

# ВСЕ ПРО ТЕКСТ

## Где брать?

В этом учебном документе я расскажу основные способы генерации текста. Пошли:

### 1. Рерайт

На самом деле, он нам не особо подходит по той простой причине, что мы не знаем алгоритм, по которому Дзен проверяет уникальность. Однако есть темы, в которых написать что-то свое будет довольно затруднительно. Одна из таких – новости.

Как рерайтить? Открываем несколько источников новости, читаем их несколько раз, по очереди. Перепечатываем то, что запомнили, в редактор Дзена, без подглядываний. Подглядел, получил пессимизацию. Все просто.

Обычное перепечатывание, типа:

*«В Москве неизвестные убили человека»*

*«В столице России, кто-то завалил биологическое существо»*

НЕ ПОЙДЕТ.

Есть вариант нанять рерайтера. Они делают это быстро и качественно.

### 2. Копирайт

Начинается самое интересное. Копирайтеров можно найти на одной из бирж труда. Ссылки на сайты также смотри в отдельном файле, в папке *Дополнительно*. Из всех там прописанных бирж, особенно рекомендую text.ru.

Прошли регистрацию, закинули деньги. Кстати говоря, за тысячу знаков вполне реально отдать всего 30-35 рублей. Т.е стоимость полноценной статьи выставляйте 40 - 50 рублей. Отдельным файлом дали вам шаблон для составления заявки на копирайт.

### 3. Текст из книг

Данная тема не новая и многие её используют для наполнения своих сайтов. Итак, вам нужно посетить библиотеку или книжные магазины и найти книги подходящие под вашу тематику.

Например, у вас канал о рецептах. Просто покупаете дешевые сборники рецептов, коих большое кол-во. Переписываете рецепт, проверяете на уникальность и публикуете.

И так с другими тематиками: красота и уход, гороскопы, изотерика в целом, полезные советы и так далее.

Небольшой совет - не берите книги крупных и известных книжных издательств, они все есть в тырнете.

**Если используешь эти три способа, обязательно проверь результат через систему антиплагиат, все ссылки будут в отдельной папке *Дополнительно*.**

#### **4. Самый крутой способ.**

Как легально заработать на чужом контенте с помощью Дзен?

Ни для кого не секрет, что ВКонтакте куча авторов, пишущих на самые разнообразные темы. Но вот почти никто из них еще не запускался на Дзен. Важно написать автору и предложить совместную работу на новой площадке. Начальными условиями ставь: возможность копирования статей на другую площадку, прибыль 50/50, автор докажет техподдержке, что у него есть авторское право на данный контент. Вполне возможно, что 90% авторов откажутся, но тут нужно научиться играть с цифрами и повышением ценности предложения. Предложите соотношение прибыли 60/40, расскажите писателю, что Дзен – это новые читатели и поклонники, что он бесплатен и т.д. Однако, важно так же убедить автора, что он не справится без вас, потому что Дзен – огромный «организм», и чтоб с ним подружиться, нужно много знать и уметь.

Когда вы договорились с тремя авторами, можете делегировать оформление статей своему помощнику, или еще кому-нибудь, главное, проверенному. Сделайте контент-план на пару месяцев наперед. Выведете на монетизацию, вуаля, у вас пассивный заработок. Да, небольшой, часть нужно отсылать автору. Но с трех каналов капает уже неплохо, а если шесть?

Основные плюсы:

- Тебе не придется тратить деньги на копирайтера
- Используя старые работы автора, можно прописать контент на месяц.
- Реклама у этого же автора в группе, или на другом ресурсе

Главное не запороть все пессимизацией, и вовремя доказать, что ты – это автор. Для этого можно дать писателю логин и пароль от почты и дать самому связаться с техподдержкой.

## 5. Сохранение контента с выпавших из индекса сайтов

Полностью бесплатный метод получения уникального контента – брать его с уже не работающих сайтов. Ежедневно тысячи доменов (по данным доменных регистраторов) освобождаются по истечении срока продления. Среди них можно найти сайты с контентом, который подойдет для наполнения ваших каналов на Дзене. Процесс заключается в следующем:

1. Поиск освободившихся доменов.
2. Анализ и отбор сайтов, с каких будем брать контент.
3. Получение статей с сайтов.
4. Проверка полученного контента на уникальность.

### Поиск доменов

Для поиска освободившихся доменов можно воспользоваться любым сервисом, предоставляющим список освобождающихся доменов. Я использую [mydrop.io](http://mydrop.io).

Регистрируемся там. Затем открываем на главной странице фильтр (ссылка «Показать фильтр»). В фильтре выбираем:

1. Зона **ru**
2. Статус – **Свободные**
3. Задаем количество **HTML файлов**.

Данный сервис сохраняет сайты у себя для продажи архивов. Поэтому в фильтре можно выбрать, сколько файлов в архиве с сайтом. Этот параметр позволяет задать ограничение на количество html файлов, т.е. по сути сколько на сайте было веб- страниц, без лишних файлов (изображения, файлы css, js и т.д.). Задавать количество страниц имеет смысл для того, чтобы отсеять огромные сайты, на тысячи и миллионы страниц, т.к. с вероятностью 99% это будут форумы, интернет-магазины, киносайты, дорвеи, что угодно, но не статейники. Если страниц слишком мало, то там вообще может не быть контента (обычно это сайты визитки компаний). Я ставлю в этом параметре от 50 до 999. Можно экспериментировать.

4. Язык – **Русский**.
5. **Не показывать с АГС**
6. **Без склеенных**

7. SimilarWeb – **от 1000**. Сколько было трафа на сайте при жизни (количество показов за месяц, т.е. от ~30 HE уников в день). Задаю, чтобы отсеять откровенный шлак. Если параметр не задавать, потенциальных сайтов будет больше, но придется их выискивать среди огромного количества шлака.

8. % поискового – **от 50**. Задаем, чтобы с поиска трафа было хотя бы от половины.

## **Анализ сайтов**

После применения фильтра появится таблица с доменами. Если навести курсор на домен в таблице – появится тайтл его главной страницы. Таким образом можно понять, подойдет домен или нет. Если домен заинтересовал, то уже можно открыть его описание. Там обычно скриншоты нескольких страниц с сайта, можно опять же визуально посмотреть, есть ли на сайте текстовый контент.

Поскольку домены, которые мы смотрим, уже не работают, то открыв их в браузере, увидим ошибку. В качестве альтернативы можно открывать сайты в веб архиве (ссылка в той же папке, где все).

В веб архиве открываем последнюю дату. Если видим заглушку регистратора домена, то смотрим более ранние даты, пока не будет видно сам сайт. Дальше необходимо просто попереходить по ссылкам, убедиться, что на сайте есть статьи, и большая часть страниц с контентом доступна.

После этого нужно статьи выборочно проверить на уникальность. Бывает и такое, что сайта вообще нет в веб архиве, либо отсутствует большинство нужных страниц (обычно сохраняются главная и категории, а внутри с нужными текстами не доступны). В этом случае можно попробовать делегировать сайт на старые DNS.

Для этого нужно узнать, на каком IP находился сайт, когда он еще работал. После чего локально делегировать домен на этот IP адрес. В случае, если сайт еще не удален с хостинга – он будет открываться с вашего компа, как обычно. Что нужно сделать:

1. Переходим на сайт <http://viewdns.info/iphistory/>. В форму вбиваем домен. Находим строку с записью, где указан какой-нибудь хостинг. Обычно он соответствует дате примерно за месяц до самой верхней записи в таблице.
2. Открываем в винде файл C:\Windows\System32\drivers\etc\hosts любым текстовым редактором.
3. В файле hosts пишем: ip domain.com
4. Чистим кеш в браузере, и открываем повторно домен.

5. Дальше возможно несколько вариантов. Либо сразу откроется сайт – тогда все ок. Либо будет редирект на www версию домена. В этом случае в hosts дописываем еще строку: ip www.domain.com. Либо будет висеть заглушка хостинга – значит сайт, увы, уже удален с хостинга, и сделать ничего не получится. Либо ничего не изменится и будет та же страница в браузере. Тогда нужно почистить кеш, если не помогло, то попробовать другой IP адрес. IP можно поменять, перезагрузив Wi-Fi роутер.

## Получение статей

Само собой, статьи можно руками скопировать, открыв сайт. Но в случае, если сайт доступен только в вебархиве, это займет слишком много времени – в вебархиве страницы очень долго грузятся, а также часть страниц не работает.

Поэтому проще выкачать сайт с вебархива любым софтом, и уже вытащить тексты с папки сайта на компе. Есть несколько платных сервисов по выкачиванию сайтов с вебархива, например, [r-tools.org](http://r-tools.org). Можно выкачивать через них. Но если на сайте много страниц, то выйдет дороговато.

К тому же, [r-tools](http://r-tools.org) у меня через раз работает. Проще и дешевле самому выкачивать вот этим скриптом:

<https://github.com/hartator/wayback-machine-downloader>.

Естественно он будет работать только на своем сервере (VDS или дедик), если его нет, то можно попробовать запускать с локального сервера (опенсервер, денвер и т.д.). Скрипт работает на ruby, нужно поставить его, если не стоит. Установка зависит от операционной системы на сервере. Нужно просто погуглить, к примеру, «install ruby centos», «install ruby debian», ну или если локальный сервер, погуглить как ставить ruby на него.

После чего устанавливаем из консоли скрипт. По ссылке выше, есть подробная инструкция на английском, если прокрутить страницу вниз. Запускать в консоли так: `wayback_machine_downloader` <http://example.com>

По умолчанию, он сохранит страницы за самую последнюю доступную дату. Если в вебархиве за последнюю дату висит заглушка хостинга вместо сайта, то нужно найти дату, когда сайт еще работал.

После этого запускать скрипт с дополнительным параметром:

`wayback_machine_downloader http://example.com --to 20100916231334` Число – это дата, которую можно скопировать из урла страницы в вебархиве. В таком виде скрипт будет выкачивать страницы с этой датой и ранее.

Информация выше актуальна для вебархива. Объективных причин для выкачивания статей каким-то специальным софтом с сайтов, делигированных на ваши днс, не так уж и много. Статьи можно просто скопировать руками.

Единственное, бывает, что весь контент сразу не нужен, а сохранять все статьи долго и лениво, и хочется оставить это все на потом. Но сайт-то уже не работает, и может быть удален с хостинга в любой момент. Тогда будет логично выкачать сайт на локалку, чтобы использовать тексты с него в дальнейшем. Есть много качалок сайтов. Отличаются только количеством мусора, правильностью воссоздаваемой структуры, полнотой сохранения файлов стилей, скриптов, картинок.

По большому счету, нам нужны только html страницы, а остальное все не важно. Поэтому подойдет любая программа, которая вам нравится. Я использую Offline Explorer.

Offline Explorer – удобная программа под винду. Основное преимущество перед wget – визуальные настройки. Можно отключить выкачивание любых ненужных файлов, чтобы сохранялись только html страницы. Программа платная, но на просторах сети есть взломанная Portable версия. Из выкачанных html файлов удаляем лишний код, оставляя только тексты и пересохраняем в формат txt.

**Это все, конечно, не просто, поэтому советуем нанять копирайтера)**

## **6. Переводы через Google Translate**

Условно-бесплатный способ получения контента на любую тематику в любое время. Мало кто знает, но уже около года (с весны 2017) гугл обновил и сильно усовершенствовал алгоритмы машинного перевода.

Это применимо только к переводам на отдельном сервисе <https://translate.google.com>. Перевод через плагин в хrome будет по старым алгоритмам, и сильно отличается по качеству в худшую сторону.

Подготовка контента для гугл перевода ничем не отличается от описанной в предыдущем разделе. Находим подходящий буржуйский сайт (я использую сайты на английском – много сайтов, адекватное качество перевода на русский с этого языка), после чего выкачиваем статьи одним из перечисленных выше способов (с помощью качалки и подготовки файлов вручную, либо шаблоном для Content Downloader'a).

Надо отметить, качество контента после перевода все равно далеко от идеального. Во избежание проблем с блокировками аккаунтов в Дзене, статьи

нужно править. Обычно все правки сводятся к форматированию текста после переводчика (разбить на абзацы, поправить теги), поправить по смыслу корявые фразы. В среднем на 1 текст уходит 5-10 минут, в зависимости от объема. Это гораздо проще и быстрее, чем писать с нуля. Качество остается прежним (но зависит от изначального источника).

Качество перевода также сильно зависит от тематики и стиля изложения. Нормально выходит переводить обычные информационные статьи, различные эмоциональные рассказы с шутками и фразеологизмами – с этим у гугла проблемы, получается бред. Из того что пробовал – нормально получается женская тематика и медицина.

Но вносить правки в тексты все равно нудно. Выход – нанять редактора на любом сайте фриланса. По оплате в районе 20-30р за статью (объемом 1500 - 2000 знаков), мотивируя низкую цену количеством контента, которое будете заказывать.

Помимо всего этого ты должен знать, что Дзен любит экспертное мнение автора больше, чем новости, проверено. Делай выводы.