

УДК 004.043

ИСПОЛЬЗОВАНИЕ СИСТЕМ АНТИПЛАГИАТА В ОБРАЗОВАНИИ

© Е.С. Чиркин

Ключевые слова: плагиат; плагиат в образовании; антиплагиат; системы антиплагиата; неправомерное заимствование; заимствование.

Представлен обзор текущего состояния борьбы с неправомерными заимствованиями (в основном, в научной и образовательной сфере), ответственности за плагиат, обзору систем антиплагиата, методов борьбы с плагиатом, а также методов борьбы с антиплагиатом, критике применения систем антиплагиата в образовании. В конце работы приведен ряд рекомендаций, которые могут оказаться полезными при эксплуатации или создании систем антиплагиата, а также при выработке нормативной базы по использованию последних.

ВВЕДЕНИЕ

К настоящему времени в мире идет усиление борьбы с неправомерным использованием чужой интеллектуальной собственности. Последний виток этой борьбы вызван повсеместным проникновением широкополосного Интернета, удешевлением доступа к нему, ростом количества сервисов и услуг, которые можно получить удаленно, доступностью систем электронной коммерции; последние два пункта форсируют совершенствованием законодательства в данной сфере.

ОСНОВНАЯ ЧАСТЬ

Определение понятия «плагиат». У плагиата существует множество определений [1], одинаковых с точностью до формулировок. В данной работе под плагиатом, как и везде [1], подразумевается умышленное присвоение авторства чужой работы или ее части. Соответственно, плагиат находится в сфере действия авторского права. Следует отметить, что возможен и «неумышленный плагиат», который наиболее известен по музыкальным произведениям. Сложность проблемы последнего заключается в возможном действии объективных факторов – в свойствах и/или расстройках человеческой памяти (например, криптомнезия – «забывание источника» информации [2]), стилизации – подгоне произведения под принятые в данной отрасли нормы, банальных совпадениях, превращает проблему определения плагиата в отдельных случаях в нетривиальную задачу.

Ответственность за плагиат в РФ, действующее законодательство. Правовое поле ответственности за плагиат – авторские и смежные права. Ответственность за неправомерное использование чужой интеллектуальной собственности в Российской Федерации регулируется (по действующим на 23 сентября 2013 г. редакциям):

- ГК РФ: ст. 1250-1253, в т. ч. 1253.1;
- КоАП РФ, ст. 7.12 «Нарушение авторских и смежных прав, изобретательских и патентных прав»;

– УК РФ, ст. 146 «Нарушение авторских и смежных прав».

Вкратце, содержание этих статей можно свести к следующим пунктам:

– по ст. 1251 ГК РФ: в случае нарушения личных неимущественных прав автора их защита осуществляется, в частности, путем признания права, восстановления положения, существовавшего до нарушения права, пресечения действий, нарушающих право или создающих угрозу его нарушения, компенсации морального вреда, публикации решения суда о допущенном нарушении;

– по ст. 7.12 КоАП РФ: нарушение авторских и смежных прав в целях извлечения дохода влечет за собой наложение административного штрафа на граждан в размере от 1500 до 2000 руб. с конфискацией контрафактных экземпляров произведений и фонограмм;

– по ч. 1 ст. 146 УК РФ: если «присвоение авторства (плагиат) причинило крупный ущерб автору или иному правообладателю, то оно наказывается штрафом в размере до 200 тыс. рублей или в размере заработной платы или иного дохода осужденного за период до 18 месяцев, либо обязательными работами на срок от 180 до 240 часов, либо арестом на срок от 3 до 6 месяцев».

Таким образом:

1) административная ответственность наступает только в том случае, если доказана цель плагиата – извлечение дохода;

2) уголовная ответственность – только если доказано причинение крупного ущерба;

3) считается, что плагиат в научных работах не связан с целью извлечения доходов и не причиняет крупного ущерба.

Вывод: присвоение авторства на научную работу реально влечет только гражданско-правовую ответственность.

Правомерное использование чужой интеллектуальной собственности. Правомерное использование чужой интеллектуальной собственности в Российской Федерации, без согласия автора и без выплаты ему

вознаграждения регулируется гражданским кодексом РФ, ст. 1274 «Свободное использование произведения в информационных, научных, учебных или культурных целях» [3].

В применении к сфере обучения и образования, ее положения можно свести к следующим (разумеется, приведенная трактовка не претендует на юридическую полноту, корректность и непротиворечивость, все детали и неупомянутые здесь нюансы следует смотреть в первоисточнике или [3]):

1) цитируемое произведение должно быть первоначально опубликовано и/или введено в гражданский оборот;

2) допускается без согласия автора или иного правообладателя и без выплаты вознаграждения, но с обязательным указанием имени автора, произведения которого используется, и источника заимствования:

а) цитирование в объеме, оправданном целью цитирования;

б) как иллюстрация в произведениях учебного характера с учетом цели;

3) возможно передавать произведения во временное безвозмездное пользование в библиотеках; произведения в цифровой форме – только в помещениях библиотек при условии исключения возможности создать копии этих произведений.

Следует обратить внимание в тексте закона на формулировки «объеме, оправданном целью цитирования» и «с учетом цели» – понятие о «целях» и «объемах» заимствования могут не совпадать у заимствующего, правообладателя и суда.

Возможная в (ближайшем) будущем ответственность за плагиат. В настоящее время во всех странах мира существует тенденция по ужесточению наказаний за правонарушения в области авторского права. Ниже перечислены факты, говорящие о подобном тренде в России и о возможных наказаниях.

1. «На заседании комиссии по координации деятельности Открытого правительства министр внутренних дел Владимир Колокольцев выступил с предложением о введении уголовной ответственности за плагиат в научных работах» [4].

2. «Министерство культуры России подготовило новый законопроект о борьбе с пиратством в Интернете <...> он распространяется, в отличие от уже вступившего в силу закона, не только на фильмы, но и на остальные объекты авторских и смежных прав – музыку, тексты, фотографии и другое <...> предполагается распространить и на сайты со ссылками на нелегальный контент. Минкульт предлагает также добавить статью «Нарушение порядка ограничения доступа к информации в случае нарушения авторского права и смежных прав» в Кодекс об административных правонарушениях. Эта статья предусматривает штрафы в размере 300 тысяч рублей для граждан, 600 тысяч – для должностных лиц и миллиона – для юридических» [5].

3. Упомянутый в предыдущем пункте законопроект в формулировках [6] внесен на обсуждение в Государственную Думу. С высокой долей вероятности он будет принят в изложенной форме в кратчайшие сроки.

Плагиат: современное состояние, общественное мнение. В последний год в РФ произошло несколько скандалов, связанных с обнаружением плагиата.

1. «Российская государственная библиотека проверила диссертации по истории, которые были защищены с 2000 года <...> Проверка показала, что почти 10 % работ почти полностью плагиат. По словам директора РГБ Александра Вислого, около полутора тысяч диссертаций по истории, защищенных за последние 13 лет, более чем на 70 % заимствованы из других работ. Каждая работа проверялась на наличие некорректных заимствований из той же базы диссертаций, и если бы ревизоры взяли для сравнения другие тексты, например, монографии и статьи, то объем плагиата был бы значительно больше» [7].

2. «Ректор МПГУ допустил многолетнее функционирование в своем вузе совета Д212.154.01 – «фабрики фальшивых диссертаций», – отметили в пресс-службе Минобрнауки» [8].

3. Общественное мнение по поводу плагиата в науке и среди должностных лиц: «Москва, 15 августа – АиФ-Москва. Данные опроса, проведенного социологами «Левада-центра» в конце июля среди 1,6 тыс. человек в 45 регионах России: за обнаружение плагиата в научной работе чиновники и депутаты должны лишаться должностей (83 %), ученые – своих степеней (82 %)» [9].

4. Сетевое сообщество на добровольных началах «Диссернет» [10] осуществляет проверку на плагиат диссертаций должностных лиц государства, депутатов, чиновников, преподавателей и администрации различных вузов. Несмотря на то, что сообщество явно преследует политические цели и что на данный момент результаты приводятся [10], но системно не интерпретируются, нет оснований полагать, что ситуация с плагиатом в диссертациях проверенных ими лиц лучше, чем в научной среде. Любопытно, что отдельные фрагменты текстов проверенных работ с плагиатом (причем, как фрагмент плагиата, так и фрагмент «источника») содержат специфичные искажения, появление которых иначе, чем попытками обойти именно автоматизированную проверку на плагиат назвать нельзя.

Борьба с плагиатом в системе образования. В настоящее время борьба с плагиатом в науке и образовании не выходит из рамки отдельных случаев, известны лишь события (например, [7]), имеющие определенный общественный резонанс. Началом системной борьбы с плагиатом можно назвать 2013–2015 гг.: «Минобрнауки обяжет все высшие учебные заведения размещать в открытом доступе на своих сайтах квалификационные, курсовые, дипломные, докторские, кандидатские и диссертационные работы каждого студента и аспиранта» [11].

Ключевые этапы создания системного подхода по повышению качества научных и квалификационных работ можно извлечь из статьи «Система открытой публикации» [12] и документа «Стратегия развития системы подготовки рабочих кадров и формирования прикладных квалификаций в Российской Федерации на период до 2020 года» [13]. Согласно им, основные этапы:

1) создание базы экземпляров всех работ: квалификационных работ обучающихся, кадров высшей квалификации и современных учебников, ее интеграция с ЕИС ВАК (2013 г.);

2) обеспечение доступа частных производителей программного обеспечения к базе контрольных экземпляров работ и их текстам, предоставление образовательным и научным учреждениям доступа к системам

контроля заимствований на любых приемлемых, в т. ч. коммерческих условиях, при этом право выбора соответствующих информационных систем должно остаться исключительно за самим учреждением (2014 г.);

3) уточнение правового определения термина «недобросовестное заимствование» (плагиат) в контексте проверки квалификационных работ. Создание организационно-методической базы для построения системы антиплагиата в научных и образовательных учреждениях. Разработка методик установления фактов неправомерного заимствования, регламентов процедур рассмотрения спорных вопросов неправомерного заимствования, систем моральных и административных санкций (срок – декабрь 2013 г.);

4) изменение требований к промежуточной и итоговой аттестации в образовательных учреждениях. Изменение процедуры регламентированного контроля качества образования в части обязательных требований к использованию систем контроля заимствований (2015 г.), аналогично диссоветы – 2015 г.;

5) реализация возможности открытого опубликования работ (квалификационных) (6 мес. с поправок в ГК), распределенных библиотек текстов работ (2014 г.), поисковые системы по ним – 2014 г.;

6) создание механизмов заинтересованности обучающихся в самостоятельной работе и развитии творческой активности через интеграцию открытых баз работ с социальными сетями (2014 г.), взаимное цитирование студентов студентами с премированием авторов самых цитируемых (2015 г.).

Всего данный этап развития состоит из 3-х периодов по одному году каждый (2013–2015 гг.).

Методы плагиата. Основные методы плагиата: полное заимствование (copy&paste-плагиат), замаскированный плагиат, пересказ, перевод, плагиат идей. На рис. 1 представлена условная шкала сравнения методов плагиата и сложности их обнаружения. Следует отметить, что методы с левой половинки рис. 1 обычно характерны для работ в областях телекоммуникации и информатике, естественнонаучных, правая – в основном, гуманитарных.

Методы антиплагиата. Единственный в РФ способ установления плагиата – экспертиза. Как имеющая правовые последствия для научных работ это должна быть судебно-лингвистическая экспертиза. Данная



Рис. 1. Условная шкала сравнения методов плагиата и сложности их обнаружения [14]

процедура для заинтересованных лиц является достаточно дорогой (от 20000 руб. [15]) и длительной во времени (от 5 рабочих дней [15]). При экспертизе обычно сравнивают множества используемых слов, последовательности фраз, стилистику текста, особенности речи и другие методы на усмотрение эксперта.

Однако с развитием средств вычислительной техники, Интернета и сетевых поисковых систем появилась возможность установления факта плагиата более простым и техничным способом – посредством поиска соответствующего фрагмента в сети. Данная процедура является сравнительно быстрой и не затратной.

К настоящему времени появились специализированные поисковые системы, направленные на поиск плагиата, т. н. «системы антиплагиата». Слово «антиплагиат» является нарицательным, в русском языке появилось при разработке системы «Антиплагиат» от ЗАО «Форексис» [16].

Принципы работы систем антиплагиата. Все системы, которые можно использовать для целей поиска заимствований, можно разделить на три больших класса:

1) поисковые системы сети Интернет. Не предназначены для поиска заимствований, но: а) с их помощью можно искать заимствования вручную; б) один из представителей – поиск от Google распознает некоторые научные работы и цитирования в них, в т. ч. и не обозначенные как заимствования (точнее, весьма обширный проект GoogleScholar постепенно интегрируется в поиск Google);

2) метапоисковые системы и системы антиплагиата, не имеющие значимой собственной базы документов. Работают посредством формирования вызовов на основе проверяемого документа к популярным поисковым машинам сети Интернет, интерпретируют их результаты. Для ускорения работы оперируют стоп-словами и проводят непоследовательную проверку документа (метод выборки) и др.;

3) специализированные системы антиплагиата с собственными алгоритмами поиска совпадений и собственными базами документов.

Особенность метапоисковых систем антиплагиата – их «базы» всегда актуальны, они не имеют доступа к документам, не размещенным в свободном доступе, либо документы в сети могут быть размещены в сводном доступе, но находиться в формате, поиск в которых не поддерживается. Скорость работы на клиенте обычно предельно низкая, возможно блокирование по IP поисковой системой (обычно – всегда и быстро блокируется) как неправомерное использование поисковой системы автоматизированными средствами, т. е. проверка будет невозможной или ее результаты будут искажены.

Особенность самодостаточных систем антиплагиата – специализированные алгоритмы поиска, собственные базы с, возможно, уникальным содержимым, не представленным в свободном доступе, относительно высокая скорость работы по сравнению с метапоисковыми системами.

Принципы работы специализированных систем антиплагиата:

1) преобразование проверяемого документа в текст;

2) индексирование полученного текста. Стадия индексирования может включать в себя:

а) выделение элементарных фрагментов текста и уменьшение их количества: исключение коротких слов, стоп-слов, слов, которых нет в словаре, приведение слов к базовой форме (например, лемматизация или стеммизация) и др.;

б) собственно, построение индекса;

3) при поиске аналогично преобразовывает проверяемый документ и происходит поиск совпадающих фрагментов, сортировка результата.

Чтобы правильно интерпретировать результаты работы автоматизированных систем антиплагиата, необходимо знать принципы их работы.

Принципы работы систем антиплагиата: морфология. При точном поиске изменение даже одного бита сказывается на результатах поиска. В поисковых системах обычно слова используются не в оригинальной форме, а преобразованными в некую базовую форму (например, для существительных это форма слова в единственном числе и именительном падеже, для глаголов – инфинитив). Данный механизм называется «поиск с учетом морфологии». Возможны следующие варианты (пример: «Кроха упала со стола на табуретку»):

1) нет морфологии + удаление коротких слов:

«кроха», «упала», «~~ее~~», «стола», «~~на~~», «табуретку».

Проблема: омонимы, одинаковое написание разных словоформ разных слов (ср. «горе» (беда) и «на горе» (место));

2) стемматизация слов (самый известный алгоритм – стеммер Портера):

«крох», «упал», «~~е~~», «стол», «~~н~~», «табурет».

Проблема: стемминги словоформ отличаются. Например, «кошка» → «кош», «кот» → «кот», «котенок» → «котен», в то время как обычно половые и возрастные характеристики объекта должны быть незначимыми;

3) лемматизация слов (лемма – базовая грамматическая категория):

«крошка», «упасть», «~~ее~~», «стол», «~~на~~», «табуретка».

Проблема: синонимы и омонимы. Например, затруднено восстановление базовой формы у: «замок», «замок», «замок»; «эти типы стали есть в литейном цехе»; «косой косошкой косил косой».

Общая проблема для всех алгоритмов «морфологии»: аббревиатуры и сокращения, особенно – контекстные. Например, слово «ПО» может быть предлогом «по» либо аббревиатурой «ПО» с одним из 8141 значений (по [27]).

Принципы работы систем антиплагиата: нечеткий поиск. Задача нечеткого поиска заключается в поиске неточно совпадающих слов или фрагментов текста. Чаще всего для нечеткого поиска используется метод N -грамм (N – количество подряд идущих символов из текста в элементе) или его вариации. Например, для слова «защищающийся» будут построены следующие биграммы ($N = 2$): «за», «аш», «щи», «ищ», «ща», «аю», «ющ», «щи» (повторно), «ий», «йс», «ся». Сравнение можно вести, например, по количеству совпавших биграмм. Например, при односимвольной ошибке: «защЕщающийся» изменятся две биграммы: «щи» → «щЕ» и «ищ» → «Ещ», остальные 9 биграмм – совпадают, что дает $9/11 = 82\%$ совпадения.

Принципы работы систем антиплагиата: шинглы и пассажи. Пословный поиск совпадений текстов – чрезвычайно ресурсоемкая операция, ее можно упростить, осуществляя каким-либо образом поиск не слов, а отдельных фрагментов. Например, предложений. Что дает очень низкую эффективность поиска, стоит лишь разбить одно из них на несколько или соединить соседние, поэтому обычно используют последовательности слов, извлеченных из текста. При последовательном извлечении данные фрагменты называются «шинглами» (характерная величина – L – длина шингла в словах), при непоследовательном (по какому-либо другому закону) – «пассажами» (характеризуется законом деления текста на пассажи). Например, пусть существует фраза «специалистов, которые занимаются копирайтингом, называют копирайтерами». Шинглы ($L = 2$) для нее следующие: «специалистов которые», «которые занимаются», «занимаются копирайтингом», «копирайтингом, называют», «называют копирайтерами». Для нормализации шингла, например, его составляющие можно упорядочивать (например, отсортировать по алфавиту). Для этой же фразы пассажи (согласно принципу деления – по знакам препинания): «специалистов», «которые занимаются копирайтингом», «называют копирайтерами». Удобство метода выделения пассажиров по сравнению с шинглами – они более крупные элементы текста, поэтому их меньше, что значительно повышает скорость работы с ними. Недостаток – при переработке текста пассажи более подвержены изменениям, чем шинглы.

Проблемы пассажиров и шинглов: речевые обороты, имена собственные и пр.: «несмотря на то, что» ($L = 5$), «в свою очередь, нельзя не заметить того факта, что» ($L = 9$), «Северный Ледовитый океан» ($L = 3$), «Министерство образования и науки Российской Федерации» ($L = 6$).

Известные характеристики систем антиплагиата. В общем случае принципы работы каждой из систем антиплагиата неизвестны, однако для каждой из них можно сделать обоснованные предположения.

eTxiAntiPlagiatv3.0.8.0. Согласно окну настроек, это метапоисковая система, используется метод выборок (выборка 20 шт. на каждую 1000 слов, размер выборки – 10 слов), выборки проверяются по шинглам ($L = 3$).

AdvengoPlagiatv 1.3.0.61. Согласно окну настроек, это метапоисковая система, используется метод выборки (размер выборки – 5 слов), выборки проверяются по шинглам ($L = 4$).

Антиплагиат от ЗАО «Форексис». По опыту использования и декларируемой независимости от языка документа [38], вероятно, используется метод шинглов на основе N -грамм (возможно, триграмм). Согласно их форуму и отчетам, можно предположить, что дополнительно используется нормализация текста по словарю синонимов и нахождение плагиата локально имеет эффект триггера для защиты от ложных срабатываний.

Поверхностный анализ результатов экспертиз сообщества «Диссернет» показывает, что в них производится поиск по точным совпадениям по шинглам с $L = 3$.

Известные системы антиплагиата. Антиплагиат от ЗАО «Форексис» (<http://antiplagiat.ru/>). Ключевая особенность системы – единственная, в чьих базах

содержатся диссертации из Российской государственной библиотеки.

1. AdvengoPlagiatus (<http://advego.ru/plagiatus/>). Простая метапоисковая система проверки на плагиат. Создана для поддержки биржи текстов Advengo.

2. eTtxtАнтиплагиат (<http://etxt.ru/>). Более совершенная, по сравнению с AdvengoPlagiatus, метапоисковая система проверки на плагиат. Создана для поддержки биржи текстов eTtxt.

3. Интеллектуальный робот контроля оригинальности и профессионализма Современной гуманитарной академии (часто встречается сокращение «ИСКОП СГА») (свободного доступа нет), является частью информационной системы вуза [19]. Ключевая особенность – согласно внутренним источникам, между собой интегрированы система проверки на плагиат, обширная электронная библиотека и личный кабинет студента, что в ходе проверки позволяет эффективно отличать самостоятельные работы, основанные на рекомендованном и учебном материале от посторонних работ.

Следует отметить, что существуют и иные программы и сервисы, имеющие в своем названии или описании слово «антиплагиат» или «уникальность текста» (например, PraideUniqueContentAnalyserII, istio.com, miratools.ru, также, многие проекты закрылись – DoubleContentFinder, FindCopy и др.), но все они являются метапоисковыми системами или оболочками к ним, направленными на поисковую оптимизацию, и для целей поиска плагиата в научных работах (курсовых, дипломных, диссертациях) абсолютно бесполезны. Также существует решение PlagiatInform [20] – продаваемая, но не работающая (внедрения не обнаружены) система антиплагиата.

Применение зарубежных решений (GoogleScholar, Copruspace, VroniPlag [21]) не имеет особенного смысла, т. к. они либо не имеют баз с русскоязычными документами, либо не поддерживают работу с русскоязычными текстами (эти два пункта принципиально важны, т. к. в России получение образования осуществляется только на русском языке), либо являются метапоисковыми системами, не имеющими решающих преимуществ перед уже упомянутыми.

Характеристики системы «Антиплагиат» от ЗАО «Форексис». Самой известной и, вероятно, самой крупной самостоятельной системой проверки на плагиат в России является «Антиплагиат» от ЗАО «Форексис». По [16] на 1 мая 2012 г. их база включает в себя:

- интернет-коллекция: свыше 40 млн существенно уникальных документов;
- коллекция полных текстов авторефератов и диссертаций РГБ: свыше 700 тыс. документов, в т. ч. практически все диссертации, защищенные в России с 2000 г.;
- коллекция нормативных и юридических документов Лекспро: свыше 7 млн.

Критика систем проверки на плагиат. Все претензии к системам антиплагиата можно свести к следующим основным положениям (приведены по порядку убывания количества претензий по каждому пункту).

1. Явное несовершенство. Не включая в себя следующие претензии, обычно данный пункт подразумевает незначительное наполнение баз, целые темы или даже отрасли не включены в базы системы, дается положительное заключение на работы, целиком списанные по 1-2-5 весьма старым источникам. Также

данный пункт включает в себя ложные срабатывания по распространенным стилистическим и речевым оборотам. Особенно сильное возмущение вызывает срабатывание системы по стилистическим оборотам, характерным для научной речи.

2. Учитываются служебные части документов. В зависимости от документа, от системы и времени проверки, проверка одного и того же документа, включающего либо не включающего список использованных источников, дает принципиальную разницу (более $\pm 15\%$ найденных заимствований).

3. Обычно с течением времени «уникальность текста» падает. Чаще всего это связано с пополнением баз системы антиплагиата, однако в некоторых случаях – с изменениями алгоритма проверки.

4. Закрытость алгоритмов проверки. Закрытость алгоритмов работы не является принципиальным недостатком, однако ввиду их явного несовершенства они, скорее, являются проблемой для уникальных творческих и научных работ, чем для списанных, причем при изменении первых для успешной «проверки на плагиат» теряется их уникальность.

5. Закрытость баз. Закрытость баз (пусть даже и неполных) автоматически предполагает злоупотребления со стороны администрации системы антиплагиата или ответственных лиц по работе с ней.

6. Учет правомерных заимствований, учет самоцитирования. Ни одна из существующих в мире систем антиплагиата не учитывает правомерное заимствование, цитирование, самоцитирование (считает их, соответственно, плагиатом), а также следующие важные моменты:

- законченная научная работа предполагает промежуточные публикации, в которых содержатся ее основные положения, полученные в процессе работы над ней. На практике содержание промежуточных публикаций полностью и неоднократно покрывает основную работу. Строго говоря, это не совсем правильно, но это повсеместная практика;

- при работе ученого над одной темой в течение некоторого промежутка времени появляются, в терминах авторского права, производные произведения, основанные на одном или нескольких произведениях того же автора – возникает проблема самоцитирования и большого количества местами повторяющихся работ. Строго говоря, это, опять же, не совсем правильно, но это – результаты новых итераций работы, все результаты которой и не должны быть новыми;

- научная работа ведется по определенным принципам, полагающим, что в основе конкретной работы лежит идея, концепция, эксперимент, работа, кем-то когда-то где-то уже высказанная или выполненная, что в обязательном порядке предполагает ее цитату, переказ или упоминание.

1) Высокие требования к работам. Конкретные цифры – см. ниже. Самый часто называемый минимально допустимый процент «оригинальности» работы – 80 % (для системы «Антиплагиат»). По опыту проверок, это очень высокая величина, достигнутая искусственным путем. Возможно, гуманитарная творческая работа типа школьного сочинения по литературному произведению способна показать столь высокий процент, но большинство уникальных научных работ, например, по информатике, покажет меньшую величину.

2) Некоторые типы работ подразумевают «плагиат» – в технических вузах есть работы, которые можно

охарактеризовать как «повторить по образцу» – речь идет, например, о вычислительных задачах и расчетно-графических работах, суть которых можно свести к «скопировать фрагмент методички, подставить числа из своего варианта, посчитать». Другой известный крайний вариант – работы юридической направленности, которые должны включать фрагменты законов. Разумеется, при проверках системами антиплагиата будет обнаружено 95–100 % плагиата.

3) Изменчивость алгоритма. Помимо того, что алгоритм закрыт, он еще и обычно меняется с течением времени, причем изменения могут быть как в сторону ужесточения проверки, так и в сторону смягчения требований. Это порождает определенное недовольство теми, для кого изменение оказалось принципиальным.

В дополнение, следует отметить повсеместное использование термина «оригинальность» («процент оригинальности» или «процент уникальности работы»), который подменяет собой изначальное значение данного слова и на самом деле означает лишь «не найдено в базе». Правильно употреблять, по означенной причине, слово «заимствование».

Критика системы «Антиплагиат» от ЗАО «Форексис». Система «Антиплагиат» от ЗАО «Форексис» ввиду систематического (с момента появления в 2005 г. [16]) наличия недостатков и ее известности в России (включая сотрудничество с вузами) удостоена дополнительной критики от общества и должностных лиц государства.

Критика № 1. Об единственном показателе «оригинальности» текста – «проценте оригинальности»: «Но ни в коем случае не должно быть механической стрижки под одну гребенку, тем более с использованием несовершенных систем контроля контента в диссертациях», – премьер-министр РФ Д.А. Медведев о системе «Антиплагиат» [22].

Критика № 2. Претензии по принципам организации работы и доступа по Д. Леванову [23]: а) неизвестно, как именно работает «Антиплагиат»; б) «Антиплагиат», будучи закрытой системой, позволяет манипулировать базами данных и исключать из баз диссертации.

Критика № 3. Почему именно «Антиплагиат»? «<...> Генпрокуратура России направила ответ на депутатский запрос Бурматова. «Документ уверил меня в мысли, что те вопросы, которые мы задавали в отношении законности деятельности этого ЗАО, были не напрасны. В частности, в своем ответе первый замгенпрокурора сообщает, что в ходе проверки, которую проводила прокуратура города Москвы, сотрудники ведомства в принципе не смогли обнаружить ЗАО «Антиплагиат» по месту его регистрации», – прокомментировал ситуацию Бурматов. <...> Бурматов считает, что в ходе судебного процесса станет известно, кто причастен к «лоббированию интересов» ЗАО «Антиплагиат». «Ливанову на суде придется объяснить, кто в миноре слал по вузам письма с требованием подключаться к услугам ЗАО, все встанет на свои места» [24].

Критика № 4. «На данном этапе идет уже не пересказ учебников своими словами, а тупой перебор различных формулировок одной и той же мысли, – какая пройдет. Сейчас студенческие работы представляют собой немислимый бред, потому что нормальные формулировки на грамотном русском языке уже кто-то однажды использовал» [25].

Борьба с системами антиплагиата. Борьбу с системами антиплагиата можно разделить на два основных подхода – технический и нетехнический.

Технические методы включают в себя: замену букв одного алфавита на букву сходного начертания другого («а» русскую на «a» английскую), перекрашивание в цвет фона пробелов, отдельных букв, точек или других символов. Вставка невидимого текста. Добавление (орфографических) ошибок. Использование синонимайзеров (программ для автоматической или полуавтоматической замены слов на синонимы). Возможно использование уязвимостей в системах антиплагиата, приводящих к требованию «проценту уникальности» текста (например, программы AntiPlagiatKiller). Более совершенный продукт (на момент своего создания) – AntiPlagiatKiller v2 анализирует текст и выводит рекомендации по изменению текста – такие, как убрать старое слово, добавить новое слово, «следует что-нибудь поменять» и др.

Достоинства технических методов обхода проверок на антиплагиат – их массовость, доступность, высокая скорость работы. Обычно технические методы обхода легко нейтрализуются создателями системы антиплагиата даже без получения в пользование соответствующего продукта. Но, опять же, на это требуется время, за которое получают положительное заключение множества работ.

Нетехнические способы обхода заключаются в переписывании текста. В настоящее время простейшие преобразования текста – соединение/разъединение предложений, перестановка слов местами, замена слов на синонимы, расшифровка сокращений или, наоборот, их использование, переписывание отдельных фрагментов – не оказывают значительного влияния на проверку. Говоря о переписывании, нельзя не упомянуть о поисковой оптимизации.

Поисковая оптимизация (searchengineoptimization) – комплекс мер по изменению формы и содержания (текста, веб-сайта) с целью повышения его позиции в выдаче поисковой системой. Помимо всего прочего включает в себя уникализацию текста и наполнение его ключевыми словами. Рерайтинг (rewriting) – переписывание текста с изменением формы изложения и сохранением оригинального смысла. Копирайтинг (copywriting) – профессиональное написание текстов (часто – рекламных). В SEO часто противопоставляется рерайтингу с точки зрения получения на выходе текста более высокого качества, с заданным наполнением и уникальностью.

Поисковая оптимизация существует более 15 лет. В настоящее время это целая индустрия, включающая в т. ч. обучение непосвященных. То есть в настоящее время существует множество руководств по переписыванию текстов непрофессионалами, по которым любое заинтересованное, никогда ранее не занимавшееся этим лицо может переписать текст, сделав его «уникальным». Разумеется, в данной индустрии существуют и услуги по «уникализации» текстов курсовых, дипломных и др. работ. Цены – от низких (30–100 руб./страница [26] и 200 руб./работа [27]) до правдоподобных (5000–7000 руб./работа [27, 28]) при ручной обработке текста, скорость выполнения заказа – высокая.

Известные граничные значения проверок текстов через системы антиплагиата. В данном пункте собраны известные граничные значения из разных ис-

точников. Помимо этого известна фраза «80 % оригинальности» (в контексте про систему «Антиплагиат»):

– положение об обеспечении самостоятельности выполнения письменных работ в МГУ им. М.В. Ломоносова на основе системы «Антиплагиат» [29] не содержит ни одной конкретной величины и отсылок на их существование;

– с форума студентов Современной гуманитарной академии: «скажите, правда что нужно получить не менее 70 % уникальности в антиплагиате? иначе из Москвы придет оценка 2??» [30];

– декан факультета социологии и политологии Финансового университета при Правительстве РФ Александр Шатилов: «Как правило, мы считаем, что когда «Антиплагиат» находит 70 % оригинального текста – это норма. Но учитывая постоянные сбои в работе программы, а также специфику некоторых тем нормой может быть и 50 %» [31];

– «без специального оформления фрагмента, полной и точной ссылки на источник, обширного объема заимствований (свыше 1500 печатных знаков)» из отчета, опубликованного на официальном сайте Росфиннадзора (отчет на сайте недоступен) – по статье «В НИОКР обнаружен научный плагиат на 160 миллионов рублей» от 19-01-2011 [32].

Замечание к последнему пункту: на одной странице формата А4 помещается, в зависимости от оформления, 1500–2000 знаков.

РАЗНОЕ

1. Согласно информационному сообщению по вопросам процедур защиты и проверки текстов диссертаций [33], ВАК РФ не пользуется системами антиплагиата. Оттуда же: «делать выводы о качестве научного исследования только по результатам компьютерной проверки невозможно и неправомерно» и «создание системы проверки использования заимствованного материала без ссылки на автора и (или) источник заимствования относится к полномочиям образовательной (научной) организации, на базе которой действует диссертационный совет, и осуществляется в инициативном порядке».

2. Изучая системы антиплагиата, можно обобщенно сделать вывод, что не все их разработчики подозревают о существовании много байтовых кодировок, манипулируя отдельными служебными символами которых, в специфичных случаях, возможно создание эффективных технических средств обхода систем антиплагиата.

3. С появлением централизованных баз с контрольными, курсовыми, дипломными, диссертациями существует следующая опасность автоматизированной проверки работ на наличие плагиата. Дело в том, что работы, логично предположить, должны храниться в их оригинальном виде – в том виде, в котором они создавались. Вероятно, это будет один из шести самых распространенных форматов (т. н. «форматы») файлов .doc, .docx, .rtf, .pdf, .odt или специально предназначенный для архивного хранения документов формат PDF/A). Общеизвестно, что в файлах форматов PDF [34] (в т. ч. PDF/A [35, 36]), а также DjVu [37] текстовый слой отделен от его представления, аналогичного назначения манипуляцию можно произвести и с другими форматами (но в них это сложнее). Таким образом, не представляя особой сложности создание документов специального формата, имеющих разное видимое (напеча-

танное) и текстовое (индексируемое) представления. Более того, учитывая поведение программ для распознавания символов [38], вероятно предположить, что этот же специальный текстовый слой и будет использован для проверки документа системой антиплагиата вместо распознавания страницы (выход один – проверять только отсканированные печатные работы либо текст, но с доработкой проблемы из предыдущего пункта).

ЗАКЛЮЧЕНИЕ

Следует напомнить, что факт плагиата устанавливается через суд, согласно действующему законодательству в области авторского и смежного права. Краткие выводы по работе:

1. Учитывая складывающуюся в стране ситуацию, использовать какую-либо систему проверки на плагиат (даже со всеми ее недостатками) – необходимо.

2. Используемая обычно методика проверки через «Антиплагиат» (от ЗАО «Форексис») в виде единого «процента» неприменима, требуется выработка более совершенных критериев, включающих в себя отношение к служебному содержимому документов (титульный лист, содержание, введение – выводы – заключение, список источников, приложения), к объему документов, стилистике речи, отрасли науки и др.

3. Учитывая принципы работы, особенности совершенствования систем антиплагиат и наполнения их баз, результаты повторных проверок, не связанные с апелляцией оценки или повторной защитой работы, не должны иметь обратной силы.

4. Обязателен контроль качества текста (особое внимание – научной стилистике и семантической грамотности) со стороны преподавателей по причине сложившейся индустрии рерайтинга и копирайтинга текстов, включающей обучение не вовлеченных в нее лиц (см. критику № 4).

5. Алгоритм применяемой системы антиплагиата должен быть подстроен под применяемую отрасль, что, в свою очередь, при правильном подходе привлекает все сопутствующие проблемы экспертных систем и машинного обучения – выбор экспертов и подготовка ими обучающей и контрольных наборов документов, а также создание или выбор адекватного по быстродействию алгоритма.

6. Даже в простейшем случае система антиплагиата должна оперировать семантическими или смысловыми единицами как противодействие переработке текстов с использованием аббревиатур и сокращений (для естественно-технических наук) и переформулировок (для гуманитарных наук).

Базовые рекомендации по повышению качества работ:

1) необходимо привитие культуры правомерного заимствования: использование соответствующих норм русского языка (помещение текста в кавычки, указание ссылки на источник в квадратных скобках после цитаты (для будущих систем) и прикладных технологий (например, использование HTML-тегов <Q> и <BLOCKQUOTE> везде, где это следует), использование, по возможности, первоисточников;

2) в учебных работах следует придерживаться мнения, что синтетическая работа реферативного типа по множеству источников, состоящая из большого количества цитат и, возможно, имеющая не очень значительную самостоятельную ценность, лучше скомпилированной по одной-двум другим работам;

3) повышение качества учебных работ возможно за счет снижения требований по величине их объема; использование ссылок и цитирования в минимально достаточном объеме;

4) следует повышать оригинальность работ с точки зрения (научной) новизны и отсутствия аналогов.

ЛИТЕРАТУРА

1. Плагиат. URL: http://dic.academic.ru/dic.nsf/fin_enc/16386 (дата обращения: 24.09.2013).
2. Криптомнезия. URL: <http://psychiatry.academic.ru/1366/Криптомнезия> (дата обращения: 24.09.2013).
3. Гражданский кодекс Российской Федерации. Ч. 4. Ст. 1274 «Свободное использование произведения в информационных, научных, учебных или культурных целях». URL: http://base.garant.ru/10164072/71/#block_41274 (дата обращения: 24.09.2013).
4. *Сабитова А.* Плагиат как уголовщина // ExpertOnline. 2013. 28 марта. URL: <http://expert.ru/2013/03/28/plagiat-kak-ugolovschina/> (дата обращения: 24.09.2013).
5. Минкульт предложил расширить действие антипиратского закона. URL: <http://lenta.ru/news/2013/08/07/nopiracy/> (дата обращения: 24.09.2013).
6. Проект Федерального закона «О внесении изменений в Гражданский кодекс Российской Федерации, Гражданский процессуальный кодекс Российской Федерации, Арбитражный процессуальный кодекс Российской Федерации» и Федеральный закон «Об информации, информационных технологиях и о защите информации» // Российская газета. 2013. 17 сентября. URL: <http://www.rg.ru/2013/09/17/internet-site-dok.html> (дата обращения: 24.09.2013).
7. *Ильин В.* Каждая десятая диссертация по истории – плагиат // Кемеровский городской сайт KemInfo. 2013. 4 сентября. URL: <http://keminfo.ru/news/9910/> (дата обращения: 24.09.2013).
8. Виктор Матросов уволен с должности ректора МПГУ // Московский день. 2013. 8 июля. URL: <http://mosday.ru/news/item.php?196778> (дата обращения: 24.09.2013).
9. Опрос: россияне предлагают увольнять чиновников за плагиат в диссертациях // Аргументы и факты – Москва. 2013. 15 августа. URL: <http://www.aif.ru/society/news/429447> (дата обращения: 24.09.2013).
10. Официальный сайт сообщества «Диссернет». URL: <http://www.dissernet.org/> (дата обращения: 24.09.2013).
11. *Сивкова А.* Дипломные и курсовые работы студентов опубликуют в Интернете // Известия. 2013. 15 января. URL: <http://izvestia.ru/news/542919> (дата обращения: 24.09.2013).
12. Система открытой публикации // Частный корреспондент. 2012. 29 ноября. URL: http://www.chaskor.ru/article/sistema_otkrytoj_publicatsii_30260 (дата обращения: 24.09.2013).
13. Стратегия развития системы подготовки рабочих кадров и формирования прикладных квалификаций в Российской Федерации на период до 2020 года // Министерство образования и науки Российской Федерации. URL: <http://минобрнауки.рф/media/events/files/41d4701ab6fda8ac35be.pdf> (дата обращения: 24.09.2013).
14. *MrScip, Krichanovsky A.* Способы обнаружения плагиата в зависимости от формы плагиата. URL: http://commons.wikimedia.org/wiki/File:Plagiarism_detection_Methods_Detection_Performance_ru.png?uselang=ru (дата обращения: 24.09.2013).
15. Услуга «Лингвистическая экспертиза» // Компания «Экспертный центр». URL: http://ceur.ru/services/lingvisticheskaja_jekspertiza/ (дата обращения: 24.09.2013).
16. Инструмент и только // Частный корреспондент. 2013. 13 мая. URL: http://www.chaskor.ru/article/antiplagiat_31963 (дата обращения: 24.09.2013).
17. ПО // Словарь сокращений, акронимов, аббревиатур и сложносоставных слов русского языка Sokr.ru. URL: <http://sokr.ru/ПО/> (дата обращения: 24.09.2013).
18. СМИ о проекте // Антиплагиат.ру. URL: <http://www.antiplagiat.ru/index.aspx?doc=massmedia> (дата обращения: 24.09.2013).
19. Электронное обучение, дистанционные и образовательные технологии // НАЧОУ ВПО Современная гуманитарная академия. URL: http://www.muh.ru/teaching/teaching_dot.php (дата обращения: 24.09.2013).
20. Плагиат-Информ. URL: <http://www.plagiatinform.ru/> (дата обращения: 24.09.2013).
21. *Ростовцев А.* Методологическое // Личный блог Андрея Ростовцева. 2013. 18 февраля. URL: <http://afrikanbo.livejournal.com/295638.html> (дата обращения: 24.09.2013).
22. Медведев назвал систему «Антиплагиат» несовершенной // Деловая газета Взгляд. 2013. 17 апреля. URL: <http://www.vz.ru/news/2013/4/17/629160.print.html> (дата обращения: 24.09.2013).
23. Блогер обвинил создателей системы «Антиплагиат» в желании давить на оппонентов // Деловая газета Взгляд. 2013. 21 марта. URL: <http://www.vz.ru/society/2013/3/21/625384.print.html> (дата обращения: 24.09.2013).
24. Единая Россия: прокуратура нашла нарушения в ЗАО «Антиплагиат» // Газета.Ru. URL: http://www.gazeta.ru/politics/news/2013/06/20/n_2986757.shtml (дата обращения: 24.09.2013).
25. Система явно изживает себя // Форум Антиплагиат.ру. URL: <http://www.antiplagiat.ru/forum/default.aspx?g=posts&t=793> (дата обращения: 24.09.2013).
26. Беспроигрышный способ обойти антиплагиат // Фактрум. 2012. 6 сентября. URL: <http://www.factroom.ru/facts/23751> (дата обращения: 24.09.2013).
27. Антиплагиату.нет. URL: <http://antiplagiatu.net/> (дата обращения: 24.09.2013).
28. Плагиата.нет. URL: <http://plagiatanet.ru/ceni/> (дата обращения: 24.09.2013).
29. Положение об обеспечении самостоятельности выполнения письменных работ в МГУ им. М.В. Ломоносова на основе системы «Антиплагиат». URL: <http://www.msu.ru/projects/antiplagiat/antiplagiat.doc> (дата обращения: 24.09.2013).
30. Как обмануть антиплагиат // Форум студентов Современной гуманитарной академии. URL: <http://forum.antimuh.ru/index.php?topic/26831-как-обмануть-антиплагиат/> (дата обращения: 24.09.2013).
31. Преподаватели вузов: «Антиплагиат» могут использовать в качестве черного пиара // Полит.ру. 2013. 26 марта. URL: <http://polit.ru/news/2013/03/26/apl/> (дата обращения: 24.09.2013).
32. В НИОКР обнаружен плагиат на 160 миллионов рублей // Служба информации Copyright.Ru. URL: http://www.copyright.ru/news/main/2011/1/19/obnarugen_na/ (дата обращения: 24.09.2013).
33. Информационное сообщение по вопросам процедур защиты и проверки текстов диссертаций // Пресс-служба Министерства образования и науки Российской Федерации. 2013. 10 апреля. URL: <http://минобрнауки.рф/новости/3281/> (дата обращения: 24.09.2013).
34. PDF 32000-1:2008: Document management – Portable document format – Part 1: PDF 1.7. // Adobe Systems Incorporated, 2008. URL: http://www.images.adobe.com/www.adobe.com/content/dam/Adobe/en/devnet/pdf/pdfs/PDF32000_2008.pdf (дата обращения: 24.09.2013).
35. ISO 19005-1:2005: Document management – Electronic document file format for long-term preservation – Part 1: Use of PDF 1.4 (PDF/A-1). URL: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=38920 (дата обращения: 24.09.2013).
36. ISO 19005-2:2011: Document management – Electronic document file format for long-term preservation – Part 2: Use of ISO 32000-1 (PDF/A-2). URL: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=50655 (дата обращения: 24.09.2013).
37. LizardtechDjVu reference. URL: <http://djvu.org/docs/DjVu3Spec.djvu> (дата обращения: 24.09.2013).
38. Извлечение текста из pdf в FineReader 10 // Форум компании AbbyyFineReader, 12-21 октября 2010. URL: <http://finereader.abbyy.ru/forum/actualthread.aspx?tid=1076&pg=-1> (дата обращения: 24.09.2013).

БЛАГОДАРНОСТИ: Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 12-07-00512).

Поступила в редакцию 10 октября 2013 г.

Chirkin E.S. USE OF SYSTEMS OF ANTI-PLAGIARISM IN EDUCATION

This article is devoted to a review of the current state of the struggle with plagiarism in scientific and educational spheres, responsibility for plagiarism, review of systems anti-plagiarism, review of problems fight with plagiarism and review of fight with system of anti-plagiarism, criticized the use of anti-plagiarism systems in education. The article contains some recommendations for using systems of anti-plagiarism and the development of normative base for the use him.

Key words: plagiarism; plagiarism in education; anti-plagiarism; systems of anti-illegal borrowing; borrowing.