работе, и ссылки из списка источников
- Читать работы студентов

И все-таки рекомендации

На уровне рецензентов и редакторов:

- Устанавливать правила использования инструментов ИИ в научных публикациях
- Внимательно изучать рукописи

Регулирование использования

The screenshot shows the ICMJE website with the following elements:

- Header:** ICMJE INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS. Search bar with "Enter search terms" and "SEARCH" button.
- Navigation:** Recommendations (highlighted), Disclosure of Interest, Journals (Stating That They Follow the ICMJE Recommendations), About ICMJE, News & Editorials.
- Left Sidebar:** Recommendations (with bookmark icon), Browse (About the Recommendations, Roles & Responsibilities, Publishing & Editorial Issues, Manuscript Preparation and Submission), Translations, Archives, Subscribe to Changes.
- Main Content:** Home > Recommendations.

Recommendations

. Text: "Read the Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals. Annotated versions of this and earlier updates to the Recommendations are available in the announcement of the respective updates in ICMJE | News & Editorials." Updated May 2023. A large graphic with a bookmark icon contains the title "Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*" and a "DOWNLOAD" button.
- Bottom of Main Content:** "Download the full PDF to print or view on your computer and/or tablet." → Looking for Translations or Archives?
- Bottom Left:** "KEEP UP-TO-DATE" icon and text: "Request to receive an E-mail when the Recommendations are updated."

<https://www.icmje.org/recommendations/>



Регулирование использования

Because acknowledgment may imply endorsement by acknowledged individuals of a study's data and conclusions, editors are advised to require that the corresponding author obtain written permission to be acknowledged from all acknowledged individuals.

4. Artificial Intelligence (AI)-Assisted Technology

At submission, the journal should require authors to disclose whether they used artificial intelligence (AI)-assisted technologies (such as Large Language Models [LLMs], chatbots, or image creators) in the production of submitted work. Authors who use such technology should describe, in both the cover letter and the submitted work, how they used it. Chatbots (such as ChatGPT) should not be listed as authors because they cannot be responsible for the accuracy, integrity, and originality of the work, and these responsibilities are required for authorship (see Section II.A.1). Therefore, humans are responsible for any submitted material that included the use of AI-assisted technologies. Authors should carefully review and edit the result because AI can generate authoritative-sounding output that can be incorrect, incomplete, or biased. Authors should not list AI and AI-assisted technologies as an author or co-author, nor cite AI as an author. Authors should be able to assert that there is no plagiarism in their paper, including in text and images produced by the AI. Humans must ensure there is appropriate attribution of all quoted material, including full citations.

B. Disclosure of Financial and Non-Financial Relationships and Activities, and Conflicts of Interest

Public trust in the scientific process and the credibility of published articles depend in part on how transparently an author's relationships and activities, directly or topically related to a work, are handled during the planning, implementation, writing, peer review, editing, and publication of scientific work.

The potential for conflict of interest and bias exists when professional judgment concerning a primary interest (such as patients' welfare or the validity of research) may be influenced by a secondary interest (such as finan-

При подаче рукописи журнал должен потребовать от авторов сообщить, использовали ли они технологии с поддержкой искусственного интеллекта (ИИ) (такие как большие языковые модели [LLM], чат-боты или создатели изображений) при подготовке представленной работы. Авторы, использующие такие технологии, должны описать, как в сопроводительном письме, так и в представленной работе, как они их использовали. Чат-боты (такие как ChatGPT) не должны указываться в качестве авторов, поскольку они не могут отвечать за точность, целостность и оригинальность работы, а эти обязанности необходимы для получения авторства (см. раздел II.A.1). Поэтому люди несут ответственность за любой представленный материал, в котором использовались технологии, основанные на искусственном интеллекте. Авторы должны тщательно проверять и редактировать результат, поскольку ИИ может генерировать авторитетно звучащие результаты, которые могут быть неверными, неполными или необъективными. Авторы не должны указывать ИИ и технологии с помощью ИИ в качестве автора или соавтора, а также ссылаться на ИИ как на автора. Авторы должны быть в состоянии утверждать, что в их работе нет плагиата, в том числе в тексте и изображениях, созданных ИИ. Люди должны обеспечить надлежащую атрибуцию всех цитируемых материалов, включая дословное цитирование.

<https://www.icmje.org/recommendations/>

Журнал «Университетская книга»

Издательства
и книжный рынок

Библиотеки

Образование
и наука

ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИЙ ЖУРНАЛ **ОКтябрь 2023**

УНИВЕРСИТЕТСКАЯ

КНИГА®

ДЛЯ ПРИНЯТИЯ РЕШЕНИЙ

УНИВЕРСИТЕТСКАЯ КНИГА № 8, Октябрь 2023



Аркадий ХАЛЮКОВ:
«Держать руку на пульсе науки»

Книжный бизнес — 2023:
ключевые события и перспективы

Креативный суверенитет:
законодательство, образование
и медиаконтент

«Цифра» без иллюзий

Будущее высшей школы:
итоги форума

**«Белый список»
научных изданий:**
альтернативная
наукометрия

Тема номера:
**КНИЖНАЯ ОТРАСЛЬ: ЦИФРОВОЙ ПРОФИЛЬ
И КЛЮЧЕВЫЕ КОМПЕТЕНЦИИ**

ISSN 1726-6726 23008 >
9 771726 672000

www.unkniga.ru
Журнал издаётся с 1996 года

16+

Октябрь 2023

УНИВЕРСИТЕТСКАЯ КНИГА • октябрь 2023

ИННОВАЦИОННЫЕ ТЕХНОЛОГИИ

ВолшебнИИк для страны ОЗ



Юрий ЧЕХОВИЧ,
исполнительный директор
компании Антиплагиат

ОБНАРУЖЕНИЕ ЗАИМСТВОВАНИЙ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Многолетний опыт разработки и использования системы обнаружения заимствований (ОЗ) указывает на неизменно завышенные ожидания пользователей в отношении результатов её работы.

Чаще всего от систем plagiarism detection (устоявшийся англоязычный термин, который проще всего перевести как «детектор плагиата») ожидают возможностей, присущих скорее бытовым детекторам, — простоты использования и понятности результата.

Это очень похоже, например, на глюкометр — прибор, которым пользуются миллионы больных сахарным диабетом для контроля уровня глюкозы в крови: надо получить каплю крови, нанести её на тест-полоску и прибор покажет уровень глюкозы. Далее в соответствии с рекомендациями врача следует решить, принимать лекарство или нет. Всё устроено очень просто, так чтобы прибор могли использовать люди, не обладающие специфическими знаниями и навыками.

От систем ОЗ в целом ожидают чего-то похожего: загружаем документ в систему, получаем оценку «уровня плагиата», принимаем решение. Такие ожидания прослеживаются и в локальных актах российских вузов [1], и в приказах регулятора: «Тексты выпускных квалификационных работ... проверяются на объём заимствования» [2, п. 38]. На оценке доли заимствований в текстах работ основаны решения многих диссертационных советов и научных журналов. Такой подход не является исключительно российским [3]. Зарубежные вузы и журналы широко используют оценки сходства, подобия, уровня плагиата и другие аналогичные метрики для принятия решения о качестве работы.

Многократно обсуждалось, что как решение проблемы этических нарушений в научных и учебных работах в целом, так и оценка качества каждого конкретного документа не могут быть сведены к определению допустимого «уровня плагиата» и измерению его в каждом проверяемом тексте [4].

Сервисы ОЗ создавались как специализированные поисковые системы. Отличия от обычных поисковиков заключались в длине поискового запроса, как правило составляющего десятки и сотни тысяч символов и в формате представления результата, позволяющем пользователю понять, с каким источником совпадает тот или иной фрагмент текста (так называемая раскраска). Очевидно, что

работа с такими инструментами даёт возможность понять, есть ли в документе непропорциональные заимствования, но требует при этом определённой квалификации проверяющих, наличия апробированных методик и затрат времени на выполнение каждого анализа. Не менее очевидно, что включение в процесс оценки экспертов повышает роль таких субъективных факторов, как уровень квалификации, внимательность, возможность конфликта интересов. Трудно исключить коррупционное влияние. Кроме того, в целом ряде ситуаций на экспертный анализ работ просто не хватает времени.

Всё это создаёт запрос на разработку идеального детектора плагиата, который в полностью автоматическом режиме может дать заключение, хорошая работа или плохая. При этом, с одной стороны, необходимо иметь возможность получать решение по работе при минимальном участии пользователей или вообще без этого участия, с другой — если возникают сомнения, должны быть средства, которые позволяют полностью «препарировать» результат автоматической проверки.



Иллюстрация создана Юлией Чехович при помощи сервиса Kandinsky (<https://kandinsky.ai.ru/>).

«Быстрое развитие технологий машинного обучения и искусственного интеллекта (ИИ) превращает несбыточные мечты в набор конкретных и посильных задач. Решение этих задач и встраивание полученных решений в конвейер проверки документов позволяет создать инструмент, в значительной степени отвечающий в глазах пользователя образу идеального детектора»

www.unkniga.ru

61

Спасибо за внимание!

Ваши вопросы

Юрий Чехович,

Исполнительный директор компании Антиплагиат, к.ф.-м.н.

chehovich@antiplagiat.ru

+7 495 223 23 84

8 800 777 81 28