



Обзор возможностей и технологий реализации систем антиплагиата

Е.М. Павлов, А.В. Рыжов, С.А. Петров

Национальный исследовательский университет «МЭИ», Москва

Аннотация: В данной статье проводится анализ и обзор современных методов и технологий, используемых в системах антиплагиата, с акцентом на российский рынок. Целью рассмотрения всего вышеперечисленного является выбор подходящей для интеграции системы антиплагиата. В статье представлены наиболее популярные российские сервисы для детекции заимствований, их бизнес-модели, алгоритмы работы, а также представлено общее описание принципов и механизмов, лежащих в основе этих алгоритмов. Было определено, что наиболее универсальной и эффективной системой поиска заимствований является сервис Antiplagiat.ru, так как в нем присутствует возможность интеграции через API, а также 34 дополнительных модуля, предоставляющих возможность адаптировать функционал системы под индивидуальные потребности.

Ключевые слова: антиплагиат, анализ текста, алгоритмы обработки текста, семантический анализ, стилистический анализ.

Введение

В современном мире, где информация становится одним из ключевых ресурсов, вопросы интеллектуальной собственности и сохранения авторских прав приобретают особое значение. Плагиат — одна из наиболее актуальных проблем в образовательной и научной среде — представляет собой использование чужой работы без должного признания или ссылки на первоначального автора. Этот феномен негативно влияет на качество образования, искажает результаты исследований и нарушает этические нормы.

С развитием цифровых технологий и расширением доступа к информации через интернет открылись более широкие возможности заимствования и использования чужого контента. Наиболее громким случаем стала история в 2014 году, когда журналисты портала "Диссернет" выявили многочисленные случаи плагиата в диссертациях некоторых высокопоставленных чиновников и политиков. Этот скандал поднял вопросы о целостности академического сообщества и привел к ужесточению требований к диссертационным работам [1]. В ответ на это появилась

потребность в разработке и внедрении эффективных систем антиплагиата, способных выявлять и предотвращать использование необоснованных заимствований в научных и образовательных работах.

Системы антиплагиата стали неотъемлемой частью образовательного процесса во многих учебных заведениях по всему миру. Они обеспечивают проверку студенческих и научных работ на наличие плагиата, помогают ученым и преподавателям поддерживать академическую честность и высокий уровень качества исследований.

Во многих ВУЗах обязательно проводится проверка всех студенческих и научных работ на наличие плагиата с использованием этих систем. Они помогают ученым и преподавателям поддерживать академическую честность и высокий уровень качества исследований.

Методы детекции плагиата

Для выявления плагиата в текстовых материалах применяются разнообразные методы, каждый из которых имеет свои особенности и области применения [2].

Одним из основных методов детекции плагиата является поиск похожих фраз [3]. Этот метод основан на прямом сравнении строк из проверяемого текста с строками из базы данных. Он подходит для выявления прямого копирования без изменений. Данный метод является одним из наиболее быстрых, но менее эффективен при маскировке плагиата (например, замене синонимами).

В случае работы с измененным текстом используется более продвинутый метод семантического анализа, который фокусируется на определении смысла фраз и предложений, несмотря на изменение отдельных слов или порядка слов. В нём применяются технологии обработки естественного языка (NLP) для интерпретации контента. Семантический

анализ позволяет обнаруживать “переформулированный” плагиат, где структура или слова изменены, но смысл сохраняется [4].

Помимо смысловой нагрузки в тексте, немалую роль в определении заимствований играет стилистика текста. В связи с этим используется стилистический анализ, основанный на идее, что каждый автор имеет уникальный стиль написания, который можно определить по выбору слов, грамматике, структуре предложений и другим параметрам. При таком подходе проверяемый текст сравнивается с известными работами автора для определения несоответствий. Анализ стилистики полезен для выявления плагиата в долгосрочной перспективе, особенно когда подозревается, что работа написана другим человеком [5].

В дополнении к вышеописанным методам используется метод сравнения документов. При его использовании сравнивается два или более документа целиком для определения сходства. Метод может быть полезен при сравнении нескольких версий одного документа или при анализе работ, представленных разными студентами на одну тему.

Одним из самых популярных методов детекции плагиата является сетевой анализ. Он подразумевает сравнение текста с содержимым веб-страниц, найденных в сети Интернет. Данный метод решает проблему заимствований из онлайн-источников, позволяя выявлять материалы на их основе.

В современных системах используются сразу несколько из вышеупомянутых методов для повышения точности и надежности результатов. При этом состав этих методов варьируется в зависимости от целей и специфики задачи [6].

Российские системы антиплагиата

В России, как и во многих других странах, проблема плагиата в академической и научной среде приобрела особое внимание. Однако

зарубежные системы выявления заимствований либо не поддерживают русский язык, либо не учитывают его специфику в полной мере. Это привело к созданию ряда систем антиплагиата, которые активно используются в учебных заведениях и научных учреждениях страны. Сравнение различных российских систем антиплагиата представлено ниже (Таблица № 1).

Таблица № 1

Сравнение российских систем антиплагиата

Название	Тариф	Методы детекции
Antiplagiat.ru	Бесплатный/платный	Поиск похожих фраз Сетевой анализ Семантический анализ
Text.ru	Бесплатный	Поиск похожих фраз Сетевой анализ Семантический анализ
Advego Plagiat	Бесплатный (поддержка прекращена)	Поиск похожих фраз Сетевой анализ

Одной из наиболее известных и широко используемых в российских ВУЗ'ах систем является antiplagiat.ru. Она имеет собственную поисковую систему и предоставляет возможность проверки научных работ, дипломов и курсовых на соответствие академической этике. Принцип работы заключается в сравнении текстов с базой данных, включая научные работы, публикации и интернет-ресурсы. Кроме того, antiplagiat.ru имеет OCR-модуль, считывающий текстовые данные с изображений [7].

Само сравнение текстов с потенциальными источниками можно разделить на несколько этапов:

1. Вычленение слов из предложений и удаление лишних символов с последующей лемматизацией каждого слова.
2. Хеширование (превращение слова в целые числа). Формирование шинглов из хеш-кодов с повторным хешированием. Шингл – это небольшой

фрагмент текста, подготавливаемый особым образом для анализа уникальности.

3. Получение упорядоченного массива чисел путем сортировки шинглов.

4. Поиск по индексу с помощью функции, превращающей хеши в документы, отранжированные по убыванию числа совпадений [8].

Сервис также предоставляет возможность открытой интеграции с API [9]. Это означает, что поиск заимствований можно встроить именно туда, где он необходим, вызвать, когда это требуется пользователю, проверять только необходимые тексты и показывать результаты так, как того требует логика информационной системы. К примерам успешной интеграции относится [7]:

- Интеграция с системой авторизации ВУЗа (ЕНУ им. Гумилева, ДНТП).
- Сервис проверки через API (ТПУ, Экономфак МГУ).
- Интеграция с СДО «Moodle» (ВГУ, ТюмГУ, РГАЗУ).
- Проверка документов по e-mail (МФЮА, СГТУ, НГУЭУ).

Другой известной системой антиплагиата является text.ru, который также является сервисом для проверки орфографии и стилистики текста. Данная система пользуется популярностью среди студентов, преподавателей и журналистов. В отличие от других систем пользователи могут оставлять комментарии к проверенным текстам и обмениваться отзывами. На сайте системы доступен инструмент для оптимизации SEO-текстов [10].

Text.ru использует схожий с antiplagiat.ru алгоритм сравнения источников (с помощью шинглов). Однако в нем сначала производится канонизация текста – из него удаляются все несущественные для смысла и грамматики слова (союзы, предлоги, междометия и т.д.), а также знаки препинания. Чем меньше шингл или фрагмент (ячейка для анализа) текста, тем точнее можно определить уникальность и элементы плагиата [11].

Text.ru также предлагает API для интеграции с другими платформами и сервисами [10].

Однако пользователи отмечают ряд недостатков данного сервиса:

- Ограниченное количество проверок. Относится это только к неавторизованным пользователям, у зарегистрированных этой проблемы нет.
- Ограничение по числу знаков. За один раз система проверяет не более 15 тысяч знаков. [12]

Advego Plagiatus предназначен для быстрой проверки текстов на плагиат и особенно полезен для авторов веб-контента и редакторов. Advego, также как и text.ru, не только определяет степень уникальности текста, но и предоставляет инструменты для его оптимизации. Сервис включает в себя ряд дополнительных функций, таких как проверка орфографии, пунктуации и семантического ядра текста [13].

Текст статьи в Advego анализируется сразу по нескольким алгоритмам:

Алгоритм шинглов — проверяются точные совпадения фраз, находятся источники копипаста, а также страницы, на которых размещены тексты с такими же словосочетаниями и предложениями [14].

Алгоритм лексических совпадений — проверяется схожесть набора лемм, терминов и значимых слов, находятся источники рерайтинга, а также страницы, совпадающие по тематике с проверяемой статьей.

Алгоритм псевдоуникализации — проверяется наличие сторонних символов и признаков обработки текста сервисами "повышения" уникальности [15].

Все три системы активно используются в России и зарекомендовали себя как надежные инструменты для выявления плагиата. Они предоставляют разнообразные функции, удовлетворяя потребности различных групп пользователей – от студентов и преподавателей до профессиональных авторов и редакторов [3].

Заключение

Системы детекции плагиата играют ключевую роль в сохранении академической и профессиональной честности, а также защите авторских прав. Они разнообразны, базируются на различных методологиях и предоставляют широкий спектр инструментов для анализа текстов.

Выбор конкретной системы антиплагиата в значительной степени зависит от конкретных задач и потребностей пользователя. Однако в общем случае Antiplagiat.ru находится на лидирующих позициях по сравнению с двумя другими рассмотренными сервисами. В первую очередь его эффективность достигается за счет собственной поисковой системы и наличия таких опций как OCR (оптическое распознавание), кольцо ВУЗов (коллекция документов университетов) и 32 других модулей [7]. Также, как и у Text.ru, присутствует интеграция по API. Доступные модули позволяют адаптировать функционал системы под индивидуальные потребности клиентов, будь то преподаватели, студенты или профессиональные писатели.

Text.ru же свою очередь имеет инструменты SEO-оптимизации, которые не являются настолько востребованными, как основной функционал сервиса. Его проблемой является ограничение на объем текста в стандартной версии. В Advego Plagiatus тоже присутствуют инструменты SEO-оптимизации, но отсутствует интеграция по API.

Литература

1. Министры и диссертации: разоблачения академического плагиата и их политические последствия в России и ЕС. URL: dissernet.org/media/ministry_i_dissertatsii_razoblacheniya_akademicheskogo_plagiata_i_ikh_politicheskiye_posledstviya_v_rossii.
2. Красников И.А., Никуличев Н.Н. Гибридный алгоритм классификации текстовых документов на основе анализа внутренней



связности текста // Инженерный вестник Дона, 2013, №3. URL: ivdon.ru/ru/magazine/archive/n3y2013/1773.

3. Севостьянова, И. О. Обзор систем проведения проверки на плагиат: общероссийские и зарубежные // Научное обозрение. Педагогические науки. – 2017. – № 5. – С. 162-166.

4. Джангаров А.И., Ахметов Х.А. Модель поиска смысла текстовых фрагментов // Инженерный вестник Дона, 2019, №4. URL: ivdon.ru/ru/magazine/archive/n4y2019/5877.

5. Хренов, А. Н. Использование методов семантического анализа текстов для выявления плагиата // XX Всероссийская студенческая научно-практическая конференция Нижневартковского государственного университета : сборник статей, Нижневартовск, 03–04 апреля 2018 года / Ответственный редактор А.В. Коричко. Том Часть 2. – Нижневартовск: Нижневартковский государственный университет, 2018. – С. 202-205.

6. Чиркин, Е. С. Использование систем антиплагиата в образовании // Вестник Тамбовского университета (Вестник российских университетов. Математика). Серия: Естественные и технические науки. – 2013. – Т. 18, № 6-2. – С. 3380-3387.

7. Антиплагиат – российская система обнаружения текстовых заимствований. URL: antiplagiat.ru/.

8. Так устроен поиск заимствований в Антиплагиате. URL: habr.com/ru/companies/antiplagiat/articles/429634/.

9. Документация ApiCorp. URL: docs.antiplagiat.ru/ru/api/.

10. Куций, О. Я. Проблемы проверки выпускных квалификационных работ на объем заимствованного текста // Alma Mater (Вестник высшей школы). – 2021. – № 1. – С. 52-56. – DOI 10.20339/AM.01-21.052.

11. Можно ли верить проверке уникальности текста на сайте text.ru? URL: maxtop.org/mozhno-li-verit-proverke-unikalnosti-teksta-na-sajte-text-ru/.

12. Пошаговая инструкция, как проверить уникальность текста на Текст.ру. URL: softolet.ru/content/tekst-ru-proverka-unikalnosti.html.

13. Advego Plagiatus – проверка уникальности текста. URL: advego.ru/plagiatus/.

14. Как работает Адвего Плагиатус. URL: advego.com/blog/read/news/753820/all1/.

15. Программа для проверки уникальности текста Advego Plagiatus URL: dev.advego.org/plagiatus/.

References

1. Ministry` i dissertacii: razoblacheniya akademicheskogo plagiata i ix politicheskie posledstviya v Rossii i ES [Ministers and dissertations: revelations of academic plagiarism and their political implications in Russia and the EU]. URL: dissernet.org/media/ministry_i_dissertatsii_razoblacheniya_akademicheskogo_plagiata_i_ikh_politicheskiye_posledstviya_v_rossii.

2. Krasnikov I.A., Nikulichev N.N. Inzhenernyj vestnik Dona, 2013, №3. URL: ivdon.ru/ru/magazine/archive/n3y2013/1773.

3. Sevost`yanova, I. O. Nauchnoe obozrenie. Pedagogicheskie nauki. 2017. № 5. pp. 162-166.

4. Dzhangarov A.I., Axmetov X.A. Inzhenernyj vestnik Dona, 2019, №4. URL: ivdon.ru/ru/magazine/archive/n4y2019/5877.

5. Xrenov, A. N. Ispol`zovanie metodov semanticheskogo analiza tekstov dlya vy`yavleniya plagiata [Using methods of semantic analysis of texts to detect plagiarism]. XX Vserossijskaya studencheskaya nauchno-prakticheskaya konferenciya Nizhnevartovskogo gosudarstvennogo universiteta: sbornik statej, Nizhnevartovsk, 03–04 aprelya 2018 goda. Otvetstvenny`j redaktor A.V. Korichko. Tom Chast` 2. Nizhnevartovsk: Nizhnevartovskij gosudarstvenny`j universitet, 2018. pp. 202-205.

6. Chirkin, E. S. Vestnik Tambovskogo universiteta (Vestnik rossijskix universitetov. Matematika). Seriya: Estestvenny`e i texnicheskie nauki. 2013. T. 18, № 6-2. pp. 3380-3387.
7. Antiplagiat – rossijskaya sistema obnaruzheniya tekstovy`x zaimstvovaniy [Antiplagiat is a Russian system for detecting textual borrowings]. URL: antiplagiat.ru/.
8. Tak ustroen poisk zaimstvovaniy v Antiplagiate [This is how the Anti-Plagiarism search is organized]. URL: habr.com/ru/companies/antiplagiat/articles/429634/.
9. Dokumentaciya ApiCorp [ApiCorp documentation]. URL: docs.antiplagiat.ru/ru/api/.
10. Kucyj, O. Ya. Alma Mater (Vestnik vy`sshej shkoly`). 2021. № 1. pp. 52-56. DOI 10.20339/AM.01-21.052.
11. Mozhno li verit` proverke unikal`nosti teksta na sajte text.ru? [Can we trust the text uniqueness check on text.ru?]. URL: maxtop.org/mozhno-li-verit-proverke-unikalnosti-teksta-na-sajte-text-ru/.
12. Poshagovaya instrukciya, kak proverit` unikal`nost` teksta na Tekst.ru [Step-by-step instructions on how to check the uniqueness of text on Text.ru]. URL: softolet.ru/content/tekst-ru-proverka-unikalnosti.html.
13. Advego Plagiatus – proverka unikal`nosti teksta [Advego Plagiatus - text uniqueness checker]. URL: advego.ru/plagiatus/.
14. Kak rabotaet Advego Plagiatus [How Advego Plagiatus works]. URL: advego.com/blog/read/news/753820/all1/.
15. Programma dlya proverki unikal`nosti teksta Advego Plagiatus [Advego Plagiatus text uniqueness checker program]. URL: dev.advego.org/plagiatus/.

Дата поступления: 29.10.2023

Дата публикации: 13.12.2023
