

Система обнаружения заимствований «Антиплагиат.ВУЗ» – решение для повышения качества образования

Юрий Викторович Чехович, к.ф.-м.н.

antiplagiat.ru



24 мая 2018 г.

г. Санкт-Петербург

Антиплагиат.Эклектика

Юрий Викторович Чехович, к.ф.-м.н.

24 мая 2018 г.

г. Санкт-Петербург

О проблеме заимствований в научных и учебных работах

24 мая 2018 г.

г. Санкт-Петербург

Проблема неправомерных и некорректных заимствований

Современная научная работа практически не может обойтись без заимствований. Важно, чтобы сделанные заимствования были правомерными и корректными.

Правомерное заимствование: «...обоснованное целями цитирования использование в своем произведении науки части чужого текста с обязательным указанием (ссылкой) на истинного автора и источник заимствования, оформленные в соответствии с установленными правилами цитирования»*

Некорректное заимствование: «...обоснованное целями цитирования использование в своем произведении науки части чужого текста с обязательным указанием (ссылкой) на истинного автора и источник заимствования, когда такое указание (ссылка) оформлена с нарушением установленных правил цитирования»*

*О плагиате в произведениях науки (диссертациях на соискание ученой степени): С.М. Шахрай, Н.И. Аристер, А.А. Тедеев. – М.: МИИ, 2014. - 176 с.

Проблема неправомерных и некорректных заимствований

- В России в 2005 году не менее **50%** дипломных работ, защищенных на «отлично», были скопированы из интернета без изменений (<https://rg.ru/2009/01/20/referaty.html>)
- Более **1 500** диссертаций по историческим наукам, защищенных в России после 2000 года, содержат значительные заимствования из других диссертаций (по материалам исследования, проведенного в 2013 году компанией Антиплагиат по заказу РГБ: http://www.chaskor.ru/article/istoricheskie_zaimstvovaniya_33399)
- В среднем **35%** студентов назвали скачивание работ из Интернета распространенной практикой (по данным проекта НИУ ВШЭ «Мониторинг студенческих характеристик и траекторий», проведенного в 2014 г. в 11 российских вузах
<https://www.hse.ru/data/2014/09/18/1315011740/%D0%9E%D1%82%D1%87%D0%B5%D1%82%D0%BF%D0%BE%D0%BF%D1%80%D0%BE%D0%B5%D0%BA%D1%82%D1%83%D0%9C%D0%A1%D0%A5%D0%A2.pdf>)

Проблема «copy-paste»

Основное противоречие:

подготовить несамостоятельную работу гораздо проще, чем обнаружить ее несамостоятельность....

...если нет специальных инструментов!

Антиплагиат

- **Антиплагиат** – это специализированная поисковая система, предназначенная для обнаружения заимствований в текстовых документах
- **Запрос** – это документ любого размера и практически любого формата
- **Результат проверки** – интерактивный отчет, в котором «подсвечен» заимствованный текст, показаны источники, рассчитан «процент оригинальности»
- **Антиплагиат.ВУЗ** – версия системы Антиплагиат, предназначенная для использования в учебных заведениях
- **Антиплагиат.Эксперт, Антиплагиат (защищенное исполнение)** – специализированные версии системы Антиплагиат

Отлаженные технологии проверки текстов

Проверка документов:

- любого размера
- на всех языках
- во всех распространенных форматах (doc, docx, pdf, rtf, txt, html, ppt, pptx, odt, xml, rar, zip, 7z, gzip, расширяется по требованию)

Обнаружение перестановок:

- предложений
- абзацев
- страниц и т.п.

Корректная обработка:

- вставок
- удалений
- замены форм слов
- правильно оформленных цитат
- перефразировок

Выявление замены

сходных по написанию букв
(снятие омоглифии)

Оповещение о попытках «обхода» системы

Область поиска содержит **430+ млн** источников

Интернет - **330+ млн** веб-страниц в индексе:

✓ Википедия (русскоязычная,
англоязычная)

✓ Архив.org

✓ Сайты рефератов

✓ Сайты СМИ

✓ Сайт вузов

✓ Сотни тысяч других сайтов

10+ млн новых
источников ежемесячно

100+ млн новых
источников за 2017 год

О переводных заимствованиях

24 мая 2018 г.

г. Санкт-Петербург

Переводные заимствования. Статьи

- В 2017 году компания Антиплагиат выпустила решения для обнаружения переводных заимствований с английского на русский язык
- Мы провели самостоятельное исследование, проверив с помощью этого модуля 2,5 млн русскоязычных статей из изданий, входящих в РИНЦ (коллекция eLibrary.ru)
- При проверках не исследовался текст, найденный «обычным Антиплагиатом», а в результатах не учитывались случаи, когда у английской и у русской статьи совпадает автор

Переводные заимствования. Статьи

- Было обнаружено более 20 тысяч статей, содержащих переводные заимствований в значительных объемах
- Часть результатов все-таки относится к переводным работам тех же авторов или выполненным в соавторстве, но...
- значительная часть результатов – это некорректные переводные заимствования.



Переводные заимствования. Пример 1

Energy-Aware Load Balancing in Content Delivery Networks

Vimal Mathew[†], Ramesh K. Sitaraman^{†‡} and Prashant Shenoy[†]

[†]University of Massachusetts, Amherst [‡]Akamai Technologies Inc.

Abstract—Internet-scale distributed systems such as content delivery networks (CDNs) operate hundreds of thousands of servers deployed in thousands of data center locations around the globe. Since the energy costs of operating such a large IT infrastructure are a significant fraction of the total operating costs, we argue for redesigning CDNs to incorporate energy optimizations as a first-order principle. We propose techniques to turn off CDN servers during periods of low load while seeking to balance three key design goals: maximize energy reduction, minimize the impact on client-perceived service availability (SLAs), and limit the frequency of on-off server transitions to reduce wear-and-tear and its impact on hardware reliability. We propose an optimal offline algorithm and an online algorithm to extract energy savings both at the level of local load balancing within a data center and global load balancing across data centers. We evaluate our algorithms using real production workload traces from a large commercial CDN. Our results show that it is possible to reduce the energy consumption of a CDN by more than 55% while ensuring a high level of availability that meets customer SLA requirements and incurring an average of one on-off transition per server per day. Further, we show that keeping even 10% of the servers as hot spares helps absorb load spikes due to global flash crowds with little impact on availability SLAs. Finally, we show that redistributing load across proximal data centers can enhance service availability significantly, but has only a modest impact on energy savings.

(CDN, for short) that delivers web content, web and IP-based applications, downloads, and streaming media to end-users (i.e., *clients*) around the world [8]. A large CDN, such as that of a commercial provider like Akamai, consists of hundreds of thousands of servers located in over a thousand data centers around the world and account for a significant fraction of the world's enterprise-quality web and streaming media traffic today [14]. The servers of a CDN are deployed in *clusters* where each cluster consists of servers in a particular data center in a specific geographic location. The clusters are typically widely deployed on the "edges" of the Internet in most major geographies and ISPs around the world so as to be proximal to clients. Clusters can vary in size from tens of servers in a small Tier-3 ISP to thousands of servers in a large Tier-1 ISP in a major metro area. A CDN's servers cooperatively deliver content and applications to optimize the *availability* and *performance* experienced by the clients. Specifically, each client request is routed by the CDN's *load balancing system* to an "optimal" server that can serve the content with high availability and performance. Content and applications can typically be replicated on demand to any server of the CDN. The load balancing system ensures high

Статья переведена из
англоязычной работы целиком,
включая таблицы и данные

ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ ПО ИССЛЕДОВАНИЮ ЭФФЕКТИВНОСТИ МЕХАНИЗМОВ ГЛОБАЛЬНОЙ БАЛАНСИРОВКИ НАГРУЗКИ В СИСТЕМАХ ДОСТАВКИ КОНТЕНТА

Нгуен Тхоай Ань, аспирант

Кравец Олег Яковлевич, д.т.н., профессор

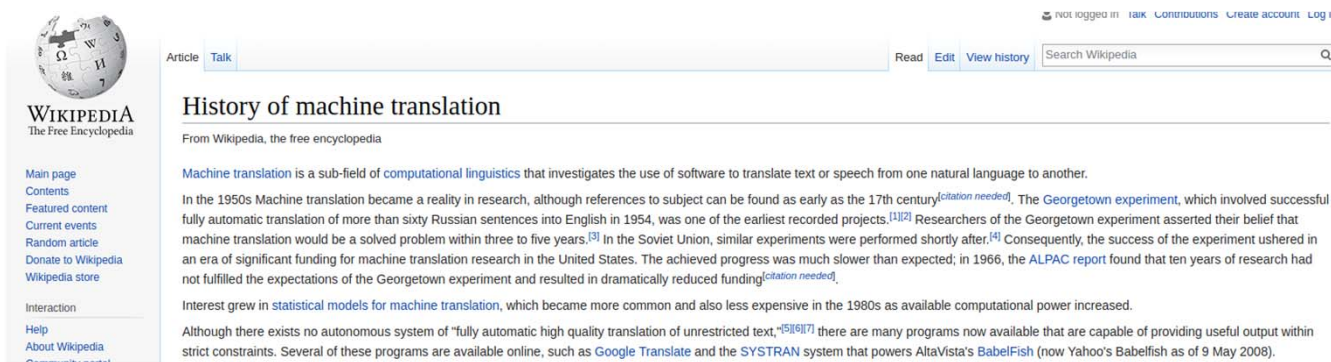
Воронежский государственный технический университет

1. Глобальная балансировка нагрузки

В предшествующих исследованиях авторов [7, 10-13] разработаны энергосберегающие механизмы для локальной балансировки нагрузки и перераспределения нагрузки между серверами в рамках одного кластера. В настоящей статье представлены результаты численного эксперимента по исследованию результатов балансировки нагрузки между различными кластерами сети доставки контента (СДК). Важное требование для глобальной балансировки нагрузки заключается в том, что каждый запрос подается из кластера, ближайшего к клиенту для обеспечения хорошей производитель-

5641v1 [cs.NI] 26 Sep 2011

Переводные заимствования. Пример 2



Статья на 78% состоит из перевода английской википедии (в русской версии такая статья отсутствует)

*Д.А. Гондоров, студ.; рук. И.В. Волкова, к. пс. н., доцент
(Филиал МЭИ в г. Смоленске)*

ИСТОРИЯ МАШИННОГО ПЕРЕВОДА

Машинный перевод - это подобласть компьютерной лингвистики, которая исследует использование программного обеспечения для перевода текста или речи с одного естественного языка на другой.

В середине 1930-х годов "машины-переводчики" были запатентованы Жоржем Арстрони, для автоматического двуязычного словаря, используя перфокарты. Петр Троянский представил более подробное предложение, включающее двуязычный словарь и способ борьбы с грамматическими ролями между языками, на основе грамматического строя эсперанто. Эта система была разделена на три этапа: на первом этапе носитель языка организовывал слова в их логических формах и проявлении синтаксических функций; второй этап требует машину "перевести" эти формы в целевом языке; и в третьем требуется носитель языка, чтобы нормализовать этот вывод. Предложение Троянского оставалось неизвестным до конца 1950-х годов, но к тому времени компьютеры были хорошо известны и реализованы.

Первая работа по компьютерному машинному переводу были представлены в 1949 году Уорреном Вивером, научным сотрудником фонда Рокфеллера, и были названы "Меморандумы перевода". Эта работа была основана на информационной теории, которая позволяла успешно взламывать

«Антиплагиат». Последние новости и ближайшее будущее

- Переводные заимствования
- Новый формат отчета
- Обнаружение перефразировок
- Использование OCR
- Общеупотребительные выражения (названия организаций, документов, объектов, устойчивые выражения)
- «Подсветка» списка литературы
- Собственная «белая коллекция» ВУЗа
- Интеграция с LMS, ЭБС, порталами
- Модернизация API

О проклятии процента

24 мая 2018 г.

г. Санкт-Петербург

Приказ МОН РФ № 636 от 29 июня 2015



МИНИСТЕРСТВО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
(МИНОБРНАУКИ РОССИИ)

П Р И К А

« 29 » ИЮНЯ 2015 г.

Москва 3/

Peru

01

Об утверждении Порядка проведения госуда
по образовательным программам высшего
бакалавриата, программам специалитета

В соответствии с частью 5 статьи 79 Конституции Российской Федерации от 29 декабря 2012 г. № 273-ФЗ «Об образе и стиле одежды в учреждениях государственной власти» (Собрание законодательства Российской Федерации, 2013, № 19, ст. 2326; № 23, ст. 2878; № 27, ст. 6165; 2014, № 6, ст. 562; № 6, ст. 566; № 2, ст. 2933; № 26, ст. 3388; № 30, ст. 4263; 2015, № 1, ст. 1) и подпунктом 5.2.36 Положения о Министре Российской Федерации, утвержденного постановлением Правительства Российской Федерации от 3 июня 2013 г. № 500 «О Положении о Министерстве Российской Федерации», 2013, № 23, ст. 2923; 2014, № 2, ст. 126; ст. 582; № 27, ст. 3776), и при

1. Утвердить прилагаемый Порядок проведения аттестации по образовательным программам высшего образования, программам бакалавриата, программам специалитета и программам магистратуры.

2. Признать утратившим силу при
Российской Федерации от 25 марта 2003
Положения об итоговой государственной а

О Положении проведения государственной итоговой аттестации - 05

38. Тексты выпускных квалификационных работ, за исключением текстов выпускных квалификационных работ, содержащих сведения, составляющие государственную тайну, размещаются организацией в электронно-библиотечной системе организации и проверяются на объём заимствования. Порядок размещения текстов выпускных квалификационных работ в электронно-библиотечной системе организации, проверки на объём заимствования, в том числе содержательного, выявления неправомерных заимствований устанавливается организацией.

Доступ лиц к текстам выпускных квалификационных работ должен быть обеспечен в соответствии с законодательством Российской Федерации, с учетом изъятия по решению правообладателя производственных, технических, экономических, организационных и других сведений, в том числе о результатах интеллектуальной деятельности в научно-технической сфере, о способах осуществления профессиональной деятельности, которые имеют действительную или потенциальную коммерческую ценность в силу неизвестности их третьим лицам.

Приказ № 636 от 29 июня 2015 и анализ практики его применения

Исследование локальных актов вузов

- 25 российских вузов из разных регионов, разного «масштаба» и направлений подготовки
- Исследовались Положения о ГИА или Положения (Регламенты) о проверке ВКР на объем заимствований

<http://www.oz2018.ru/materiali2017>



Сведения о результатах государственной итоговой аттестации по основной образовательной программе

Вид государственных аттестационных испытаний						
Защита выпускной квалификационной работы (ВКР)						
количество выпускников, всего	из них:			Результаты проверки ВКР на наличие заимствований		
	получивших оценку "удовлетворительно"	получивших оценки "отлично" и "хорошо"	выполнявших ВКР по заявкам предприятий	Средняя доля оригинальных блоков в работе	Доля работ с оценкой оригинальности текста менее 50%	Доля работ с оценкой оригинальности текста более 70%



Требования к оригинальности ВКР в отдельных российских вузах

	Бакалавриат	Магистратура	Специалитет
Региональный аграрный вуз	40%	60%	40%
Региональный технический вуз	50%	60%	50%
Региональный педагогический вуз	60%	60%	-
Региональный медицинский вуз	-	-	70%
Один из ведущих технических вузов	75%	75%	-
«Топовый» столичный вуз	80%	80%	80%

Закон Гудхарта

«Когда показатель становится целью, он перестает быть хорошим показателем»*.

**Goodhart, C.A.E. Problems of Monetary Management: The UK experience // Papers in Monetary Economics. — Reserve Bank of Australia, 1975. — Vol. I.*

Проблема неправомерных и некорректных заимствований

Современная научная работа практически не может обойтись без заимствований. Важно, чтобы сделанные заимствования были правомерными и корректными.

Правомерное заимствование: «...обоснованное целями цитирования использование в своем произведении науки части чужого текста с обязательным указанием (ссылкой) на истинного автора и источник заимствования, оформленные в соответствии с установленными правилами цитирования»*

Некорректное заимствование: «...обоснованное целями цитирования использование в своем произведении науки части чужого текста с обязательным указанием (ссылкой) на истинного автора и источник заимствования, когда такое указание (ссылка) оформлена с нарушением установленных правил цитирования»*

*О плагиате в произведениях науки (диссертациях на соискание ученой степени): С.М. Шахрай, Н.И. Аристер, А.А. Тедеев. – М.: МИИ, 2014. - 176 с.

Несколько практических рекомендаций по организации проверки

- Профилактика (курсы по академической этике и академическому письму)
- Проверка на заимствования с помощью систем обнаружения заимствований – не финал, а начало процесса по выявлению и квалификации заимствований
- Право на апелляцию результатов
- Развитие нормативной базы

Об открытом доступе к квалификационным работам

24 мая 2018 г.

г. Санкт-Петербург

ОД к квалификационным работам в России: начало

Проект Федерального Закона РФ «О внесении изменений в статью 7 Федерального закона «О рекламе» и статью 59 Федерального закона «Об образовании в Российской Федерации»

Полный текст выпускной квалификационной работы должен быть доступен для ознакомления на официальном сайте образовательной организации в сети Интернет для любых лиц в течение не менее 6 месяцев со дня защиты выпускной квалификационной работы.

Какие цели?

Цель квалификационной работы – подтвердить соответствующую квалификацию автора:

- ✓ Для любого желающего
- ✓ В любой момент времени
- ✓ Из любой точки мира

Единый подход к ОД квалификационных работ



http://www.ubackground.com/wallpaper/films/the_matrix_matrica_agent_smit_klony/22-0-6588



Сроки размещения



Будь хоть бедой в моей судьбе, Но кто б нас
ни судил, Я сам пожизненно к тебе Себя
приговорил.

(Константин Михайлович Симонов)

tsitaty.com

<https://tsitaty.com/%D1%86%D0%B8%D1%82%D0%B0%D1%82%D0%B0/170595>



АНТИПЛАГИАТ
ТВОРИТЕ СОБСТВЕННЫМ УМОМ

Вопросы авторского права



<http://zibainis.lt/sample-page/>



АНТИПЛАГИАТ
ТВОРИТЕ СОБСТВЕННЫМ УМОМ

Коммерческая тайна



<https://www.iphones.ru/iNotes/764823>

24.05.2018

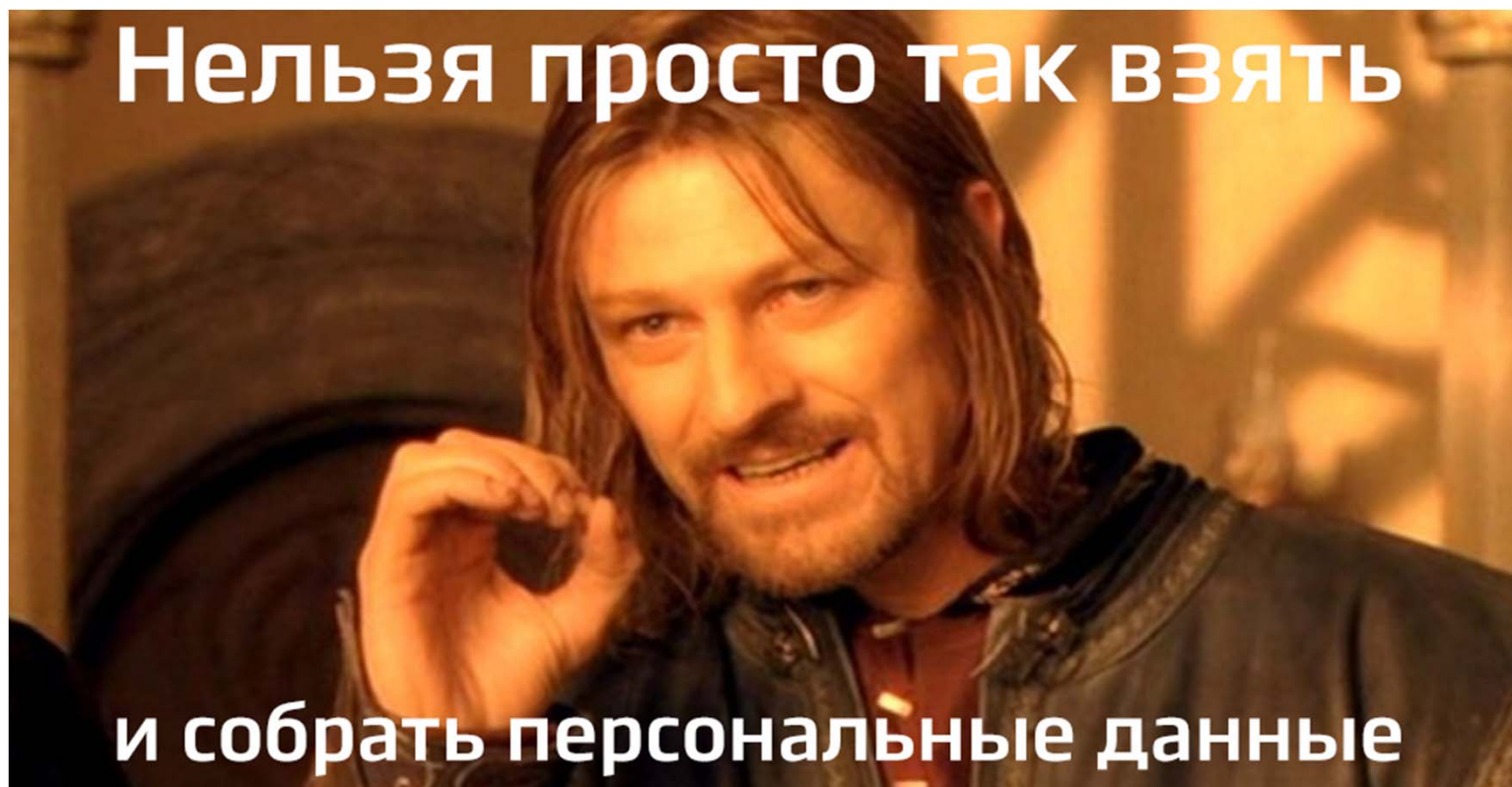
31



АНТИПЛАГИАТ
ТВОРИТЕ СОБСТВЕННЫМ УМОМ

Персональные данные

Нельзя просто так взять



и собрать персональные данные

<https://habrahabr.ru/company/zarlaw/blog/298882/>

Что получим?

- ✓ Повысим прозрачность присуждения квалификационных степеней;
- ✓ Повысим уровень ответственности вузов и научных организаций за результаты их работы;
- ✓ Повысим мотивацию студентов и соискателей при подготовке работ;
- ✓ Устраним проблему некорректного использования источников (научного плагиата и нарушения академической этики);
- ✓ **Улучшим качество квалификационных работ.**

Несколько слов о НОРА

24 мая 2018 г.

г. Санкт-Петербург

НОРА

- Составление «продвинутых» отчетов о заимствованиях, которые:
 - учитывают информацию об авторе публикации (самоцитирование, в том числе, выполненное в составе авторских коллективов учитывается отдельно)
 - сигнализируют о ситуациях, в которых авторы ссылаются, но не используют текст статьи и наоборот – используют текст, но не ссылаются [это пока немного космос – надо проверять будет ли это работать]
- «Хитрая разметка» списка источников выделение ссылок на статьи:
 - того же автора (в том числе, в составе авторского коллектива)
 - с такой же аффилиацией
 - из того же издания
 - может быть какие-то еще
- Ранжирование статей по методике, предложенной в проекте Научный архив
- Для новых статей:
 - Выявление ошибок в оформлении ссылок или случае неканонического оформления ссылок
 - Автоматическая рекомендация подходящих кодов различных классификаторов: ВАК, ГРНТИ, УДК и т.п.
 - Автоматическая рекомендация подходящих издания для подачи статьи
 - Автоматический поиск коллег, работающих в этой же научной области
 - Подбор кандидатов для литературного обзора

Все! Заканчиваю!

24 мая 2018 г.

г. Санкт-Петербург



Расписание вебинаров

регистрация: corp.antiplagiat.ru

Тема	Дата
NEW! Новые модули "Антиплагиат" - новые возможности для пользователя	22.05
Организация взаимодействия Преподаватель-Студент в системе Антиплагиат.ВУЗ	23.05
Проверка ВКР на заимствования с помощью системы Антиплагиат.ВУЗ	30.05, 14.06
Авторам учебной литературы: главное об авторском праве и системе Антиплагиат	24.05
Интеграция системы Антиплагиат.ВУЗ с информационными системами вуза через API	29.05
Система Антиплагиат как инструмент повышения качества научных работ	17.05
NEW! Заимствования в научных документах: просветительская лекция. Из опыта экспертов РГБ	06.06
Эффективное администрирование системы Антиплагиат	19.06
7 актуальных вопросов об организации проверок на заимствования в вуз	20.06



МИНИСТЕРСТВО
ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ



Российская ассоциация
ЭЛЕКТРОННЫХ
БИБЛИОТЕК

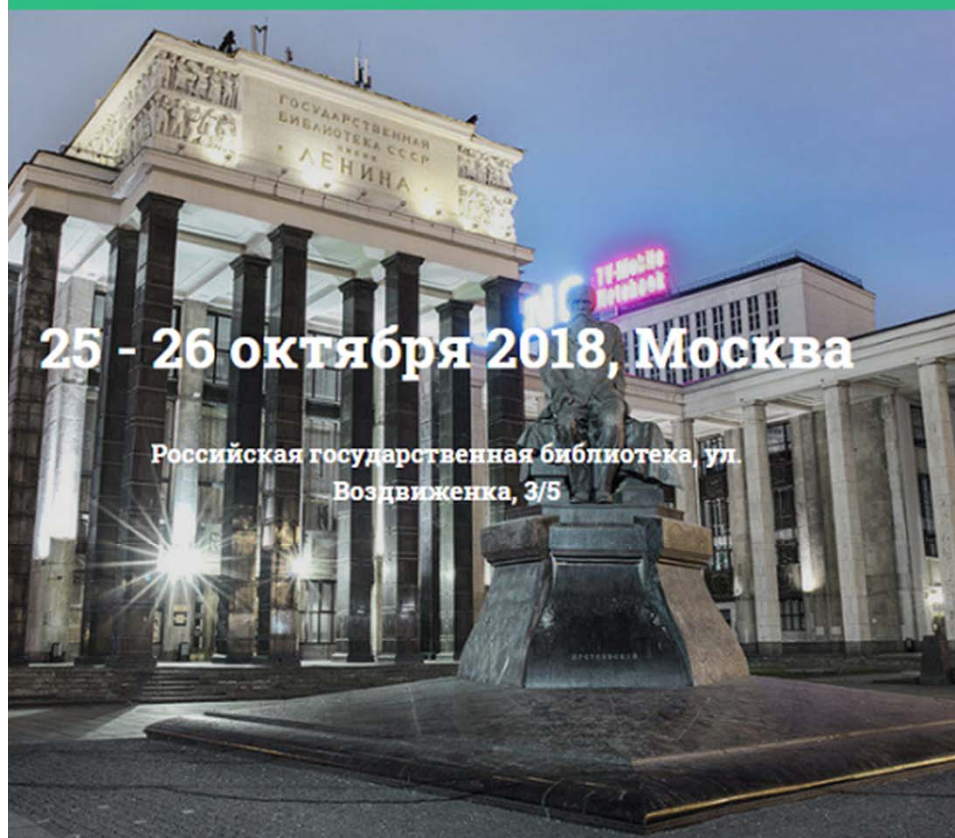


АНТИПЛАГИАТ
ТВОРИТЕ СОБСТВЕННЫМ УМОМ



ИРО
ГАОУПО ЛИПЕЦКОЙ ОБЛАСТИ
ИНСТИТУТ РАЗВИТИЯ ОБРАЗОВАНИЯ

Международная научно-практическая конференция «Обнаружение заимствований – 2018»



25 - 26 октября 2018, Москва

Российская государственная библиотека, ул.
Воздвиженка, 3/5



1 - 2 ноября 2018, Липецк

Институт развития образования, ул.
Циолковского, 18

orgkomitet@oz2018.ru

oz2018.ru

+7 (495) 223-23-84 доб. 107

ВАШИ ВОПРОСЫ

Юрий Чехович

antiplagiat.ru

8 (800) 777 81 28

+7 (495) 223 23 84

chehovich@antiplagiat.ru