

Как переводить тексты при помощи MateCat

Маркиев Владимир

Содержание:

Введение	1
Что такое MateCat?	1
Почему машинный перевод?	1
Почему гуглоперевод?	1
1. Настройка Google Cloud Platform	3
1.1. Регистрация в Google Cloud Platform	3
1.2. Создание проекта в Google Cloud Console	4
1.3. Создание учётных данных	5
1.3.1. Закрались сомнения?	5
1.4. Добавление учётных данных	5
2. Настройка MateCat	7
2.1. Начало работы с MateCat	7
2.2. Подключение машинного перевода к MateCat.	7
3. Перевод текста при помощи MateCat	9
3.1. Настройка проекта	9
3.2. Перевод документа в MateCat	9
Заключение	11
Немного об альтернативах	11

Введение

Если содержание уже пугает, то советую прочитать краткую информацию вот в [этом пункте](#).

Что такое MateCat?

MateCat - это бесплатная утилита для переводов. Основные преимущества:

- бесплатность,
- скорость работы,
- простота использования,
- богатый набор функций, включая интеграцию с машинным переводом и работу с гуглодиск

Подкупает именно возможность бесплатно работать одновременно с гуглодиск, машинным переводом. При этом настроить всю шушуеру не составит труда.

Почему машинный перевод?

Я предлагаю для перевода в MateCat использовать машинный перевод, потому что он облегчает жизнь.

Но машинный перевод делает очень много ошибок, он совсем плохой и смешной. Любой сможет отличить машинный перевод от человеческого.

На самом деле - это миф. Да, гугл переводчик может выдать ошибки, но это устаревшая информация и ошибки несерьёзные.

Я пользуюсь машинным переводом почти всегда, когда мне нужно перевести хоть сколько объёмный текст, т.к. это упрощает и ускоряет процесс. После современного машинного перевода даже мне остаётся только немного отредактировать, а я всё-таки являюсь некоторым специалистом в этой области. Так что для бытовых целей машинный перевод от гугла подойдёт отлично.

Почему гуглоперевод?

Почему я вообще предлагаю использовать машинный перевод, я объясняю в [другом пункте](#). Почему именно от гугла? Потому что его проще настроить. Для этого понадобится получить один ключ, потом добавить его в настройки MateCat. Это совсем не страшно. Есть большое количество систем машинного перевода, но с гуглом я знаком лучше и новичков он должен пугать меньше, чем майкрософт. Я не использую яндекс, потому что яндекс мне не нравится.

У меня есть уже настроенный ключ, которым я мог бы поделиться, но всегда лучше использовать отдельные аккаунты с собственными настройками на случай, если кто-то

что-то решит изменить.

Chapter 1. Настройка Google Cloud Platform

1.1. Регистрация в Google Cloud Platform

Машинный перевод настраивается через Google Cloud Platform, для которой нужен гугл аккаунт. Если гуглоаккаунта ещё нет или не хочется использовать основной, можно зарегистрировать ещё один, инструкции доступны в интернете.

1. Первым делом необходимо перейти в [Google Cloud Console](#).



Чтобы начать работать с Google Cloud Platform нужно перейти в Google Cloud Console - это немного сбивает с толку, но так вот гугл придумал.

Гугл предложит ознакомиться с условиями предоставления услуг. Можно ознакомиться при желании, а можно просто принять условия. Соглашаться на рассылку необязательно.

2. После ознакомления нажать **Agree And Continue**.

Наверху появится уведомление с предложением активировать подарок. Нужно нажать **Activate**.

3. Будет открыта новая страница, где нужно снова ознакомиться с условиями и принять их. Нажать **Принять и продолжить**.
4. Откроется следующая страница **Шаг 2 из 2**.

На этой странице нужно заполнить все поля настоящей информацией.

- a. В разделе **Тип аккаунта** выбрать ничего нельзя, поэтому оставляем бизнес.
- b. В разделе **Наименование и адрес** нужно указать что-нибудь типа **Моя лучшая компания**, в трёх строках адреса можно написать свой домашний адрес.

Где это возможно, заполнять поля важно реальной информацией, чтобы пройти проверку.

- c. Потом нужно заполнить раздел **Основной контакт** своими контактными данными и перейти к разделу оплаты.
- d. В разделе оплаты нужно заполнить данные банковской карты.



Заполнить данные банковской карты придётся в любом случае. Данные должны быть настоящими, отвертеться не получится. Но не стоит пугаться, всё нормально, гугл не будет списывать деньги за перевод в объёмах до 250 миллионов символов в месяц. Это очень большой лимит. Когда я постоянно переводил, у меня не получалось его превысить. К тому же гугл даёт возможность израсходовать 300 долларов просто по доброте душевной.

После успешного завершения регистрации необходимо перейти к следующей части.

1.2. Создание проекта в Google Cloud Console

После регистрации система сама перенаправит на нужную страницу, но если этого не произошло, нужно ещё раз ввести адрес [Google Cloud Console](#). На главной странице нужно создать новый проект.

Чтобы создать проект, необходимо выполнить следующее:

1. На синей полоске в верхней левой области страницы найти **My First Project**, раскрыть меню, в появившемся окошке выбрать **Создать проект**.
2. Ознакомиться с информацией на странице, при желании изменить название проекта. Местоположение проекта менять не требуется. После ознакомления нажать **Создать**.

Снова откроется главная страница Google Cloud Console, но уже для другого проекта.

Чтобы убедиться, в том, что проект правильный, нужно посмотреть на синюю полоску в верхней левой области страницы. Если там указано имя нужного проекта, то всё замечательно. В противном случае нужно выбрать проект с правильным именем.

3. Затем нужно открыть меню, нажав кнопку в левом верхнем углу страницы. В меню выбрать пункт **API и сервисы**.
4. На следующей странице нажать на кнопку + **Включить API и сервисы**.

Откроется библиотека API. Всё нормально.

5. В строке поиска набрать **Cloud Translation API**, если его нет на главной странице. Перейти на страницу API.
6. На странице **Cloud Translation API** в разделе **Стоимость** выбрать модель оплаты.



Из всех моделей оплаты нужна та, которая называется **AutoML Translation Online Predictions**. Именно она будет бесплатной, если не переводить больше 250 миллионов символов.

7. После выбора модели оплаты, нажать синюю кнопку **Включить**.

Система скажет, что для использования этого API нужно создать учётные данные (Credentials).

Если система не предлагает создать учётные данные, а предлагает создать платёжный аккаунт (Billing account), его просто нужно выбрать из раскрывающегося списка. Он будет под названием **My Billing Account**.

1.3. Создание учётных данных

Если на экране появилась страница с запросом создать учётные данные, нужно их создать. Для этого нужно нажать кнопку **Создать (Create Credentials)** в правом верхнем углу страницы.

Если закрались сомнения, рекомендую прочитать пункт [Закрались сомнения](#). Если пока сомнений нет, можно сразу перейти [дальше](#). Если совсем ничего не понятно, рекомендую начать [сначала](#).

1.3.1. Закрались сомнения?

Где-то на этом моменте могут появиться сомнения. На всякий случай разберу несколько логичных вопросов в формате вопрос/ответ:

В: Зачем мне всё это?

О: Чтобы настроить машинный перевод и не тратить лишние силы и время.

В: Почему так сложно?

О: Да, эта часть может показаться сложно, но если следовать инструкции проблем не возникнет. В следующий раз будет проще. Ну и ещё это полезное знание.

В: А что если с меня спишут все деньги?

О: Не спишут. Нужно понимать, что пока мы ничего не делаем, мы ничего не платим. Создание ключа - бесплатное. Можно выбрать даже самый дорогой тарифный план, но пока что мы ни за что не платим.

1.4. Добавление учётных данных

После нажатия кнопки **Создать** откроется страница для добавления учётных данных. Здесь сложностей быть не должно.

Нужно выполнить следующие действия:

1. Ответить на вопрос: **"Какой API вы используете?"**

Очевидно, мы используем Cloud Translation API.

2. Ответить на вопрос (*может не появиться, это нормально*): **"Планируется ли использовать этот API для App Engine или Compute Engine?"**.

Выбрать "Нет". Все эти Engines для бытового применения не нужны.

3. Ответить на вопрос: **"С какими данными вы будете работать?"**

Ответить "**С данными приложения**".

4. После этого может быть предложено ответить на вопрос из [пункта 2](#).

Ответить отрицательно.

5. Нажать на кнопку **Выбрать тип учётных данных**. Кнопка может также называться **Далее**.

Развернётся следующий пункт — **Создание сервисного аккаунта**.

6. Дальнейшие действия могут различаться:

- Раньше нужно было выбрать роль **Проект** → **Владелец** из раскрывающегося списка, потом выбрать тип ключа и т.д.
- Сейчас нужно ввести имя сервисного аккаунта, его описание и нажать **Создать**.

7. А следующие шаги вообще опциональны и их можно пропустить:

- Дать сервисному аккаунту права доступа к проекту.
- Дать пользователям доступ к этому сервисному аккаунту.

8. Нажать кнопку **Готово**.

Откроется страница учётных данных.

9. На странице учётных данных нужно нажать кнопку + **Создать (Create Credentials)**.

10. В появившемся меню выбрать **Ключ API**.

Откроется окно с готовым ключом API. Можно ограничить доступ к ключу, чтобы его никто не похитил и не начал использовать для перевода многих сотен миллионов символов. Но проще просто не передавать ключ никому. Кто-то просит ключ? Пусть создаст себе свой. Это несложно.

Итак, ключ создан. Не нужно записывать этот ключ на бумажку или сохранять в файл. Ключ всегда доступен для копирования из Google Cloud Console. Чтобы копировать ключ, нужно перейти на главную страницу Google Cloud Console, затем навести мышку на APIs & Sercvices, а потом выбрать Credentials. Все ключики будут показаны на экране. Теперь его можно использовать, но пока что негде. Значит пора переходить к настройке MateCat.

Chapter 2. Настройка MateCat

2.1. Начало работы с MateCat

Настроить MateCat проще, чем Google Cloud Platform.

Для этого нужно следующее:

1. Перейти на главную страницу MateCat - [MateCat.com](https://matecat.com).
2. В правом верхнем углу страницы нажать на иконку человечка.

Откроется окошко входа.

На регистрацию можно забить, ведь у нас есть гуглоаккаунт!

3. Выбрать Sign in with Google.
4. Затем выбрать свой гуглоаккаунт и разрешить доступ к аккаунту в новом окне, нажав **Разрешить** или **Allow**.
5. MateCat предложит сгенерировать ключ API для аккаунта, но нам это не нужно. А вот доступ к гуглодиску пригодится, поэтому ставим переключатель под Google Drive в положение **ON**.
6. Снова откроется окно для выбора аккаунта. Это нормально, выбираем аккаунт, с диска которого нужно будет переводить файлы.
7. Снова нажимаем **Разрешить** или **Allow**.

Окно окажется пустым. Это нормально. Его нужно закрыть. Теперь у MateCat есть доступ к гуглодиску.

2.2. Подключение машинного перевода к MateCat.

Теперь, когда у нас всё настроено, можно переходить к объединению машинного перевода и системы для перевода.

1. На странице проекта MateCat (главной странице), в строке настроек нужно найти шестерёнку и нажать на неё.

Откроется меню настроек.

2. В меню настроек нужно выбрать вкладку **Machine Translation**.
 - Затем нажать на кнопку + **Add a MT Engine**.
3. Вы раскрывающемся списке выбрать провайдера — Google Translate.
4. В поле **Engine name** вписать имя машинного перевода. Можно указать любое, хоть "Единороги скачут по радуге", если потом это имя не вызовет путаницу.
5. В поле **API-key** ввести **ключ API**, который мы сгенерировали при настройке Google Cloud

Platform.

6. Копируем ключ с Google Cloud Console, вставляем в MateCat, нажимаем **Confirm**.

7. Поставить галочку там, где сказано **Use in this project** и закрыть меню.

Ура! Теперь MateCat знает о существовании ключа API Google Machine Translation! Можно приступать переводу.

Chapter 3. Перевод текста при помощи MateCat

3.1. Настройка проекта

Прежде чем переводить текст, потребуется создать проект. Это совсем не сложно.

На главной странице MateCat:

В поле **Project name** нужно указать имя проекта. Имя лучше вводить латинскими буквами, без пробелов и специальных символов, кроме нижнего подчёркивания (_) и дефиса (-). Иначе MateCat может работать некорректно.

Список **From** содержит языки, с которых можно перевести документ. Если перевод с английского, нужно выбрать любой English доступный в списке.

Аналогично поступить со списком **To**.

В списке **Select subject** нужно выбрать тему перевода. Это не обязательно, но рекомендуется.

В списке **TM & Glossary** можно ничего не выбирать, тогда по умолчанию будет использоваться общая для всего MateCat память переводов^[1]. Можно создать свою собственную память переводов, нажав на кнопку **Create resource** → **New resource**. А можно импортировать уже существующую, нажав **Import TMX**.^[2]

3.2. Перевод документа в MateCat

После того как все поля проекта заполнены, можно начинать работу с файлами из гуглодиска.



Лучше отключить AdBlock на сайте MateCat. Иначе сайт может работать некорректно.

Для этого:

1. Под надписью **"Drop your files to translate them with MateCat"** нужно найти фразу **"MateCat supports 78 file formats and Google Drive files"** и нажать на ссылку **Google Drive files**.

Откроется окошко гуглодиска.

2. Нужно выбрать файл(ы) из гуглодиска, затем нажать **Select**.



Поддерживаются только файлы в формате гуглодиска. Это значит, что любой формат, который не был изначально создан в гуглодиске, не получится добавить на перевод. Обходное решение — скопировать всё содержимое документа в любом формате и вставить его в новый документ на гуглодиске. Структура документа от этого не пострадает.

3. После добавления документов нужно нажать кнопку **Analyze**.

В новой вкладке откроется страница анализа проекта, нужно нажать **Translate**.

Благодаря подключенному машинному переводу MateCat автоматически подставит перевод в каждый сегмент. Останется только посмотреть каждый сегмент и, при желании, отредактировать неточности.

Если MateCat не подставил перевод, нужно пройти по каждому сегменту и выбрать перевод вручную.

[1] Что такое память переводов, можно посмотреть в интернете.

[2] Вряд ли такие детали понадобятся при обычном бытовом переводе, но знать полезно.

Заключение

После перевода на гуглодиске в той же папке появятся ещё по два файла для каждого исходного. Один файл — исходный документ, с которым будет работать MateCat, чтобы не испортить оригинал. Второй файл — переведённый исходник на языке перевода.

Если нужно получить доступ к предыдущим переводам в MateCat, их можно найти, нажав на аватар в верхнем правом углу страницы. В появившемся меню нужно выбрать пункт **My Projects**.

Немного об альтернативах

Файлы можно добавлять на перевод и с компьютера, просто перетаскив их на страницу проекта, но это не так мобильно. Например, если нужно скачать перевод на другой компьютер, то придётся его как-то переносить руками или через тот же гуглодиск. А так перевод можно открыть даже с телефона.

В MateCat есть встроенный машинный перевод. Можно использовать его, если хочется, и не настраивать дополнительно никакой другой. Но как же дух энтузиазма?!