

# Анализ алгоритмов и программных средств сравнения документов

Выполнил: Студент группы ДА-32

Ищенко Ярослав Иванович

Научный руководитель: доцент, кандидат технических наук

Капшук Олег Алексеевич

# Цель работы

- Проанализировать алгоритмы которые используются для сравнения документов.
- Протестировать программные средства для сравнения документов.
- Проанализировать использование программных средств для сравнения документов в борьбе с плагиатом.

# Актуальность

Исследования показывают, что почти 60% офисных работников сталкиваются с задачей сравнения документов, при этом каждый пятый сотрудник занимается такой деятельностью регулярно, например чтобы снизить риск подписания некорректной версии договора из-за изменений, которые могли быть внесены контрагентом. В 78% случаев юристы, представители договорных отделов, логисты, менеджеры по продажам, финансисты и другие офисные сотрудники сравнивают бумажные экземпляры документов с электронными, в 61% - файлы Microsoft Word между собой, немного реже - копии в формате PDF с документами MS Office ( 51%) и отсканированные изображения документов с их электронными дубликатами (37%). В наличии сильная разрозненность форматов данных, вкупе с низкой степенью автоматизации существенно усложняет процесс поиска несоответствий в документах, особенно в случае их порядковой сверки. Неудивительно, что в некоторых организациях закрывают глаза на подобные вещи, совершенно не подозревая, насколько рискованным может быть такой подход к организации документооборота в компании.

# Алгоритм diff

Работа diff основана на нахождении наибольшей общей подпоследовательности (англ. longest common subsequence, проблема LCS). Например, имеются две последовательности элементов:

- a b c d f g h q z
- a b c d e f g l j k r x y z

и надо найти наиболее длинную последовательность элементов, которая представлена в обеих последовательностях в одинаковом порядке. Это означает, что необходимо найти новую последовательность, которая может быть получена из первой последовательности удалением некоторых элементов или из второй последовательности удалением других элементов. В данном случае такой последовательностью будет являться.

- a b c d f g j z

После получения наибольшей общей последовательности остаётся только небольшой шаг до получения похожего на diff вывода:

```
e h i k q r x y
+ - + + - + + +
```

# Алгоритм шинглов

- Канонизация текста
- Разбиение на шинглы
- Вычисление хэшей шинглов с помощью 84х статических функций
- Случайная выборка 84 значений контрольных сумм
- Сравнение, определение результата

# Сравнение двух Word файлов

Данный параметр позволяет сравнить два документа и вывести на экран только несовпадающие фрагменты. Сравнимые документы не изменяются. При таком способе сравнения различия между документами всегда отображаются в новом, третьем документе.

Если нужно сравнить исправления, сделанные несколькими рецензентами, не следует использовать данный параметр. Выберите команду **Объединение исправлений от нескольких авторов в одном документе**.

1. Откройте документы, которые нужно сравнить.
2. На вкладке **Рецензирование** в группе **Сравнить** нажмите кнопку **Сравнить**.
3. Выберите пункт **Сравнение двух версий документа (юридическое примечание)**.
4. В поле **Исходный документ** укажите документ, который будет использоваться в качестве исходного.
5. В поле **Измененный документ** выберите документ, который нужно сравнить с уже открытым документом.
6. Щелкните **Больше**, а затем укажите параметры сравнения документов. Рядом с **Показывать изменения** выберите отображение изменений на уровне знаков или на уровне слов.
7. Если результаты сравнения не требуется выводить в третьем документе, укажите документ, в котором должны быть отображены изменения.
8. Нажмите кнопку **ОК**.
9. Если какая-либо из версий документа содержит записанные исправления, на экране отобразится соответствующее сообщение. Чтобы принять исправления и сравнить документы, нажмите кнопку **Да**.
10. Откроется новый документ, в котором будут приняты исправления, записанные в исходном документе, а изменения, отмеченные во втором из сравниваемых документов, отобразятся в виде исправлений.
11. Сравнимые документы не изменяются.

# Сравнение файлов в стандартном выпуске Acrobat DC (PDF)

Для обнаружения различий между двумя версиями файлов PDF можно использовать функцию Сравнить документы. После сравнения двух файлов Acrobat создаст подробный отчет обо всех различиях, включая различия в содержании текста, в шрифтах, изображениях и даже в порядке страниц.

1. Выбрать Инструменты > Сравнить документы.
2. Выбрать документы для сравнения. Сравнение можно ограничить фрагментом документов, введя номера страниц в поля «Первая страница» и «Последняя страница» под каждым именем файла.
3. В разделе Описание документа выберите тип сравниваемых документов. Чтобы игнорировать различия между графическими элементами, выберите Сравнить только текст. Нажмите кнопку ОК. Acrobat отобразит результаты в новом документе. Первая страница содержит сводку результатов сравнения файлов.
4. Выбрать способ просмотра различий.
5. Использовать панель **Сравнить** с левой стороны (по умолчанию).
  1. Страницы отображаются на панели **Сравнить** в виде миниатюр. Значок под миниатюрой обозначает различие. Зеленая стрелка обозначает перемещенные или удаленные страницы. Пурпурная стрелка указывает на наличие изменений на странице.
  2. Вернитесь в первую страницу и щелкните ссылку **Начать**, чтобы перейти к первому различию.
  3. Изменения выделяются на основном экране. Наведите указатель мыши на выделенный элемент, чтобы просмотреть подробные сведения. В правом верхнем углу находится описание цветовых обозначений.
6. Использовать параллельный просмотр документов.
  1. На панели **Сравнить** слева нажмите значок параметров и выберите **Одновременный просмотр**.
  2. Изменения выделяются на основном экране. Наведите указатель мыши на выделенный элемент, чтобы просмотреть подробные сведения. В правом верхнем углу находится описание цветовых обозначений.

# Сравнение документов Word с PDF

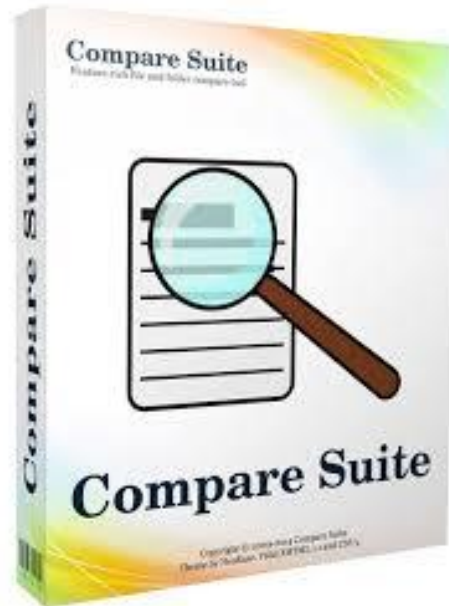
Документ PDF можно сравнить с документом Word. Типичный сценарий имеет место, если из документа Word был создан PDF-файл, после чего документ был изменен, и теперь нужно найти эти различия.

PDF-файл передается в программу Convert Assistant для преобразования в документ Word. Сравнение двух документов выполняется в Word, что позволяет выполнить визуальное или текстовое сравнение с результатом, отображаемым во временном файле, который при желании можно сохранить.

1. Выбрать пункты «Домой» > «Обработка» > Сравнение документов. Откроется диалоговое окно «Сравнение документов».
2. Применить открытый в данный момент документ PDF в качестве более старого или нажмите кнопку Обзор, чтобы выбрать более новый файл в диалоговом окне «Открыть». В поле Редакция выберите любую из доступных редакций документа.
3. Нажмите кнопку Обзор и выберите в качестве типа файла документ Word (DOC или DOCX), а затем выберите нужный документ Word в диалоговом окне «Открыть».
4. Выбрать тип отчета для представления результата сравнения: Рядом или Совмещено.
5. Нажмите кнопку ОК. Подтвердите преобразование из формата PDF и нажмите кнопку Преобразовать все в программе PDF Converter Assistant. Копия исходного PDF-файла, доступная только для чтения, будет отображена в Microsoft Word с расширением PDF.
6. При необходимости подтвердите преобразование файла из формата RTF в файл Word.



# Программные средства для сравнения документов



# ABBYY Comporator

ABBYY Comporator - Договор\_1.doc / Договор\_2.tif

ФАЙЛ ПРАВКА ВИД СПРАВКА

ДОКУМЕНТ 1 D:\Documents\4. COMPARATOR\Демо-документы\NE ДОКУМЕНТ 2 D:\Documents\4. COMPARATOR\Демо-документы

РАЗЛИЧИЯ (4)

Текст исправлен стр. 1 <> 1  
"без" на "с возможностью"

Текст удален стр. 1 <> 1  
2.4.2. Отказаться полностью или частично от услуг Исполнителя, ...

Текст исправлен стр. 2 <> 2  
"0,1%" на "0,01%"

Текст исправлен стр. 2 <> 2  
"3 (Трех)" на "4 (четырёх)"

Искать различия в языке:  
Русский

Искать различия в пунктуации  
 Искать различия в одной букве

НАЙТИ РАЗЛИЧИЯ

ДОГОВОР № 65-1201/059- 26.01.R1.12 на рекламные услуги

г. Москва «26» января 2012 г.  
ООО «Ромашка», именуемое в дальнейшем «Заказчик», в лице Генерального директора Иванова Г.Г., действующего на основании Устава, с одной стороны, и ООО «Солнышко», именуемое в дальнейшем «Исполнитель», в лице Генерального директора Петровой Е.А., действующего на основании Устава, с другой стороны, далее вместе именуемые «Стороны», заключили настоящий Договор о нижеследующем:

**1. ПРЕДМЕТ ДОГОВОРА**

1.1. Исполнитель обязуется оказать услуги по обеспечению размещения рекламных материалов, предоставленных Заказчиком, на ресурсах, выходящих в распоряжение Исполнителя, а Заказчик обязуется принять и оплатить услуги Исполнителя.

1.2. Перечень рекламных ресурсов, сроки и объем размещения рекламных материалов, а также иные условия, которые Стороны сочтут существенными для размещения рекламных материалов, определяются в Приложениях к настоящему Договору. Приложения подписываются уполномоченными представителями обеих Сторон и после подписания становятся неотъемлемой частью настоящего Договора.

**2. ПРАВА И ОБЯЗАТЕЛЬСТВА СТОРОН**

**2.1. Исполнитель обязан:**

2.1.1. Обеспечить размещение предоставленных Заказчиком рекламных материалов в соответствии с Приложениями к настоящему Договору.

2.1.2. Оказывать услуги, предусмотренные настоящим Договором, своевременно и с надлежащим качеством.

2.1.3. Информировать Заказчика (по его требованию) о ходе оказания услуг по настоящему Договору.

2.1.4. Обеспечить Заказчика статистическими данными по размещению рекламных материалов Заказчика. Перечень предоставляемых данных, если таковые необходимы, указывается в Приложениях.

**2.2. Исполнитель вправе:**

2.2.1. Требовать оплаты услуг, оказываемых по настоящему Договору, оставаясь при этом ответственным перед Заказчиком за действия третьих лиц.

**2.3. Заказчик обязан:**

2.3.1. Осуществлять оплату услуг Исполнителя в соответствии с условиями настоящего Договора и Приказаний к нему.

2.3.2. Предоставлять Исполнителю информацию, необходимую последнему для качественного и своевременного оказания услуг по настоящему Договору.

2.3.3. Предоставлять рекламные материалы, соответствующие техническим и иным требованиям Исполнителя, указанным в Приложениях к настоящему Договору, а также требованиям действующего законодательства РФ о рекламе.

2.3.4. В случае досрочного расторжения Договора оплатить фактически оказанные Исполнителем услуги.

**2.4. Заказчик вправе:**

2.4.1. Проверять ход и качество оказания услуг, предусмотренных Договором, **без вмешательства в деятельность Исполнителя.**

2.4.2. Отказаться полностью или частично от услуг Исполнителя, предупредив последнего не менее чем за 30 (Тридцать) дней до предполагаемой даты отказа. В этом случае Стороны составляют Акт об оказании услуг, отражающий фактический объем оказанных услуг, их качество и сумму, на которую они оказаны.

**3. СТОИМОСТЬ УСЛУГ И ПОРЯДОК РАСЧЕТОВ**

3.1. Стоимость услуг по обеспечению размещения каждого вида рекламных материалов и порядок их оплаты определяются в Приложениях к настоящему Договору.

3.2. Все платежи по Договору осуществляются в рублях путем перечисления Заказчиком денежных средств на расчетный счет Исполнителя.

**4. ПОРЯДОК ПРИЕМКИ ОКАЗАННЫХ УСЛУГ**

4.1. Факт выполнения обязательств Стороны фиксируют Актом об оказании услуг.

4.2. Исполнитель ежемесячно отправляет Заказчику счета-фактуры и Акты об оказании услуг.

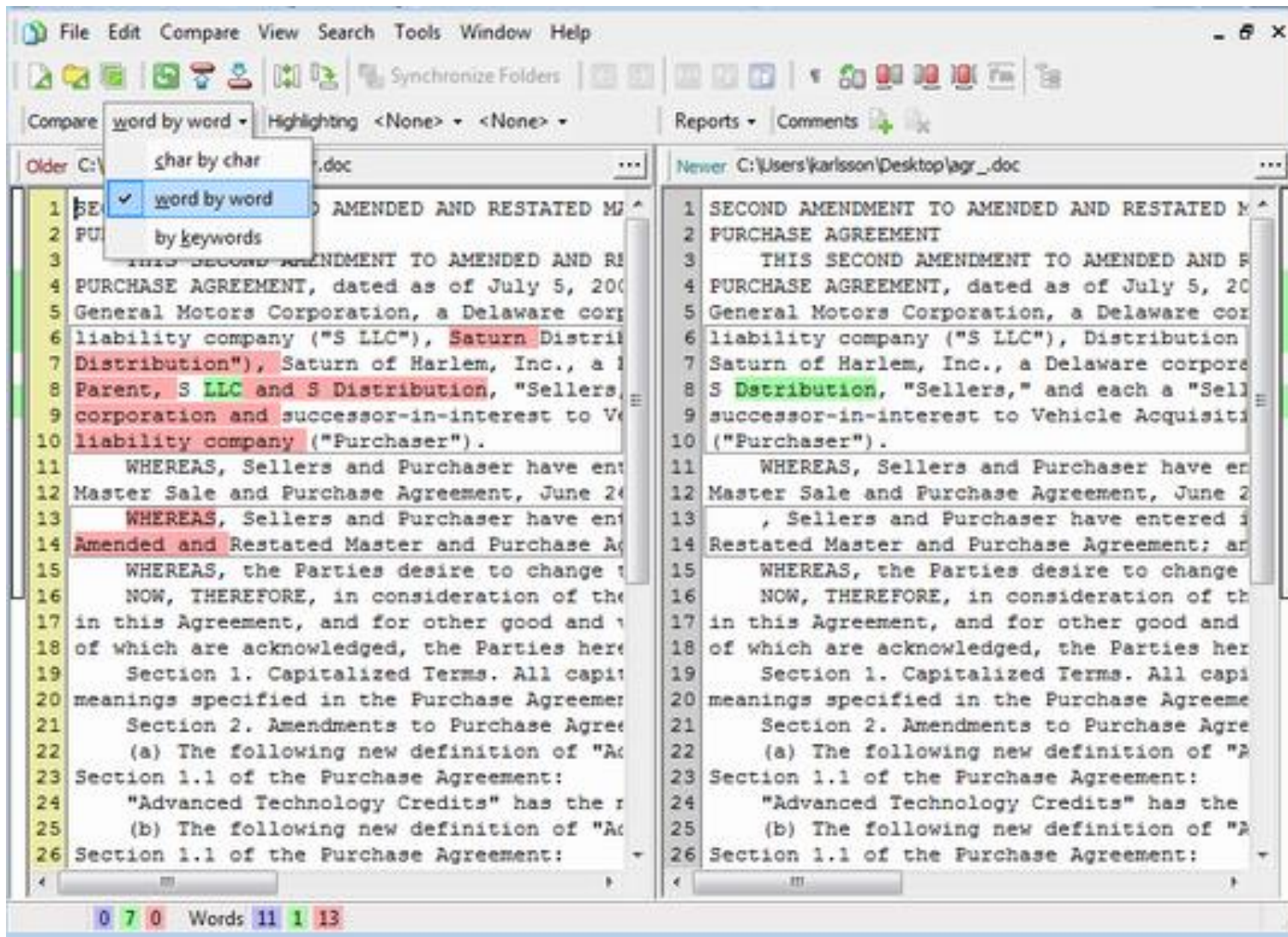
4.3. Заказчик должен принять счета-фактуры и Акты об оказании услуг, подписать и отправить или передать Исполнителю вторые экземпляры выставленных Актов об оказании услуг не позднее 5 (пяти) дней после их получения, или направить в указанный срок мотивированный отказ от подписания Акта об оказании услуг.

4.4. Услуги, оказанные Исполнителем, будут считаться реализованными и принятыми ежемесячно на основании Акта об оказании услуг. При получении Исполнителем подписанного Заказчиком Акта об оказании услуг и мотивированного отказа от подписания Акта об оказании услуг более 5 (пяти) дней факт выставления Исполнителем Акта будет считаться как принятие услуг Заказчиком и закрытие очередного расчетного периода услуг по настоящему Договору.

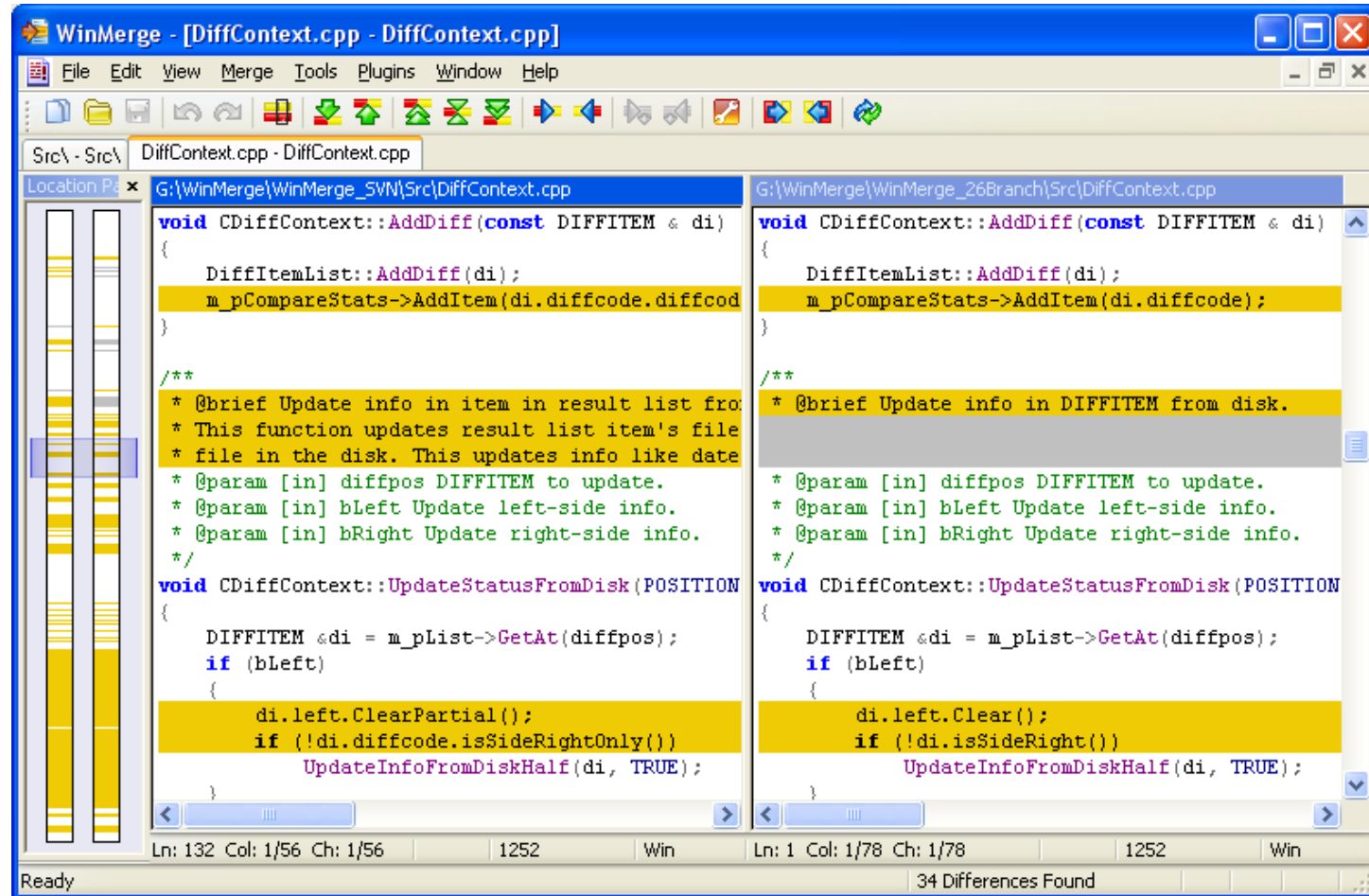
**5. ОТВЕТСТВЕННОСТЬ СТОРОН**

5.1. В случае нарушения условий настоящего Договора виновная Сторона несет ответственность в

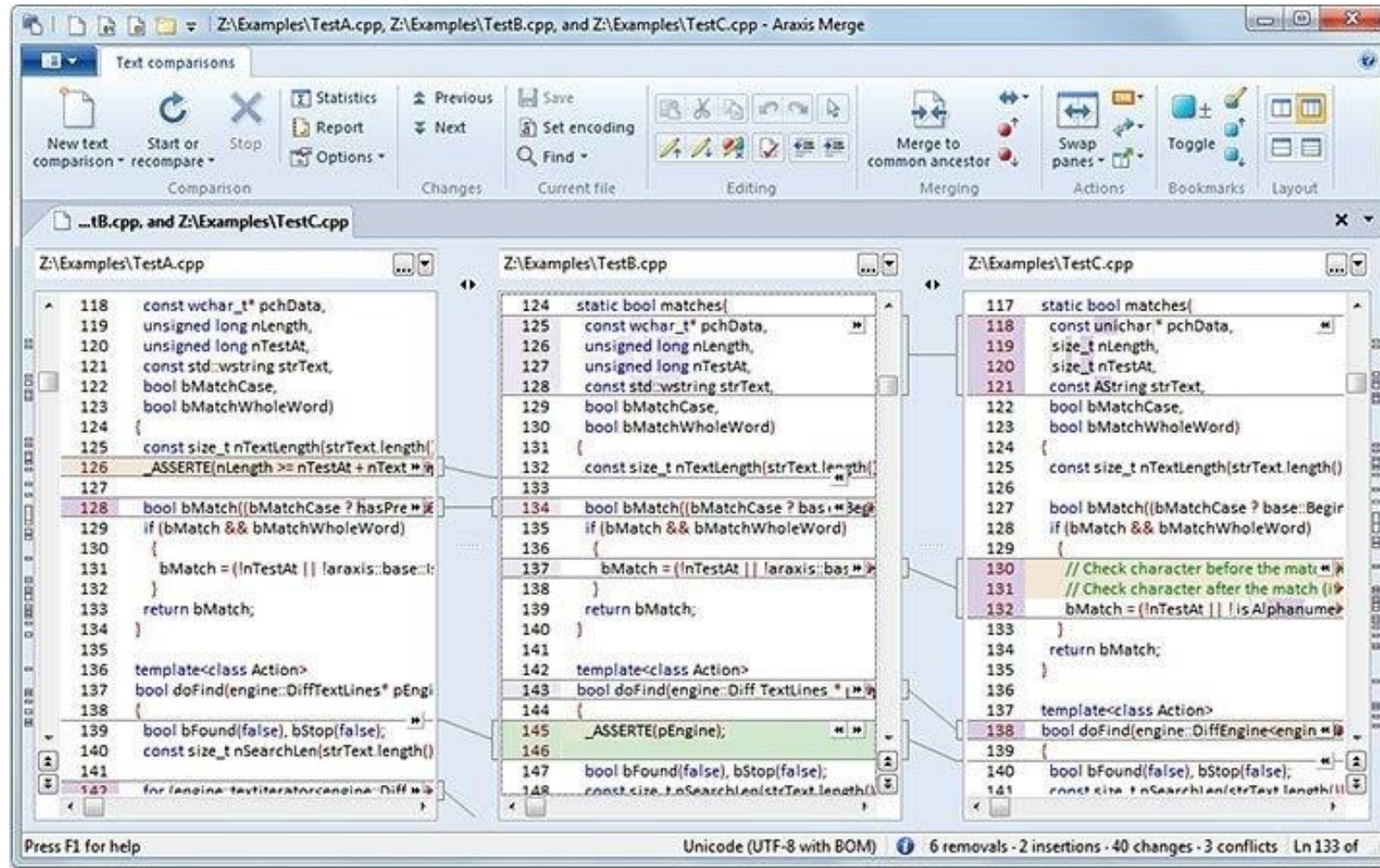
# Compare Suite Pro



# WinMerge



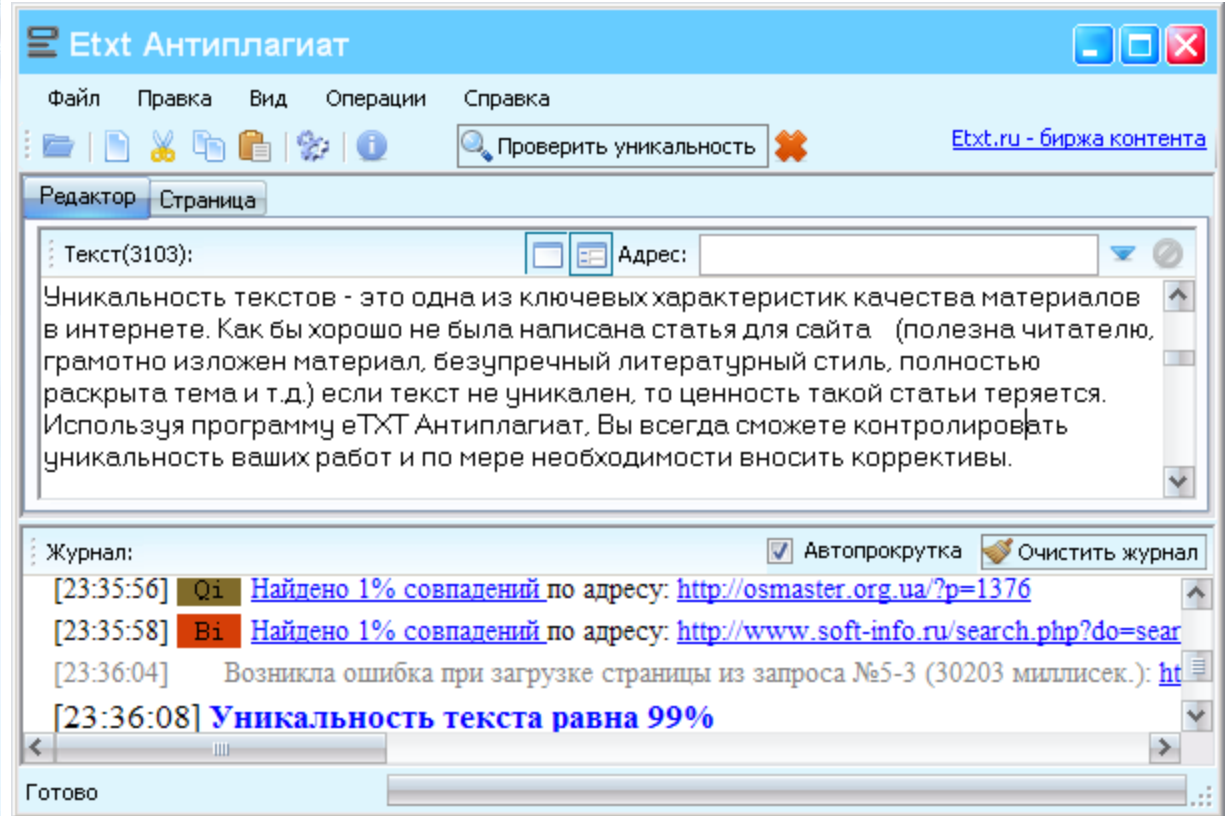
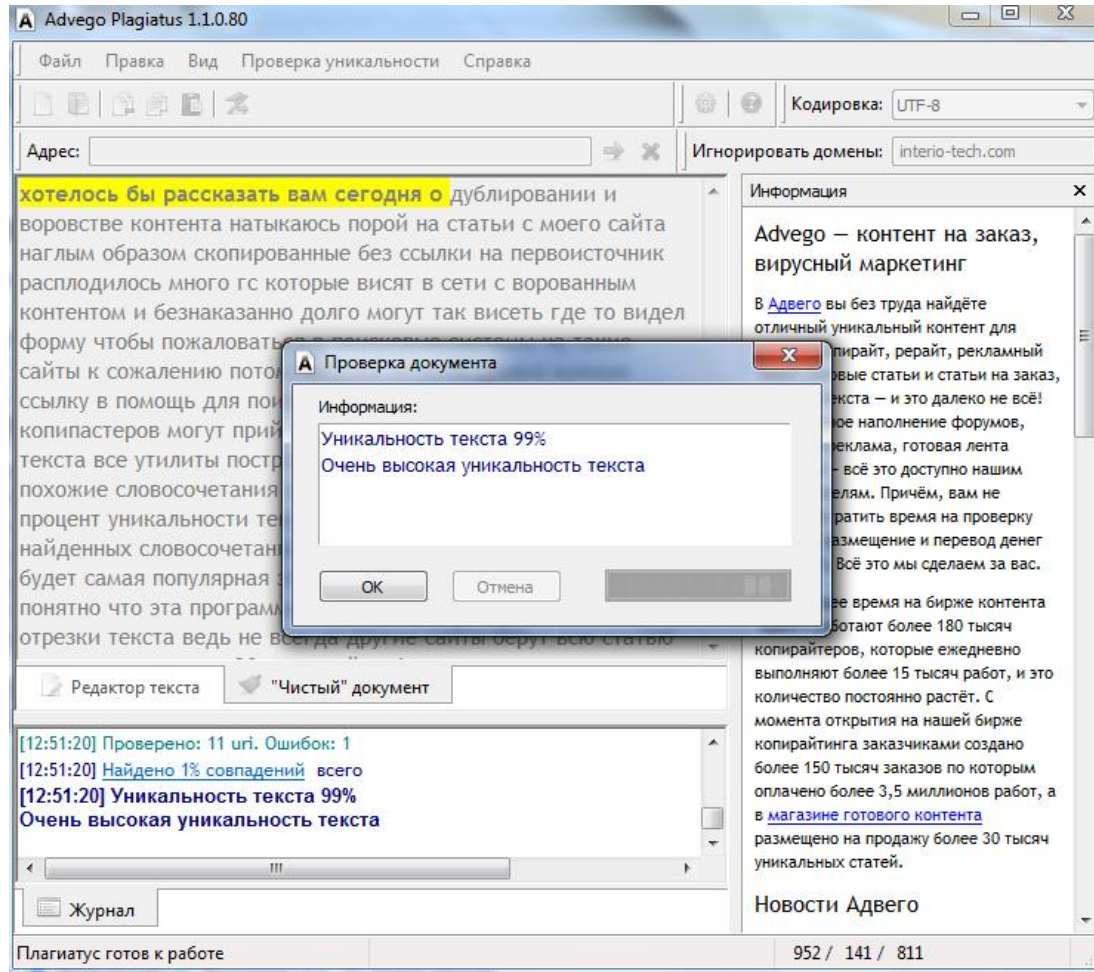
# Araxis Merge



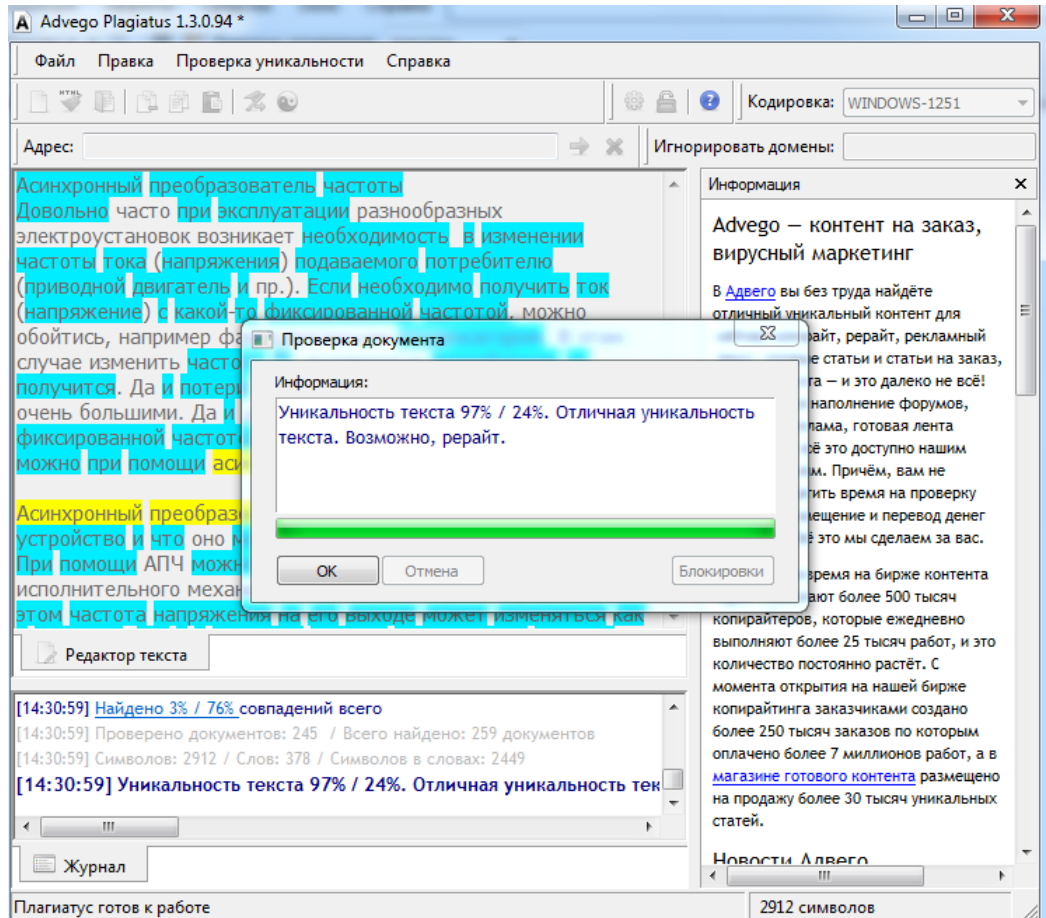
# Использование программ для сравнения документов в борьбе с плагиатом

- Система «Антиплагиат»
- Программа Advego Plagiatus
- Сервис Unplag
- Сервис [www.miratools.ru](http://www.miratools.ru)
- Сервис [www.istio.com](http://www.istio.com)
- Программа Praide Unique Content Analyser II
- Система Plagiatinform
- Сервис Copyscape

# Программы для проверки на плагиат



# Advego Plagiatus





# Etxt Антиплагиат

The screenshot displays the Etxt AntiPlagiat software interface. It features a main window with two panes for text comparison and a bottom panel for a journal of findings.

**Text Comparison:**

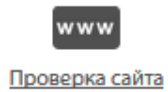
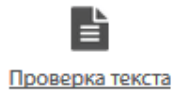
- Текст(2994):** энергией в том случае, когда все электроны будут находиться в самом низком состоянии. На самом деле это не так, поскольку, согласно принципу запрета Паули, в данном энергетическом состоянии, определяемом всеми четырьмя квантовыми числами, не может находиться больше одного электрона. Учитывая, что спиновое квантовое число, может быть равным только  $\pm 1/2$ , принцип Паули иногда определяют следующим образом: в данном энергетическом состоянии, определяемом квантовыми числами  $n, l$  и  $m$ , может находиться не более двух электронов с противоположными спинами. Сочетание принципа запрета Паули с ограничениями величин приводит к следующим выводам: Максимальное число электронов, которое может находиться в состоянии, характеризующемся главным квантовым числом  $1, 2, 3, \dots, n$ , соответственно равно  $2, 6, 18, 32, \dots, 2n^2$  степени.
- Канонизированный текст(4060):** иногда определяют следующим образом: в данном энергетическом состоянии определяемом квантовыми числами  $n$  и  $l$  может находиться более двух электронов противоположными спинами. Сочетание принципа запрета Паули с ограничениями величин приводит к следующим выводам: Максимальное число электронов которое может находиться в состоянии характеризующемся главным квантовым числом  $1, 2, 3, \dots, n$  соответственно равно  $2, 6, 18, 32, \dots, 2n^2$  степени данного значения  $n$ . Максимальное число электронов которые могут находиться в состоянии, соответственно равно  $2, 6, 18, 32, \dots, 2n^2$  степени.

**Журнал:**

- [10:17:24] Найдено 2% совпадений (78 миллисек.) по адресу: [http://dic.academic.ru/dic.nsf/enc\\_physics/2988/%D0%AD%D0%9B%D0%95%I](http://dic.academic.ru/dic.nsf/enc_physics/2988/%D0%AD%D0%9B%D0%95%I)
- [10:17:24] Найдено 1% совпадений (500 миллисек.) по адресу: <http://www.pereplet.ru/obrazovanie/stsoros/529.html> (78 миллисек.)
- [10:17:26] Найдено 1% совпадений (531 миллисек.) по адресу: <http://www.b-i-o-n.ru/theory/atom/himicheskaja-svjaz> (281 миллисек.)
- [10:18:14] Найдено 5% совпадений (2718 миллисек.) по адресу: <http://www.istu.edu/export/sites/istu.v3/ru/education/faculty/zvf/download/con>
- [10:18:16] Найдено 3% совпадений (1109 миллисек.) по адресу: [http://chem.msu.su/rus/school/zhukov1/06.html\(Сохраненная копия\)](http://chem.msu.su/rus/school/zhukov1/06.html(Сохраненная копия)) (578 мил
- [10:18:16] **Уникальность текста равна 89%**

Готово

# Онлайн сервисы для проверки на плагиат




Проверить текст на уникальность


Длина текста: 0 (без пробелов: 0)

Игнорировать сайт:   Запомнить


Проверить

 **TEXT.RU**  
Текст.ру


О проекте: [Новости](#) | Служба поддержки: +7 (495) 789-02-33 | 0,00 рублей | [Пополнить / Вывод средств](#) | [PRO-Аккаунт](#) | [Выйти](#) | [RU](#)

 ПОДАРКИ!


- Биржа копирайтинга
- Биржа рерайтинга
- Магазин статей
- Магазин новостей



- Уникальность текста
- Проверка орфографии
- SEO-анализ
- Синонимы к слову



- Уникальность сайта
- Уникальность документа
- Регулярная проверка
- API уникальности
- Пакеты символов 15 тыс



## Онлайн-сервис проверки текста на уникальность

Сервис онлайн проверки текста на уникальность TEXT.RU покажет процент уникальности текста. Глубокая и качественная проверка найдет дубликаты и рерайт.

Эффективные алгоритмы позволили сделать проверку глубокой и точной. Автоматическая проверка орфографии подскажет, где были допущены ошибки.

Бесплатное определение уникальности текстов.

Нет ограничений. Зарегистрировавшись у нас на сайте, вы полностью снимете ограничения на проверку и получите возможность проверить неограниченное количество текстов.

Вставьте текст...

Проверить на уникальность

Всего символов 0 | Без пробелов 0 | Количество слов 0

# Content-watch.ru

Вход Регистрация



Проверка текста



Проверка сайта



Защита сайта



Магазин подписок



API

## Проверить текст на уникальность

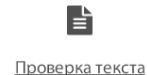
Длина текста: 1369 (без пробелов: 1167)

Добро пожаловать в кошачий мир! Если вы любите или любите кошек и не можете устоять перед этими разнообразными, характерными, но в то же время очень нежными животными, тогда эта социальная сеть то, что нужно. Здесь вы сможете узнать много нового и интересного о самых разнообразных породах кошек, об их особенностях и отличиях. Эта социальная сеть создана для настоящих поклонников кошек и котят и призвана помочь поближе узнать своего домашнего питомца или же подобрать для себя наиболее подходящую породу по характеру и внешнему виду. Социальная сеть для тех, кто любит кошек приглашает всех желающих узнать самую актуальную информацию о кошках и о новых выставках с ними в главной роли. Более того, на страницах сети можно узнать о всевозможных нюансах содержания котят и о тонкостях ухода за определенными породами. Также здесь вы узнаете актуальные адреса, номера телефонов ветеринарных клиник и ближайших питомников. Нельзя не отметить тот факт, что в социальной сети вы лично сможете вести блог и делиться с другими посетителями сайта полезной информацией, общаться с людьми, разделяющими любовь к кошкам, а также оставлять комментарии и отзывы под статьями, фотографиями и видеозаписями других пользователей. Присоединяйтесь к сообществу любителей кошек и получайте море полезной информации, а также позитив от просмотра фото- и видеороликов с участием котят.

Уникальность текста: 100.0%

ПРАВИТЬ ЭТОТ ТЕКСТ

НОВАЯ ПРОВЕРКА



Проверка текста



Проверка сайта



Защита сайта



Магазин подписок



API

## Проверить текст на уникальность

Длина текста: 972 (без пробелов: 850)

Сварочный трактор удобный и функциональный аппарат для дуговой сварки, оснащенный автоматической подачей проволоки. Устройство может самостоятельно передвигаться вдоль вектора свариваемого шва по поверхности материала или специальным рельсам, и легко переносится из одного места в другое. Такая мобильность позволяет использовать сварочный агрегат практически в любых условиях. Все параметры сварки можно задать в автоматическом режиме, выбрать скорость подачи проволоки, тока, напряжения дуги и сделать тонкие настройки. Блок управления и проволока крепятся к специальной тележке. Сварочные трактора могут работать под флюсом или в среде защитных газов с использованием разных видов проволоки, а в памяти современных устройств может храниться более 200 программ сварки. Использование автоматических агрегатов всегда увеличивает скорость работы, повышая эффективность производства. Таким образом, роль человека в данном процессе сводится к управлению и контролю аппарата.

Игнорировать сайт:   Запомнить

Уникальность текста: 88.8%

[показать все совпадения](#)

URL страницы, где найдено совпадение	Сколько совпало	Что совпало
<a href="http://www.arsenal-orel.ru/index.php?productID=318431">http://www.arsenal-orel.ru/index.php?productID=318431</a>	6.1%	<a href="#">показать</a>
<a href="http://www.one-power.ru/p/...">http://www.one-power.ru/p/...</a>	5.0%	<a href="#">показать</a>

НУЖЕН УНИКАЛЬНЫЙ КОНТЕНТ?

ПРАВИТЬ ЭТОТ ТЕКСТ

НОВАЯ ПРОВЕРКА

# Text.ru

Проверка уникальности  
Загрузка

Проверка орфографии  
Загрузка

SEO-анализ текста  
Загрузка

завантажити вихідний файл. Потім з меню "Сервіс" відкрити команду "Порівняти і об'єднати виправлення", вказати файл, порівнюваний з вихідним, і включити прапорець "Чорні рядки". Тільки після цих маніпуляцій кнопка "Об'єднати" перетвориться в кнопку "Порівняти", і при натисканні на даній кнопці програма і проведе порівняння файлів. Результати порівняння будуть показані в новоствореному документі в традиційному режимі рецензування. З появою Word 2007 все стало набагато простіше, оскільки тепер досить перемкнутися на вкладку "Рецензування", клацнути по кнопці "Порівняти" і вказати порівнювані версії документа. Теоретично, в Excel теж можливо порівняння документів вбудованими засобами, правда, тільки при роботі в режимі фіксації змін. Однак це незручно, оскільки кожну з змінних осередків доведеться переглядати, наводячи на неї миша, так як зміни, внесені в документ, відображаються в спливаючих вікнах (приблизно таких, як звичайні примітки). Подруге, якщо названий режим не буде попередньо включений (команда "Сервіс">"Виправлення">"Виділити виправлення", прапорець "Відстежувати виправлення"), то зробити порівняння XLS-файлів потім виявиться неможливо. Що стосується швидкого порівняння PDF-документів, то така можливість, звичайно, мається на Acrobat 9 Pro і Acrobat 9 Pro Extended, але ці рішення встановлені далеко не на кожному комп'ютері. Тому при необхідності швидкого порівняння

Версии текста:  
4 минуты назад (UTC +03:00)  
Пустой текст

7 минут назад (UTC +03:00)  
Уникальность 0% Орфография 34  
Всего символов 3767 Заспамленность 85%..  
Без пробелов 3302 Вода 6%  
Количество слов 456

**Проверка уникальности**  
Уникальность: **0.00%**  
[818.ru-wiki.ru/wiki/Латентно-семантически...](#) 100%  
[kurp.ru-wiki.ru/wiki/Латентно-семантическ...](#) 100%  
[Подробнее](#)

**Проверка орфографии**  
В тексте найдены 34 ошибки:

- нейросети
- термы-на-документы
- SVD-разложения

[Подробнее](#)

**SEO-анализ текста**  
Всего символов: **3767** Заспамленность: **85%**  
Без пробелов: **3302** Вода: **6%**  
Количество слов: **456**  
[Подробнее](#)

Подсвечено: ■ Ошибки в тексте

ЛСА можно сравнить с простым видом **нейросети**, состоящей из трех слоев: первый слой содержит множество слов (термов), второй – некое множество документов, соответствующих определенным ситуациям, а третий, средний, скрытый слой представляет собой множество узлов с различными весовыми коэффициентами, связывающих первый и второй слои.

В качестве исходной информации ЛСА использует матрицу **термы-на-документы**, описывающую набор данных, используемый для обучения системы. Элементы этой матрицы содержат, как правило, веса, учитывающие частоты использования каждого термина в каждом документе и участие термина во всех документах (TF-IDF). Наиболее распространенный вариант ЛСА основан на использовании разложения матрицы по сингулярным значениям (SVD – Singular Value Decomposition). С помощью **SVD-разложения** любая матрица раскладывается во множество ортогональных матриц, линейная комбинация которых является достаточно точным приближением к исходной матрице.

Говоря более формально, согласно теореме о сингулярном разложении[8], любая вещественная прямоугольная матрица может быть разложена на произведение трех матриц:

$$\{\displaystyle \{\begin{matrix} A=USV \end{matrix}\} \{\begin{matrix} A=USV \end{matrix}\}$$

где матрицы  $\{\displaystyle \{\text{bf}\{U\}\} \{\text{bf}\{U\}\}$  и  $\{\displaystyle \{\text{bf}\{V\}\} \{\text{bf}\{V\}\}$  – ортогональные, а  $\{\displaystyle \{\text{bf}\{S\}\} \{\text{bf}\{S\}\}$  – диагональная матрица, значения на диагонали которой называются сингулярными значениями матрицы  $\{\displaystyle \{\text{bf}\{A\}\} \{\text{bf}\{A\}\}$ . Буква Т в выражении  $\{\displaystyle \{\text{bf}\{V\}\} \{T\} \{\text{bf}\{V\}\} \{T\}$  означает транспонирование матрицы.

Такое разложение обладает замечательной особенностью: если в матрице  $\{\displaystyle \{\text{bf}\{S\}\} \{\text{bf}\{S\}\}$  оставить только  $\{\displaystyle \{\text{bf}\{k\}\} \{\text{bf}\{k\}\}$  наибольших сингулярных значений, а в матрицах  $\{\displaystyle \{\text{bf}\{U\}\} \{\text{bf}\{U\}\}$  и

Версии текста:  
1 минуту назад (UTC +03:00)  
Уникальность 0% Орфография 34  
Всего символов 3767 Заспамленность 85%..  
Без пробелов 3302 Вода 6%  
Количество слов 456

© 2014 [Доступность проверки](#)

# Выводы

- 1. Проанализированы распространенные алгоритмы, которые на сегодняшний день используются в программных средствах для сравнения документов. В современных программных средствах для сравнения документов используются всего несколько алгоритмов, а именно Diff и шинглы, а во многих программных средствах рассмотренных в работе вообще не документированы.**
- 2. Алгоритм Diff реализован в MS Office, файлы с текстами программ (средства контроля версий CVS и Subversion, утилита UNIX diff) и т. д. Этот алгоритм используется в качестве составной части алгоритмов слияния (merge) двух различных версий файлов в системах версионного контроля, при синхронизации данных в системах с ограничениями на трафик, при объединении серверных и клиентских данных (например, в случае обрыва соединения) и т. д. Однако при всей своей популярности этот алгоритм имеет лимит на применение, главный из которых - удобная визуализация результатов. Например, если попытаться воспользоваться функцией diff для сравнения Word-документов, встроенной в MSOffice, то если сравниваемые документы немалые (10 страниц и более), а сделанных изменений достаточно много, то принять такую разницу оказывается очень сложно - пользователь видит огромное количество выделенной информации, среди которых и изъяты абзацы, и различия в правилах пунктуации.**
- 3. Проблема алгоритма шинглов заключается в количестве сравнений, ведь это напрямую сказывается на производительности. Увеличение количества шинглов для сравнения характеризуется ростом операций, что критически сказывается на производительности.**

# Выводы

- 4. Функционал платных и бесплатных программ почти не отличается. Хотя главной их преимущественной особенностью является то, что они могут сравнивать бумажные и электронные документы.**
- 5. На сегодняшний день среди программ проверки текстов на уникальность (плагиат) можно выделить два лидера: Text.ru и Content-watch.ru.**
- 6. Text.ru лучше всего подойдет редакторам, заказчикам сайтов, преподавателям, так как позволяет провести максимально глубокую проверку (пусть иногда и с ложными срабатываниями).**
- 7. А вот авторам статей, создателям сайтов лучше все-таки подойдет Content-watch.ru так как в нём гораздо меньше ложных срабатываний. Анализировать уникальность технического рерайта в Text .ru - почти невозможно, так как сервис реагирует на малейшие совпадения по поводу и без.**

Спасибо за внимание