

КАЧЕСТВО ИННОВАЦИИ ОБРАЗОВАНИЕ

№ 12
2013



журнал в журнале

КАЧЕСТВО и ИПИ (CALS)-технологии

www.quality-journal.ru

Д.А. Королев, А.В. Паволоцкий

АНАЛИЗ АУТЕНТИЧНОСТИ ПИСЬМЕННЫХ РАБОТ СТУДЕНТОВ

В статье поднимается вопрос об аутентичности студенческих работ и о понимании всеми участниками образовательного процесса ситуации с цитированием и заимствованием материалов в работах. Рассмотрена одна из типичных студенческих работ, успешно прошедшая проверку в системе "Антиплагиат".

Ключевые слова: плагиат, вёрстка, академическая культура, электронные документы, оценка, преподавание, высшее образование

Если в начале-середине 2000-х необходимость проверки дипломных и курсовых работ на аутентичность приходилось доказывать, то позже требования к работам самих вузов стали включать обязательную проверку работ в системе "Антиплагиат" [1]. Например, в методических рекомендациях НИУ ВШЭ такой пункт присутствует [2], и все выпускные работы действительно в обязательном порядке проверяются, и студентов информируют доступными средствами [3].

В данной статье мы рассмотрим проверку домашних работ, которые студенты сдают во время учебного модуля, то есть рядовой, но тот самый случай, который формирует у студента понимание академической культуры и показывает границы дозволенного.

Студентам второго курса МИЭМ НИУ ВШЭ, обучающимся по дисциплине "Инженерная и компьютерная графика", была предложена письменная работа в форме эссе. Задание добровольное, дающее дополнительные баллы в накопленной оценке. Помимо требований к содержанию работ, в задании указывались четкие критерии по оформлению и аутентичности [4]. В зависимости от учебной дисциплины, содержание может меняться, но требования к оформлению и аутентичности должны быть относительно одинаковыми - ГОСТ [5] един для всех, а плагиат - это общепризнанное нарушение. Тонкости оформления работ мы рассматривать не будем, но вопросы понимания аутентичности встают столь остро, что игнорировать их нельзя.

Каждый раз, формулируя задание на письменную работу, преподаватели заявляют, что "копипаст" недопустим, что источники следует указывать надлежащим образом [6], что существуют правила цитирования, и что за плагиат студента могут серьезно наказать. При этом, ни студенты, ни, зачастую, преподаватели не в состоянии сформулировать, где проходит граница допустимого. Получается, что пороговый уровень (например, 75%) по показателям "Антиплагиата" является для всех универсальным мериллом - и преподаватель может снять с себя труд по идентификации степени авторства студента, и студент получает официальную индульгенцию, если "прокатило".

В интернет можно легко найти советы, "как обойти Антиплагиат" (например, [7]), но практика показыва-

ет, что это не обязательно.

Далее мы разберём одну работу, которая попала в эту статью не по причине своего выдающегося в каком-либо смысле качества, она лишь содержит достаточно примеров, чтобы по ним написать небольшой иллюстрированный обзор, в целом же она не уникальна, и многие студенты узнают свой стиль написания работ. Приведенное перечисление признаков плагиата не претендует на всесторонний охват - и автор, и читатели вспомнят другие "методы и средства" написания текстов сомнительного происхождения, применяемые в написании курсовых, дипломных и прочих работ. Приведенный пример рассмотрим именно с позиций отношения к источникам. При этом есть все основания предполагать, что автор рассматриваемой работы совершенно искренне полагал, что "всё сделал правильно".

Источники, цитирование и заимствование

Читая работы студентов, можно сделать вывод о том, что они просто не подозревают о том, что возможны иные способы создания текстовых работ, кроме копирования текста и (в знак особого уважения к правилам) - расстановки ссылок на источники. Но источники не вдохновения или справочной информации, подтверждающие авторитетность и достоверность приведенных данных, а источники самого текста. При этом система Антиплагиат выдаёт высокие показатели аутентичности для большинства работ - действительно, они не скачаны из базы рефератов, они взяты из выдачи поисковой системы по запросу, содержащему ключевые слова из названия темы.

Кратко разберем способы идентификации заимствований в работах.

1. Выборочная проверка на заимствования

Задолго до повсеместного установления моды на проверку работ на Антиплагиате (и, как следствие, всеобщего успокоения по поводу их аутентичности), автором применялся простой, но довольно действенный способ идентификации происхождения текста: если студенты ищут тексты в поисковых системах и вставляют фрагменты найденных страниц (статей) в свои работы, то выборочная проверка в обратном порядке тоже должна рабо-

тать. Достаточно взять фрагмент из середины абзаца длиной 5-7 слов (рис. 1) и вставить в поисковую строку (рис. 2).

Поскольку поисковые системы неплохо адаптируются к произвольной замене порядка слов и другим нехитрым приёмам, то этот метод остаётся эффективным и в случае "предварительной обработки" текста. Проверим: поставим кавычки и заменим

часть букв латинскими (слова с заменёнными буквами подчёркиваются в браузере, как неизвестные словарю). Результат поиска представлен на рис. 3.

Разумеется, более вдумчивое изменение текста будет лучше скрывать его происхождение и, по крайней мере, потребует уменьшения длины запроса, а, следовательно, повышения количества нерелевантных нашей задаче результатов.

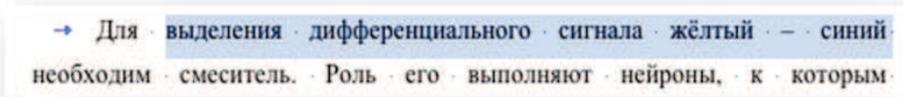


Рис. 1. Выделение фрагмента в тексте документа

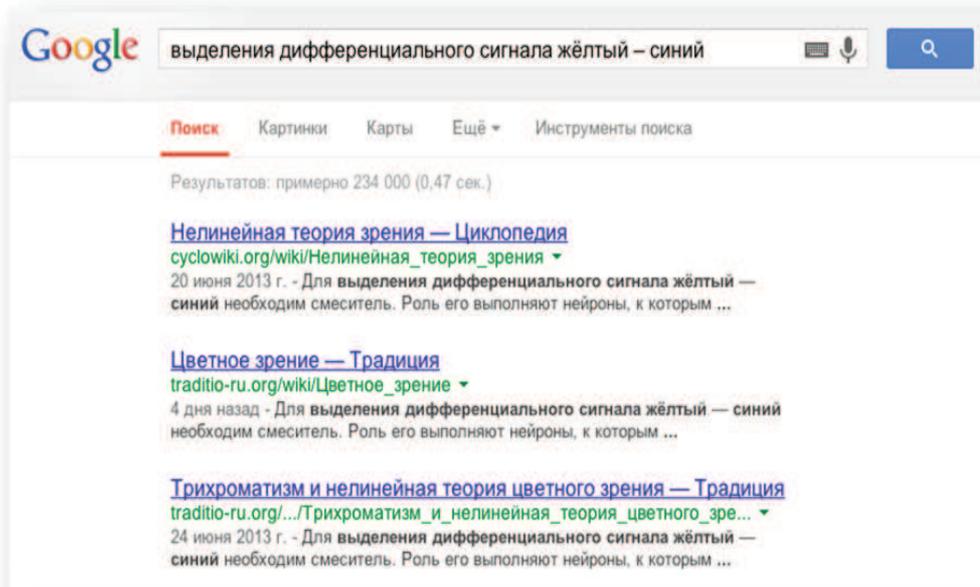


Рис. 2. Поиск выделенного фрагмента в Google

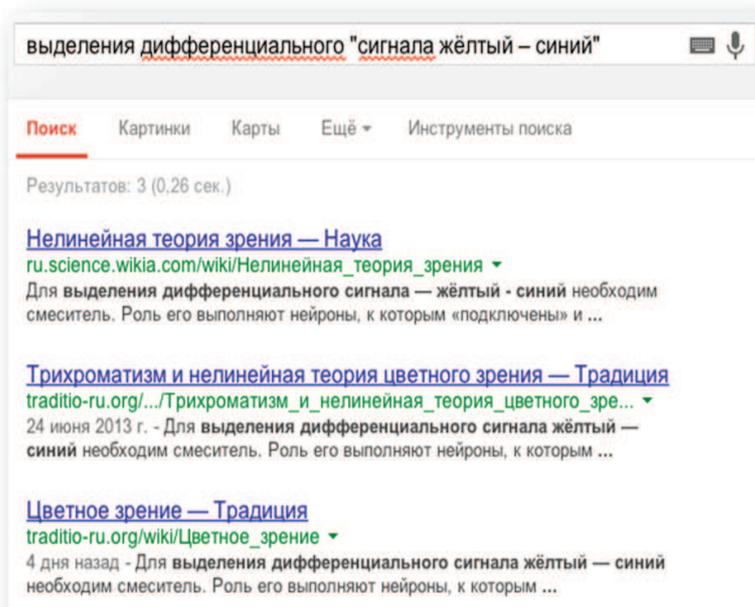


Рис. 3. Изменение запроса для маскировки заимствования (подмена латиницей, дополнение кавычками)

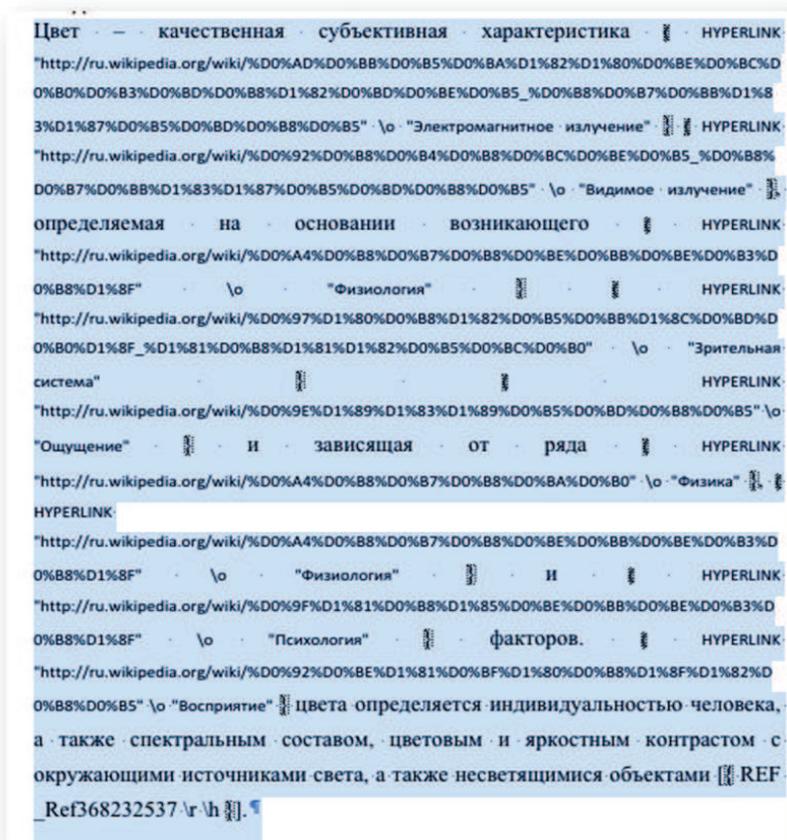


Рис. 6. Отображение кодов полей в том же тексте

5. Ременко, С. Нелинейная модель измерения цвета и уточнение терминов колориметрии. Всеакадемический семинар по проблемам стандартизации и метрологии. // С. Ременко. — Ташкент, 1986. — С. 41–42.

Рис. 7. Источник, который автор работы вероятнее всего никогда не читал

Цвет — качественная субъективная характеристика электромагнитного излучения оптического диапазона, определяемая на основании возникающего физиологического зрительного ощущения и зависящая от ряда физических, физиологических и психологических факторов. Восприятие цвета определяется индивидуальностью человека, а также спектральным составом, цветовым и яркостным контрастом с окружающими источниками света, а также несветящимися объектами. Очень важны такие явления, как метамерия, индивидуальные наследственные особенности человеческого глаза (степень экспрессии полиморфных зрительных пигментов) и психики.

Рис. 8. Оригинал текста с Рис. 5.

Поскольку все ссылки в рассматриваемом фрагменте ведут на Wikipedia, откроем наиболее вероятную статью (Цвет, Рис. 8).

Здесь подобной ссылки не обнаруживаем, поэтому ищем её в поисковой системе (Рис. 9).

На рисунке показана далеко не вся выдача, но понятно, что ссылка взята из реферативного материала одной из этих статей, тем более что эти сайты фигурируют в остальных фрагментах текста представленной работы.

4. Внезапные отклонения в узкоспециальные области

Теоретически, студент, проводящий самостоятельное исследование, может увлечься темой, стоящей в стороне от основной, и раскрыть её - вполне логичное развитие исследования, если нет жестких ограничений на границы предметной области. Но в общем случае такие отходы могут быть следствием копирования без осмысления.

Взятый из середины абзаца текст заметно отлича-

<input checked="" type="checkbox"/>	Сохраненная копия	Ссылка на источник	Хранилище	Доля в отчёте	Доля в тексте
<input checked="" type="checkbox"/>	[1] Якорение -реферат Пс...	http://www.char.ru/398/107940.htm	Интернет (Антиплагиат)	4,02%	4,02%
<input checked="" type="checkbox"/>	[2] Цветовое зрение. Осн...	http://otherreferats.allbest.ru/physics/00161335_0.html	Интернет (Антиплагиат)	3,39%	3,39%
<input checked="" type="checkbox"/>	[3] Влияние цвета на пси...	http://knowledge.allbest.ru/psychology/2c0b65635a3ac78b5c43b...	Интернет (Антиплагиат)	1,02%	1,78%

Рис. 11. Отчёт системы Антиплагиат

<input checked="" type="checkbox"/>	Сохраненная копия	Ссылка на источник	Хранилище	Доля в отчёте	Доля в тексте
<input checked="" type="checkbox"/>	[1] Якорение -реферат Пс...	http://www.char.ru/398/107940.htm	Интернет (Антиплагиат)	4,02%	4,02%
<input checked="" type="checkbox"/>	[2] rsl01003316309.txt	http://dlib.rsl.ru/rsl01003000000/rsl01003316000/rsl01003316...	РГБ, диссертации	0%	3,76%
<input checked="" type="checkbox"/>	[3] Факультет психологии...		ВШЭ	1,8%	3,62%
<input checked="" type="checkbox"/>	[4] Цветовое зрение. Осн...	http://otherreferats.allbest.ru/physics/00161335_0.html	Интернет (Антиплагиат)	3,39%	3,39%
<input checked="" type="checkbox"/>	[5] rsl01003320425.txt	http://dlib.rsl.ru/rsl01003000000/rsl01003320000/rsl01003320...	РГБ, диссертации	0%	3,02%
<input checked="" type="checkbox"/>	[6] rsl01005492521.txt	http://dlib.rsl.ru/rsl01005000000/rsl01005492000/rsl01005492...	РГБ, диссертации	1,02%	1,78%
<input checked="" type="checkbox"/>	[7] Влияние цвета на пси...	http://knowledge.allbest.ru/psychology/2c0b65635a3ac78b5c43b...	Интернет (Антиплагиат)	0%	1,78%
<input checked="" type="checkbox"/>	[8] Отделение программно...		ВШЭ	0%	0,88%

Частично оригинальные блоки: 0%
 Оригинальные блоки: 89,79%
 Заимствование из "белых" источников: 0%
 Итоговая оценка оригинальности: **89,79%**

Рис. 12. Отчёт системы "Антиплагиат" при использовании вузовской подписки

существующих), то эти искажения не имеют существенного влияния на идентификацию заимствования. Таким образом, сам метод показывает высокую эффективность. Вопрос лишь в том, к какой базе исходных данных он применяется.

Как отмечалось выше, студенты ищут ответы на свои вопросы отнюдь не в базе системы "Антиплагиат", а в крупных поисковых системах - Google и Яндекс, реже - на тематических сайтах с благородными миссиями, например Twirpx [9]. Отсюда вытекает логичное решение - искать тем же методом, но через поисковые системы. При отправке большого количества запросов с одного IP-адреса поисковая система выводит запрос CAPTCHA, придется ввести текст с картинки.

Описанный выше способ поиска источников проверяемой работы используется сервисом eTXT Антиплагиат [10]. Для пользователей доступна и веб-версия, и загружаемое приложение. Именно результаты проверки этим приложением позволили проверить сданные работы и получить численные результаты, близкие к субъективно определяемому уровню. Сравнение результатов проверки работ приведено в таблице 1.

Номер работы (A) - это её идентификатор при проверке, далее по столбцам (все показания даны в процентах):

- В - бесплатная версия Антиплагиат [1];
- С - вузовская версия того же сервиса

(hse.antiplagiat.ru);

- D - проверка программой eTXT Антиплагиат [10];
- E - разница между бесплатной и вузовской версиями Антиплагиат.ru;
- F - разница между наименьшим показателем систем Антиплагиат.ru (В и С) и eTXT.

Разница между Антиплагиат.ru и eTXT вычислялась по формуле $=\text{abs}(\min(Bx:Cx)-Dx)$, а значения этой разницы, превышающие 10%, выделены серым фоном в таблице. То есть, брались наиболее пессимистичные значения системы Антиплагиат.

Легко заметить, что особой разницы между публичной бесплатной и вузовской версиями нет, по крайней мере, для проверяемых 23 текстов.

При разнице с eTXT в пределах от 2% до 64% среднее значение составляет 20%. Подобные наблюдения проводились не только в приведенном примере, аналогичная разница была замечена при проверке работ по курсу "Информатика" в 2012/13 учебном году (тот же состав студентов).

Контекст применения

Рассмотренная студенческая работа (в таблице она имеет номер 15) была написана в качестве эссе на потоковом курсе "Компьютерная графика" [11], и разбор эссе, приведенный в данной статье, опубликованный непосредственно после проверки для студентов, можно найти в одном из сообщений [12]

Таблица 1. Сравнение оценки Antiplagiat и eTXT (указан % авторского текста)

Номер работы	Antiplagiat	Antiplagiat ВУЗ	eTXT	Разница Antiplagiat	Разница Antiplagiat - eTXT
A	B	C	D	E	F
1	83	83	70	0	13
2	93	93	71	0	22
3	94	94	85	0	9
4	80	79	76	1	3
5	82	81	70	1	11
6	74	74	69	0	5
7	82	81	47	1	34
8	100	100	96	0	4
9	75	75	49	0	26
10	82	82	49	0	33
11	76	75	25	1	50
12	83	83	78	0	5
13	84	84	59	0	25
14	100	99	97	1	2
15	92	90	26	2	64
16	99	99	95	0	4
17	96	96	61	0	35
18	96	94	78	2	16
19	87	87	39	0	48
20	97	97	88	0	9
21	76	76	64	0	12
22	91	91	85	0	6
23	92	91	72	1	19

на указанном сайте.

В порядке более масштабного эксперимента, включающего обширное применение электронных ресурсов и средств малой автоматизации (используется платформа GoogleDocs и другие сервисы Google), проверка этих эссе производится методом кросс-рецензирования (peer review), подобно тому, как такие работы проверяются в массовых открытых онлайн-курсах (МООС). В данном случае "пирами" (рецензентами) выступают студенты потока, на котором читается курс, и это один из способов снижения влияния преподавателя на итоговую оценку с оговоркой на обязательную компенсацию недостатков метода пирингового рецензирования. Критики спросят: "а судьи кто?" и будут правы: уровень экспертизы при проверке работ студентами крайне низкий, но не по причине незнания предмета, а по той же причине низкого уровня ответственности, а это не профессиональный, а культурный и психологический фактор. Соответственно, противостоять ему можно, задавая такие критерии и такие условия проведения проверки работ, чтобы минимизировать негативное влияние и выявить максимум пользы. Как это реализовано в данном случае:

- В курсе введена накопительная система оценок (10 баллов шкалы оценок ВШЭ = 100 баллов в курсе), при этом есть обязательные и дополнитель-

ные (необязательные) задания. Обязательные задания позволяют набрать максимальный балл, общая же сумма заданий в полтора раза превышает балл, необходимый для получения максимальной оценки.

- Задания имеют установленный максимальный балл. Например, эссе - 20 баллов, рецензирование - 10. Но есть и средства противодействия халтуре: любое сданное эссе (это не касается тех, кто не сдавал работу на проверку) имеет оценку X-6, то есть, нулевая оценка принесет "-6" баллов. Получить нулевую оценку можно только при выявлении плагиата, поскольку в других случаях она является суммой оценок по 4 критериям, каждый из которых принесет хотя бы по 1 баллу. Это правило стимулирует не сдавать заведомо ворованные тексты и не писать заведомо некачественные работы (в накопительной системе оценок слабым местом является возможность набрать высокий балл, взяв количеством, а не качеством), появляется риск поглотиться. Основной посыл - "не нагружайте рецензентов недостойными внимания работами".

- Для рецензий действует следующее правило: каждому студенту после приёма написанных эссе назначается проверка трех работ, проверка необязательная, но оценка начинается с "-3" баллов. Можно не делать эту работу, но это понизит итоговую

вую оценку на 9 баллов - жизненно, но отучившимся 11 лет в школе и год в институте студентам это не очевидно: необязательно не значит бесплатно.

- Поскольку эссе пишут далеко не все студенты потока, а проверяют их ввиду вышеназванных стимулов существенно большее число студентов, на каждую работу приходится значительное число рецензентов. На практике написали эссе 20% студентов, а рецензии прислали - более 80%. Каждая работа получила в среднем 10 рецензий. Это обстоятельство очень важно: при таких соотношениях начинают быть применимы методы статистики.

- Оценка за эссе выставляется по сумме усредненных оценок рецензентов по каждому критерию.

- Адекватность рецензий (именно этот вопрос является спорным в применении метода кросс-рецензирования в студенческой среде и даже в открытых онлайн-курсах) должна быть оценена в первом приближении статистическими методами. Для этого вычисляется средний балл, выставленный каждой работе всеми её рецензентами, далее это значение сравнивается с каждой конкретной оценкой, что даёт отклонение от среднего. Рецензии с наибольшим отклонением рассматриваются вручную преподавателем.

- В рамках курса заявляется максимальная дистанцированность преподавателя от выставления оценок: пиринговая оценка письменных работ, автоматическая проверка тестов - всё это сводит известное субъективное влияние преподавателя на оценки студентов к минимуму и проводит в жизнь принцип разделения функций, уводя оценочную функцию подальше от преподавателя. Но оценка рецензий на текущий момент требует авторитетного экспертного мнения, и в роли такого эксперта выступает преподаватель. В данном случае преподаватель создает свою матрицу оценок для всех эссе, рецензии на которые показали большие значения отклонения. Далее оценки рецензентов и преподавателя сравниваются с допуском ± 1 балл. Для рецензий, в которых более половины критериев не совпали с экспертной оценкой, выставляется "красный флаг", то есть, рецензия аннулируется. Практика показала, что статистическое выявление неадекватных рецензий прошло успешно, и в обсуждении со студентами было достигнуто взаимопонимание. Крайне важно с воспитательной точки зрения сопровождать любые санкции детальными и терпеливыми объяснениями с примерами. Искреннее недоумение с аргументом "я же старался, потратил время и силы" должно получить в ответ профессиональный, спокойный и детально аргументированный ответ, в таком случае требуемый воспитательный и образовательный эффект будет достигнут.

Анализ этого эксперимента - тема другой статьи, но, касательно рассматриваемого вопроса, можно сказать, что и рецензенты из числа студентов не были едины во мнениях, оценивая работы с явными признаками плагиата, несмотря на четкие описания критериев оценки [4] (работы оцениваются по

четырем критериям, каждый балл для каждого критерия имеет словесное описание достаточных характеристик работы).

Заключение

Приведенные в Таблице 1 данные говорят сами за себя. Выводы о ценности результатов проверки системой "Антиплагиат" каждый читатель сделает сам. Возможно, в гуманитарных областях она показывает лучшие результаты, но в заключение хочется обратить внимание на вопросы, выходящие за рамки выбора инструментов контроля.

Если смотреть на вопрос списывания работ с первых попавшихся в поисковой выдаче источников шире, то налицо проблема отсутствия базового уровня академической культуры, и винить в этом только студентов - это лицемерие. Для привития этой культуры на протяжении всего времени обучения студент должен встречать более-менее унифицированные требования и пояснения о правилах написания работ и качественную экспертизу своих работ, в том числе и по рассматриваемому вопросу. Если требования можно поместить на сайте, как это сделано в ВШЭ [3], довести до студентов "правила игры" (это тоже делается в начале обучения), то экспертиза всё ещё остаётся слабым местом, и начинать воспитание студентов по рассматриваемому вопросу следует, вероятнее всего, с настойчивой и неотвратимой работы с преподавателями. И здесь встаёт ребром большой вопрос высшего образования - триединство власти (содержание, преподавание и аттестация по курсу всецело зависят от одного человека), которое вызвало бы недоумение в других областях жизни. Как минимум, оценка результатов должна быть отделена, в противном случае она теряет объективность.

Наблюдение процесса оценки работ студентами (а они оставляли свои оценки с комментариями, что позволяет лучше понять мотивы выставления того или иного балла) поднимает следующий вопрос - способность студентов к рефлексии. Осмыслить с заданных позиций предложенную работу оказались способны далеко не все. Эти навыки должны прививаться не отдельным инструктажем, а в ходе обучения в рамках существующих курсов.

Литература:

1. Система анализа текстов на наличие заимствований "Антиплагиат". www.antiplagiat.ru
2. Кузьминов Я.И. О внесении изменений в "Методические рекомендации по подготовке и защите выпускных квалификационных работ" Приказ от 11.06.2008 г. № 31-06/1077 <http://www.hse.ru/data/929/831/1236/antiplagiat.doc>
3. Плагиат в письменных работах. Справочник учебного процесса НИУ ВШЭ. Студентам МИЭМ. <http://urlid.ru/becr>
4. Таблица критериев оценки письменной учебной работы (материалы курса). <http://urlid.ru/becr>
5. ГОСТ 7.32-2001 Межгосударственный стандарт "Отчет о научно-исследовательской работе. Структура и правила оформления". Межгосударственный

совет по стандартизации, метрологии и сертификации. - Минск, 2001. <http://urlid.ru/beco>

6.ГОСТ 7.1-2003. Межгосударственный стандарт "Библиографическая запись. Библиографическое описание. Общие требования и правила составления". - М.: Стандартинформ, 2010. <http://urlid.ru/becp>

7.Что такое Антиплагиат, зачем он нужен и как его обмануть? Рефком-Урал. <http://urlid.ru/becp>

8. Алгоритм шинглов. Wikipedia. <http://urlid.ru/bect>

9.TWIRPX <http://www.twirpx.com/about/>

10.eTXT Antiplagiat. <http://www.etxt.ru/antiplagiat/>

11."Компьютерная графика". Сайт поддержки курса "Инженерная и компьютерная графика". 2013 год. <http://urlid.ru/becq>

12.Королев Д.А. "Эссе: первые впечатления". Сайт поддержки курса "Инженерная и компьютерная графика". 2013 год. <http://urlid.ru/bed1>

Королев Денис Александрович,

к.т.н., доцент,

Кафедра Информационно-коммуникационных технологий МИЭМ НИУ ВШЭ.

e-mail: d.korolev@gmail.com

Паволоцкий Александр Владимирович,

к.п.н., доцент,

Кафедра Информационно-коммуникационных технологий МИЭМ НИУ ВШЭ.

e-mail: apavolotsky@hse.ru

D.A. Korolev, A.V. Pavolotsky

ANALYSIS OF THE AUTHENTICITY OF WRITTEN WORK OF STUDENTS

This article raises the problem of authenticity of the works by students and the comprehension of acceptable quoting by both students and teachers. All examples are shown from one typical essay that has successfully passed "Antiplagiat" verification.

Keywords: Plagiat, proofing, academic culture, electronic documents, evaluation, teaching, HE

References:

1.Plagiat detection service www.antiplagiat.ru

2.Y. I. Kuzminov, Changes to "Methodic recommendations to preparation of graduate works". Higher School of Economics, Moscow, 11.06.2008 #31-06/1077 <http://www.hse.ru/data/929/831/1236/antiplagiat.doc>

3.Plagiat in written works. Directory of studying at NRU HSE. For MIEM students. <http://urlid.ru/becr>.

4.Table of criteria of evaluation of written work (course materials: Computer Graphics, Fall 2013). <http://urlid.ru/becm>

5.Standard 7.32-2001. Inter-State Standard "Report for Scientific Research. Structure and Proofing". Inter-State Committee for Standardization, Metrology and Certification. Minsk, 2001. <http://urlid.ru/beco>

6.Standard 7.1-2003. Inter-State Standard "Bibliographic record. Bibliographic description. Common requirements and rules of composition". M. Standardinform, 2010. <http://urlid.ru/becn>

7.What is "Antiplagiat" and how to cheat it. Refcom-Ural. <http://urlid.ru/becp>

8.Shingles algorithm. Wikipedia. <http://urlid.ru/bect>

9.TWIRPX <http://www.twirpx.com/about/>

10.eTXT Antiplagiat service. <http://www.etxt.ru/antiplagiat/>

11."Computer Graphics" Course site, Fall 2013. <http://urlid.ru/becq>

12.Korolev D. "Essay: first impressions". Computer Graphics course site, Fall 2013. <http://urlid.ru/bed1>

Korolev Denis Alexandrovich,

Ph.D., Associate Professor,

Department of Information and Communication Technologies MIEM HSE.

e-mail: d.korolev@gmail.com

Pavolotsky Aleksandr Vladimirovich,

Ph.D., Associate Professor,

Department of Information and Communication Technologies MIEM HSE.

e-mail: apavolotsky@hse.ru