

УДК 37.026.1: 004.043

ОБЗОР СИСТЕМ ПРОВЕДЕНИЯ ПРОВЕРКИ НА ПЛАГИАТ: ОБЩЕРОССИЙСКИЕ И ЗАРУБЕЖНЫЕ

Севостьянова И.О.

ФГБОУ ВО «Магнитогорский государственный технический университет им. Г.И. Носова»,
Магнитогорск, e-mail: irinasevostyano@yandex.ru

Обозначены проблемы, связанные с нарушением авторского права, в частности плагиат. Рассматривается понятие плагиата и средства защиты от него. Для защиты от плагиата существуют различные методы: дактилоскопия; анализ встречаемости термина; выявление плагиата, основанное на цитировании; изучение языковых стилей. Эти методы являются действенными, некоторые проводят проверку текста по частям, а другие – полностью весь текст. Но самыми популярными средствами для проверки текста на плагиат являются онлайн-сервисы и программы, которые называются «системами антиплагиата». Многие страны имеют свои системы антиплагиата («Copyscape», «Istio», «Articlechecker», «Duplichecker»), это связано с тем, что каждая страна имеет свой язык. Например, в России все работы и обучение проходит на русском языке. А зарубежные системы для проверки плагиата имеют в своих базах в основном издания, написанные на их национальном языке. Поэтому в России имеются свои системы для проверки плагиата, самые известные из них: «Антиплагиат», «Детектор плагиата», «Advego Plagiatus», «Etxt Антиплагиат», «Плагиат.НЕТ». Несмотря на большое количество различных систем и программ, самым точным методом для выявления нарушения авторского права является экспертиза. Экспертиза хоть и дает высокий результат проверки, но является дорогостоящей процедурой, которая требует работы специально обученных людей.

Ключевые слова: нарушение авторского права, плагиат, антиплагиат, обнаружение плагиата, российские системы антиплагиата, зарубежные системы антиплагиата

REVIEW OF SYSTEMS OF INSPECTION TO PLAGIARISM: ALL-RUSSIAN AND FOREIGN

Sevostyanova I.O.

Nosov Magnitogorsk state technical University, Magnitogorsk, e-mail: irinasevostyano@yandex.ru

The problems related to the infringement of copyright, especially plagiarism. The concept of plagiarism and means of protection against it. To prevent plagiarism, there are various techniques: fingerprinting; analysis of the occurrence of the term; plagiarism detection based on citation; study of language styles. These methods are effective, some are checking text in parts, others - completely whole text. But the most popular means to check the text for plagiarism are online services and programs, which are called «Anti-plagiarism system». Many countries have their own Anti-plagiarism system («Copyscape», «Istio», «Articlechecker», «Duplichecker»), due to the fact that each country has its own language. For example, in Russia all the work and training takes place in Russian. And foreign systems to check for plagiarism have in their bases in the main edition, written in their national language. Therefore, Russia has its own system to check for plagiarism, the most famous of them: «Antiplagiat», «Plagiarism Detector», «Advego Plagiatus», «Etxt Antiplagiat», «Plagiata.NET». Despite the large number of different systems and programs, the most accurate method for the detection of copyright infringement is expertise. Expertise though gives a high test result, but it is an expensive procedure that requires the work of specially trained people.

Keywords: copyright infringement, plagiarism, plagiarism, plagiarism detection, the Russian system of Antiplagiat, foreign system of Antiplagiat

С появлением Интернета жизнь людей сильно изменилась. В настоящее время нет необходимости проводить много времени за книгами в поисках необходимой информации. Люди получили неограниченный доступ практически к любой информации, что позволяет получать знания и использовать их по своему усмотрению [3]. Но человек, который делится в Интернете какой-то информацией или своими знаниями, не имеет защиты от незаконного копирования и распространения. Информация, попадая в Интернет, становится общедоступной, а вопрос о соблюдении авторского права на сегодняшний день остается нерешенным [6].

Нарушение авторского права – это правонарушение, суть которого составляет использование произведений науки, литературы и искусства, охраняемых авторским

правом, без разрешения авторов или правообладателей или с нарушением условий договора об использовании таких произведений. К числу основных способов нарушения авторских прав относится незаконное копирование и распространение произведения, а также плагиат [2].

Плагиат – это заимствование чужой работы, будь то преднамеренно или непреднамеренно.

Плагиат – это использование, перефразирование и подведение итогов работы в любой форме без подтверждения ссылками на источники и представление её как своей собственной работы [1].

Неограниченный доступ к любой литературе, огромное количество публикаций в Интернете, большое число различных изда-

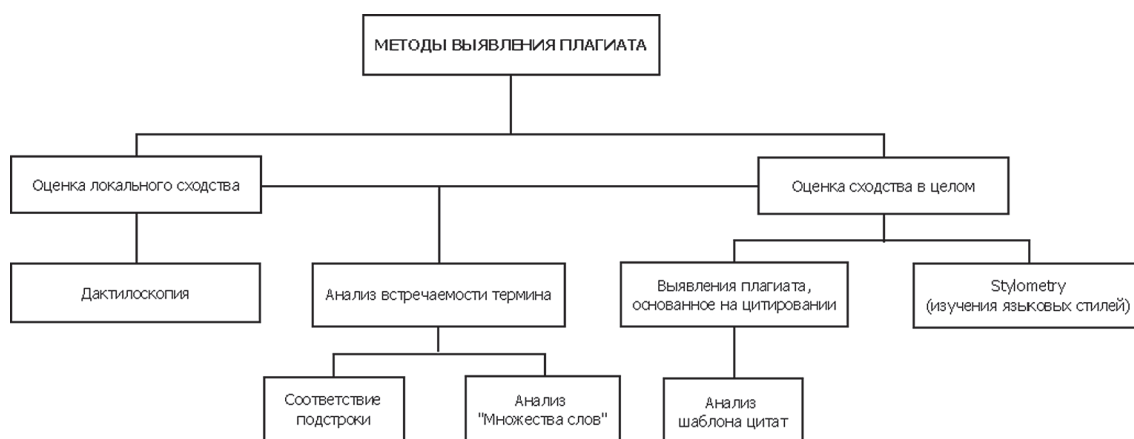


Рис. 1. Методы для обнаружения плагиата

ний – все это усложняет процесс определения первоначального автора [5].

Поэтому на сегодняшний день для борьбы с плагиатом в статьях, работах студентов и преподавателей, а также в других изданиях существуют специальные сервисы и программы. Например: онлайн-сервисы – Copyscape, Miratools, Антиплагиат, Istio.com, программы – Advego Plagiatus, Praide unique content analyzer, Double Content Finder.

Помимо сервисов и программ имеются специальные компьютерные методы для обнаружения плагиата (рисунок 1).

Методы для обнаружения плагиата можно разделить по типу оценки сходства: оценка локального сходства и оценка глобального сходства (таблица 1) [11].

Для компьютерного выявления плагиата существуют различные способы, в таблице 2 представлены некоторые из них [15].

Самым действенным способом для установления плагиата в РФ является экспертиза. Суть экспертизы заключается в сравнении нескольких текстов. При сравнении высчитывается процент контекстуальных совпадений, перефразирования и других форм переработки текста для определения факта, является ли текст оригинальным или нет. Экспертизу проводят специально обу-

ченные эксперты, в основном по решению суда.

Но экспертиза является дорогой и длительной процедурой. Поэтому для установления плагиата многие используют более простые и быстрые способы – онлайн-сервисы и программы. Такие системы получили название «системы антиплагиата».

Принципы работы систем антиплагиата

Системы антиплагиата используют следующие системы для поиска заимствований:

1) метапоисковые системы – системы, которые не имеют собственной базы документов. Метапоисковые системы отправляют запрос в поисковые системы сети Интернет, а также проверяют документ специальными методами (например, метод выборки);

2) поисковые системы сети Интернет – системы, которые не позволяют обнаружить заимствования. Но некоторые поисковые системы (например, Google) позволяют найти статьи, которые опубликованы в мировых изданиях, и цитирования в них, хотя и не обозначенные как заимствования;

3) специализированные системы – системы, которые имеют собственный алгоритм для поиска заимствований и свою базу документов.

Таблица 1

Методы выявления плагиата

Критерии сравнения	Оценка локального сходства	Оценка сходства в целом (оценка глобального сходства)
Описание	Для обнаружения плагиата используется небольшая часть текста.	Для обнаружения плагиата используются значительная часть или весь текст [9].
Способы выявления плагиата	дактилоскопия; анализ встречаемости термина.	анализ встречаемости термина; выявление плагиата, основанное на цитировании; изучение языковых стилей.

Таблица 2

Способы выявления плагиата

Способы выявления плагиата	Описание
Метод дактилоскопии	Метод дактилоскопии заключается в сравнении рассматриваемого документа с «отпечатками». «Отпечатками» является набор из нескольких строк, которые позволяют описать содержание текста и выбираются из некоторого количества текстов. При сравнении все найденные совпадения указывают на общие части текста.
Метод анализа «множества слов»	Метод анализа «множества слов» заключается в сравнении эталонного множества слов с двумя исследуемыми текстами. Для проверки текстов с помощью данного метода сначала с помощью специального алгоритма производится предварительная обработка текста, то есть удаляются все слова, которые не несут никакой смысловой нагрузки в тексте. После этого происходит подсчет количество пересечения слов, которые одновременно имеются в эталонном множестве и в проверяемом тексте. Для генерации эталонного множества существует 3 способа: статическое множество слов, созданное на основании научных текстов по нескольким направлениям; статическое множество слов, созданное на основании текстов по определенному направлению; динамическое формирование слов [7].
Анализ шаблона цитат	Анализ шаблона цитат является способом, который представляет собой подпоследовательности, содержащие не только общие цитаты для двух документов, но и подобный порядок и близость цитат в тексте, являющихся основными критериями для определения шаблона цитат.
Метод изучения языковых стилей	Метод изучения языковых стилей является статическим методом, который заключается в сравнении моделей. Стилометрические модели — модели, построенные для разных фрагментов текста, которые стилистически отличаются друг от друга.

Отличительной чертой метапоисковых систем является то, что они имеют доступ только к документам, которые имеются в свободном доступе. Но даже если доку-

мент находится в открытом доступе, то это не гарантия того, что он будет найден, так как часть форматов не поддерживается [10]. Еще одной особенностью таких систем яв-

Таблица 3

Популярные российские и зарубежные системы поиска плагиата

Наименование системы для поиска плагиата	Оплата за использование системы	Пользователи системы	Описание системы
Российские системы для поиска плагиата			
Антиплагиат (Антиплагиат. Ру)	Имеются бесплатные и платные версии. Бесплатная версия имеет ограничения функциональности, а платная – имеет расширенный функционал.	Пользоваться системой могут любые зарегистрированные пользователи, а также высшие учебные заведения.	Система имеет набор услуг, которые реализуют технологию поиска нарушения авторских прав с помощью проверки текстовых документов на наличие заимствований из открытых источников сети Интернет и других источников [12].
Detector-plagiata.ru («Детектор плагиата»)	Система имеет бесплатные и платные версии.	Систему в основном используют издательства, учебные заведения, владельцы информационных и web-ресурсов, юридические организации и практикующие юристы, специализирующиеся на защите интеллектуальной собственности, авторы и правообладатели любых текстовых материалов.	Система, направленная на выявление заимствований текста, осуществляет поиск плагиата не только в Интернете, но и в документах, расположенных на жестком диске или съемном носителе. Отличительной чертой «Детектор плагиата» является возможность поиска контрафактного использования произведений, находящихся под авторскими правами.

Окончание табл. 3

Наименование системы для поиска плагиата	Оплата за использование системы	Пользователи системы	Описание системы
Advego Plagiatus	Проверка уникальности текста является бесплатным.	Систему могут использовать любые пользователи.	Advego Plagiatus — программа, которая проводит проверку на поиск плагиата. Данная система позволяет выбирать тип проверки: глубокая или обычная. А также в данной программе есть возможность выбора поисковых систем, в которых будет проводиться проверка [14].
Etxt Антиплагиат	Система является бесплатной.	Пользоваться системой могут любые пользователи.	Etxt Антиплагиат — система для проверки уникальности текста. Проверку текста с помощью Etxt Антиплагиат можно осуществить 2 способами: скачав программу на сайте или использовав онлайн-версию.
Плагиата.НЕТ	Программа является бесплатной.	Систему могут использовать все пользователи.	Плагиата.НЕТ — программа для поиска плагиата. Принцип работы данной программы заключается в разбиении статьи на фразы и их поиска в Интернете.
Зарубежные системы для поиска плагиата			
Coryscape [13]	Система является бесплатной, но имеются платные версии для автоматической проверки.	Пользоваться системой могут любые пользователи, но ограниченное число раз. Но для безграничного количества попыток необходимо пройти регистрацию.	Coryscape — система, которая осуществляет проверку на наличие заимствований с использованием таких поисковых систем, как Google и Yahoo!
Istio	Существует платные и бесплатные версии Istio	Системой могут пользоваться все незарегистрированные пользователи.	Istio.com — система для анализа документов, которая позволяет проводить проверку на наличие заимствований и показывает сайты, откуда был скопирован текст.
Articlechecker	Сервис является бесплатным	Система доступна для всех пользователей	Articlechecker — онлайн-сервис, позволяющий проверять докуменеты с помощью Google или Yahoo поиска.
Duplichecker	Система является бесплатной	Систему могут использовать все зарегистрированные пользователи. Если пользователь не зарегистрировался, то он имеет ограниченное количество проверок текста.	Duplichecker — онлайн сервис, который проверяет каждый фрагмент текста на построчной основе для поиска в нем плагиата.

ляется то, что проверка может быть невозможна или ее результаты будут неверны. Это происходит из-за неправомерного использования поисковой системы автоматизированными средствами и в результате возможна блокировка поисковой системы по IP.

А отличительной чертой специализированных систем является их самодостаточность, то есть они имеют свои алгоритмы,

собственную базу, которая может иметь документы, не представленные в открытом доступе. К тому же специализированные системы имеют высокую скорость работы в отличие от метапоисковых систем [8].

Существует множество различных российских и зарубежных систем и программ для выявления плагиата и большинство из них используют метапоисковые системы

или оболочки к ним, направленные на поисковую оптимизацию.

Следует отметить, что использовать зарубежные системы для проверки российских научных работ на заимствование не имеет особого смысла, так как они не имеют базы с русскоязычными документами [4].

Популярные российские и зарубежные онлайн-системы и программы для проверки текстов на плагиат представлены в таблице 3.

Таким образом, можно сделать вывод, что на сегодняшний день плагиат является большой проблемой во всем мире. Существует множество систем и программ для выявления заимствований, но ни одно из них не может дать 100% гарантию, что в работе найдены все не уникальные фрагменты. И поэтому для получения более точного результата проверки лучше проверить текст несколькими системами или же обратиться за экспертизой, хотя этот способ является платным и долговременным.

Список литературы

1. Аушра А. Научная электронная библиотека как средство борьбы с плагиатом // Образовательные технологии и общество. – 2006. – Т.9, № 3. – С.270-275.
2. Бабкин С.А. Интеллектуальная собственность в Сети «Интернет» – М.: Центр ЮрИнфоР, 2005. – № 52 – 215 с.
3. Вязова Н.В., Назарова О.Б. Использование информационных технологий для повышения эффективности процесса формирования контингента студентов первого курса вуза / Н.В. Вязова, О.Б. Назарова // Новые информационные технологии в образовании. Материалы VII международной научно-практической конференции. Российский государственный профессионально-педагогический университет. Екатеринбург, 2014. С. 538-542.
4. Давлеткиреева Л.З. Информационно-предметная среда как средство профессиональной подготовки будущих специалистов в университете. Дис.. канд. пед. наук Текст. Магнитогорск, 2006. -184 с.
5. Давлеткиреева Л.З. Информационно-предметная среда как средство профессиональной подготовки будущих специалистов в университете: автореф. дис.. канд. пед. наук. Магнитогорск, 2006. -23 с.
6. Махмутов М.М. Теоретические аспекты применения информационно-предметной среды для профессиональной подготовки будущих специалистов по информационным технологиям / Л.З. Давлеткиреева, М.М. Махмутов // Сибирский педагогический журнал. - 2009, № 5. - С. 78-91.
7. Махмутова М.В. Инновационная модель подготовки ИТ-специалиста в образовательной среде вуза / М.В. Махмутова, Л.З. Давлеткиреева // Современные информационные технологии и ИТ-образование. Сборник избранных трудов VII Международной научно-практической конференции. Под ред. проф. В.А. Сухомлина. - М.: ИНТУИТ.РУ, 2012. -1050 с. - С. 118 -129.
8. Чусавитина Г.Н. Индивидуальная траектория профессиональной подготовки ИТ-специалистов при переходе на стандарт третьего поколения [Текст]: Перспективы развития информационных технологий / Л.З. Давлеткиреева, Г.Н. Чусавитина. - 2011. -№ 5. - С. 22-27.
9. Чусавитина Г.Н. Подготовка будущих ИТ-специалистов в области обеспечения interoperability электронной науки и образования / Г.Н. Чусавитина, Л.З. Давлеткиреева, О.Е. Масленникова // Разработка инновационных механизмов повышения конкурентоспособности выпускников ИТ-специальностей вуза в условиях монопромышленного города: Сб. статей/под ред. Г.Н.Чусавитиной, Л.З. Давлеткиреевой. -Магнитогорск: Изд-во Магнитогорск. гос. Ун-та, 2012. - С. 132-140. -Библиогр.: с. 140. -200 экз. - ISBN 978-5-86781-982-8.
10. Чусавитина Г.Н., Давлеткиреева Л.З. Анализ и установление уровня зрелости информационной инфраструктуры организации для управления непрерывностью бизнеса // Современные информационные технологии и ИТ-образование / Сборник избранных трудов VII Международной научно-практической конференции. Под ред. проф. В.А. Сухомлина. -М.: ИНТУИТ.РУ, 2012. -1050 с.-С. 529-544.
11. Шарапов Р. В. Анализ подходов к обнаружению заимствованных текстов // Современные наукоемкие технологии. – 2011. – № 3. – С.47-49.
12. Шарапова Е.В. Исследование возможностей системы «Антиплагиат» для обнаружения заимствований // Перспективы науки и образования. – 2013. – № 3. – С.215-219.
13. About Copyscape [Электронный ресурс] – Режим доступа: <http://www.copyscape.com/about.php> (дата обращения: 09.01.16).
14. Advego Plagiatius – проверка уникальности текста [Электронный ресурс] – Режим доступа: <https://advego.ru/plagiatius/> (дата обращения: 28.12.16).
15. Broder A. On the resemblance and containment of documents // Compression and Complexity of Sequences (SEQUENCES'97). – IEEE Computer Society, 1998. – P. 21-29.