АНТИПЛАГИАТ
ТВОРИТЕ СОБСТВЕННЫМ УМОМ

antiplagiat.ru

О некоторых результатах компании Антиплагиат в 2017 году

Юрий Викторович Чехович, к.ф.-м.н.

4 декабря 2017 г. г. Казань КПФУ

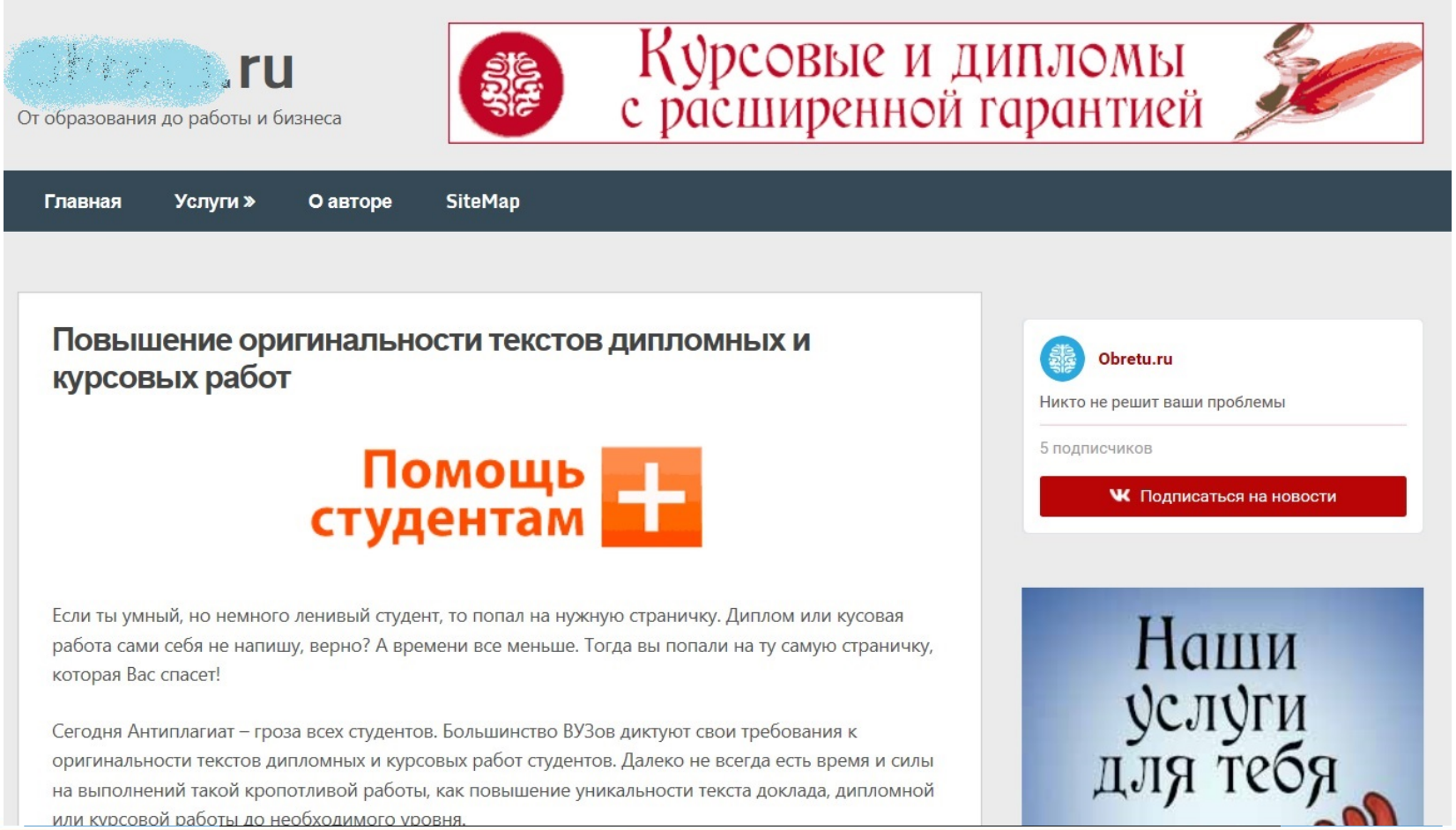
«Когда Румата миновал могилу святого Мики – седьмую по счету и последнюю на этой дороге, было уже совсем темно…»

Аркадий и Борис Стругацкие
«Трудно быть богом»

Антиплагиат

- **4 сентября 2017** нам исполнилось **12 лет**
- **Более 350 млн документов** в индексе: интернет, диссертации, статьи, патенты, юридические и нормативные документы
- **Более 5 млн новых документов** в индексе ежемесячно
- **Более 1000 организаций в Росси, Казахстане, Белоруссии, Кыргызстане:** вузы, НИИ, издательства, библиотеки и многие другие
- **Более 35 млн проверок ежегодно**
- **Более 85% студентов России** учатся в вузах, использующих Антиплагиат

Как обмануть Антиплагиат?



Как обмануть Антиплагиат? Типовые подходы

- Вставка «текста-призрака»
- Использование переводного плагиата
- Автоматизированная перефразировка

Проблема «текста-призрака»

Решение: использование OCR при извлечении текста из пользовательских документов

Проблема переводных заимствований

- Массовому пользователю стали доступны средства машинного перевода – работают google, yandex
- Другие способы текстовых заимствований становятся все более и более опасными – работает Антиплагиат

В. Писляков

Оно облегчает включение информации о контексте, например, (заученный) стиль поведения пользователя, время дня, местоположение, события из календаря и т. д. Более того, оно учитывает изменения в движении водителя/пассажира по пути к транспортному средству при помощи модели стохастического движения. Принятая концепция также может работать с нечёткими и нерегулярными замерами местоположения пользователя с помощью модели непрерывных во времени наблюдений со случайным шумовым компонентом. Таким образом, это универсальный и достаточно гибкий подход.

В. Богоров

Он обеспечивает включение контекстной информации, такой как (усвоенные) паттерны поведения, время дня, местонахождение, календарные события и т.д. Более того, он учитывает вариации в движениях водителей и пассажиров по дороге к автомобилю с учетом стохастической модели движения. Используемая формулировка также может учитывать нерегулярно размещенные и неточные оценки местоположения используя модель непрерывного временного наблюдения с компонентой случайного шума. Тем самым, представлена универсальная и весьма гибкая модель.

Яндекс

Это облегчает учет контекстуальной информации, такой как пользователя (узнал) о поведении, времени суток, местоположения, события календаря и т. д. Кроме того, он обслуживает изменчивости в движении водителя/пассажира на пути к транспортному средству через принятие стохастической модели движения. Принятая формулировка может также обрабатывать нерегулярно расположенные и неточные измерения местоположения пользователя с помощью модели непрерывного наблюдения времени со случайным шумовым компонентом. Таким образом, это общая и значительно гибкая структура.

Google

Это облегчает включение контекстной информации, такой как пользовательские (узнанные) шаблоны поведения, время суток, местоположение, события календаря и т. д. Кроме того, он обслуживает переменные в движении водителя/пассажира по пути к транспортному средству, предполагая стохастический модель движения. Принятая формулировка также может обрабатывать нерегулярно разнесенные и неточные измерения местоположения пользователя с помощью модели непрерывных наблюдений со случайным шумовым компонентом. Следовательно, это общая и значительно гибкая структура.

Как это работает?

- Вместо слов, которые являются основой при обнаружении «обычных» заимствований, используются понятия – автоматически формируемые кластеры слов
- Каждому понятию с помощью нейросетевых алгоритмов ставится в соответствие точка в пространстве очень большой размерности
- Заимствованным фрагментам соответствуют точки, расположенные неподалеку от искомой

Проблема переводных заимствований

- В начале 2017 года компания Антиплагиат начала подключения вузов к тестированию нового модуля переводных заимствований

Переводные заимствования. Статьи

- 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.

.5641v1 [cs.NI] 26 Sep 2011

***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-

Переводные заимствования. Пример 2

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal

History of machine translation

From Wikipedia, the free encyclopedia

Machine translation is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one natural language to another.

In the 1950s Machine translation became a reality in research, although references to subject can be found as early as the 17th century[citation needed]. The Georgetown experiment, which involved successful fully automatic translation of more than sixty Russian sentences into English in 1954, was one of the earliest recorded projects.[1][2] Researchers of the Georgetown experiment asserted their belief that machine translation would be a solved problem within three to five years.[3] In the Soviet Union, similar experiments were performed shortly after.[4] Consequently, the success of the experiment ushered in an era of significant funding for machine translation research in the United States. The achieved progress was much slower than expected; in 1966, the ALPAC report found that ten years of research had not fulfilled the expectations of the Georgetown experiment and resulted in dramatically reduced funding[citation needed].

Interest grew in statistical models for machine translation, which became more common and also less expensive in the 1980s as available computational power increased.

Although there exists no autonomous system of "fully automatic high quality translation of unrestricted text,"[5][6][7] there are many programs now available that are capable of providing useful output within strict constraints. Several of these programs are available online, such as Google Translate and the SYSTRAN system that powers AltaVista's BabelFish (now Yahoo's Babelfish as of 9 May 2008).

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 Уоренном Вивером, научным сотрудником фонда Рокфеллера, и были названа "Меморандумы перевода". Эта работа была основана на информационной теории, которая позволяла успешно взламывать

Переводные заимствования. Пример 3

Quantum Key Distribution with Qubit Pairs

Mohd Asad Siddiqui, Tabish Qureshi
Centre for Theoretical Physics, Jamia Millia Islamia, New Delhi, India.

Abstract

We propose a new Quantum Key Distribution method in which Alice sends pairs of qubits to Bob, each in one of four possible states. Bob uses one qubit to generate a secure key and the other to generate an auxiliary key. For each pair he randomly decides which qubit to use for which key. The auxiliary key has to be added to Bob's secure key in order to match Alice's secure key. This scheme provides an additional layer of security over the standard BB84 protocol.

| 17 Jul 2014

В русской статье содержится 39% текста, переведенного из английской статьи

КВАНТОВОЕ РАСПРЕДЕЛЕНИЕ КЛЮЧЕЙ С ИСПОЛЬЗОВАНИЕМ СПАРЕННЫХ КУБИТОВ

Кротова Елена Львовна
канд. физ.-мат. наук, доцент Пермского национального исследовательского политехнического университета, РФ, г. Пермь
E-mail: lenkakrotova@yandex.ru
Абдулов Дмитрий Игоревич
студент Пермского национального исследовательского политехнического университета, РФ, г. Пермь
E-mail: dmitry.igorevich.one@gmail.com
Андреев Роман Александрович
студент Пермского национального исследовательского политехнического университета, РФ, г. Пермь
E-mail: abusedroman@gmail.com
Бадртдинов Артём Сергеевич
студент Пермского национального исследовательского политехнического университета, РФ, г. Пермь
E-mail: asbadrtd@gmail.com
Салимзебаров Эльдар Дамирович
студент Пермского национального исследовательского политехнического университета, РФ, г. Пермь
E-mail: omeldarl@rambler.ru
Феофилова Полина Андреевна
студент Пермского национального исследовательского политехнического университета, РФ, г. Пермь
E-mail: feofilovap@gmail.com

Область поиска содержит 350 млн источников

Интернет – более 250 млн веб-страниц в индексе:

- ✓ Википедия (русскоязычная, англоязычная)
- ✓ Arxiv.org
- ✓ Сайты рефератов
- ✓ Сайты СМИ
- ✓ Сайт вузов
- ✓ сотни тысяч других сайтов

+ 5 млн новых источников ежемесячно

+ 100 млн новых источников в 2017 года

Область поиска

Русский	**Английский**	**Французский**	**Немецкий**	**Испанский**
> 168 млн	> 50 млн	> 1,8 млн	> 1,8 млн	> 432тыс.

Казахский	**Украинский**	**Белорусский**	**Киргизский**	**Таджикский**	**Азербайджанский**
> 1,6 млн	> 9,2 млн	> 142 тыс.	> 305 тыс.	> 22 тыс.	> 19 тыс.

Область поиска. Внешние коллекции

Коллекция диссертаций **Российской государственной библиотеки**

> 920 тыс.

Коллекция диссертаций **Национальной библиотеки Беларуси**

Новинка **> 12 тыс.**

Коллекция научных статей **eLIBRARY.RU** (РИНЦ)

> 9,9 млн

Коллекции нормативно-правовых документов

Новинка **ГАРАНТ > 54 млн**

LEXPRO > 12,3 млн

Сводная Коллекция ЭБС

«Университетская библиотека онлайн», «Лань», «БиблиоРоссика», «Юрайт», «Айбукс», «Book.ru»

> 430 тыс.

Коллекция **Патенты**

> 11,5 млн

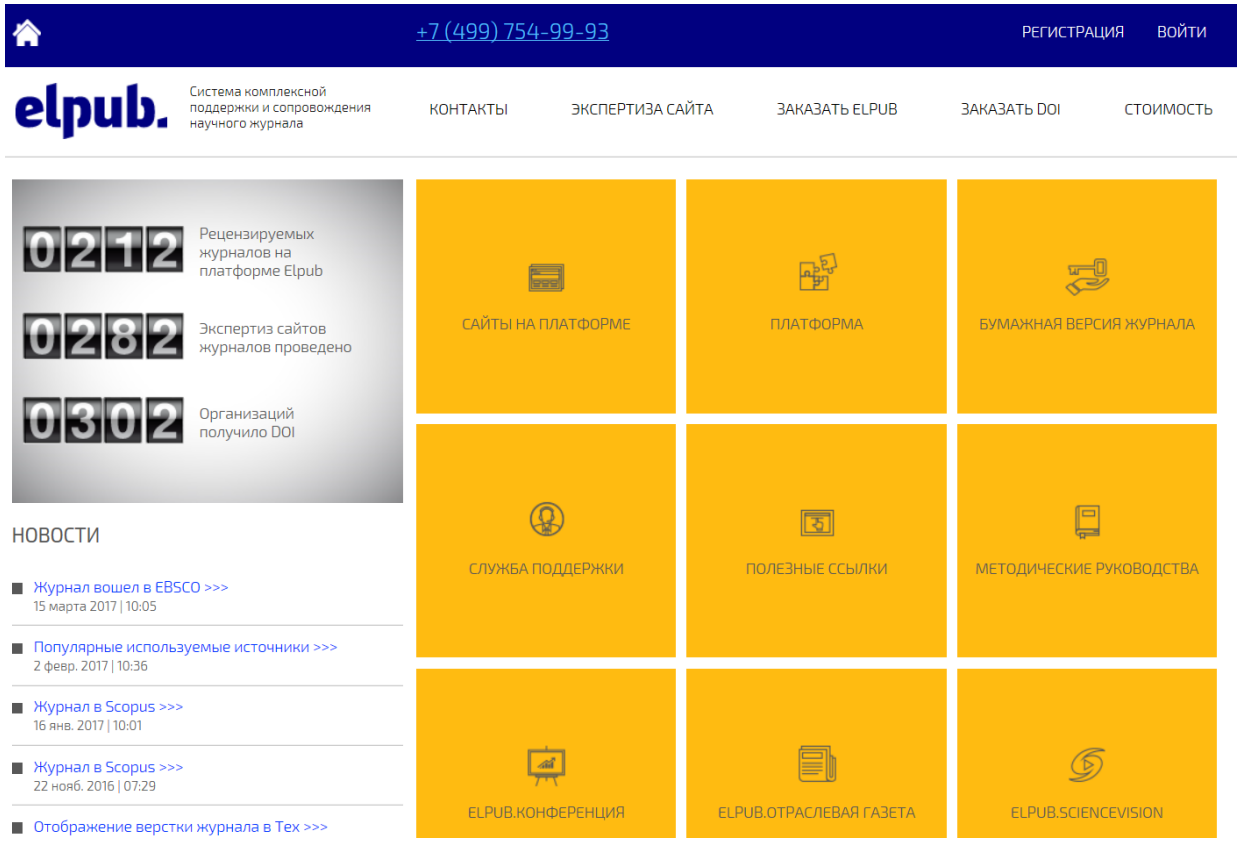
Коллекция **Медицина**

Новинка **> 52 тыс.**

Коллекция **Кольцо ВУЗов**

около 1 млн

Слайд про НЭИКОН



Антиплагиат теперь доступен подписчикам Elpub…

Конференция «Обнаружение заимствований – 2017» 26-27 октября 2017 года, www.oz2017.ru

Цель конференции – формирование экспертной среды по обсуждению вопросов, обмену мнениями и выработке решений в области обнаружения заимствований на образовательном и научном пространствах

Тематика конференции: Нормативные основы использования средств ОЗ. Приказ № 636. Требования ВАК. Рекомендации Рособрнадзора. Правовые основы и проблемы систем и средств ОЗ. Методология внедрения и использования технических средств ОЗ. Методология экспертизы и принятия решений. Этические вопросы ОЗ. Программно-технические средства ОЗ. Алгоритмы и технологии ОЗ, поиска, обработки и анализа текстов. Средства и системы информационной поддержки ОЗ, электронно-библиотечные системы, системы открытого доступа. Образовательные программы и технологии формирования компетентностей ОЗ.

Место проведение: г. Липецк, Россия

Обнаружение и профилактика некорректных заимствований

- Автоматизированные системы обнаружения заимствований
- Законодательная база
- Нормативная база
- Методическая база
- Экспертная поддержка
- Академическая этика

Artificial Intelligence in Russia

First Edition Analytical Report

2017

Table of Content

Landscape Overview 2017

84. Antiplagiat

АНТИПЛАГИАТ
твори собственным умом

Base: Moscow, Russia

Employees: 10 - 50

Established: 2005

Platform that helps to find plagiat in works.

Industries / Functions

Education / Analytics

AI Sphere

Natural Language Processing

ВАШИ ВОПРОСЫ

Юрий Чехович

antiplagiat.ru

8 (800) 777 81 28
+7 (495) 223 23 84

chehovich@antiplagiat.ru