

Входные параметры для проекта Субтитры Ютуб v1.7.xmlz

Прокрутка вниз 20

Использовать свои ссылки ☐ Да ☒ Нет

Записывать ссылку на ролик в текст ☐ Да ☒ Нет

Цикл ожидания субтитров 3

Количество символов 300

Выбор языка парсинга текста Русский

Длина ролика От 3 До 10

1. **Прокрутка вниз** – прокручивает заданное количество роликов по ключевому слову и забирает ссылки на них
2. **Использовать свои ссылки** (по умолчанию отключено) формата - <https://www.youtube.com/watch?v=UQSh4oUyvBU>. Ссылки на ролики должны лежать в файле Y_link.txt
3. **Записывать ссылку на ролик в текст** - записывает ссылку вверху текста. По умолчанию стоит «Нет»
4. **Цикл ожидания субтитров** - служебный. Можно менять. Отвечает за проверку доступности субтитров.
5. **Количество символов** - Минимальное количество текста, которое должно быть в статье. Если менее заданного значения, текст игнорируется и берется следующий ролик.
6. **Выбор языка парсинга текста** - Выбор языка, на которых должен быть текст. Рекомендуется использование расширенных ключевых слов для более точного парсинга. Например – (неправильно) – «женщины». Правильно – «женская психология», «психология в «России».
7. **Длина ролика** в минутах. Рассчитывается по примерной формуле $x/60/2$. Всё, что находится в диапазоне менее 3 минут и более 10 минут, отсеивается. Берется только диапазон от 3 до 10 минут.

| | | | | |
|--------------------------|------|---------|------------------|-------------------------------|
| 📁 [..] | | | | <Папка> 23.02.2020 21:30 ---- |
| 📁 [Responsibility] | | | | <Папка> 23.02.2020 21:29 ---- |
| 📁 [Systems] | | | | <Папка> 23.02.2020 21:29 ---- |
| 📁 [Тексты] | | | | <Папка> 23.02.2020 21:29 ---- |
| 📄 Парсинг субтитров ютуб | pdf | 460 360 | 21.02.2020 17:08 | -a-- |
| 📄 key | txt | 0 | 23.02.2020 17:53 | -a-- |
| 📄 Y_link | txt | 0 | 23.02.2020 13:08 | -a-- |
| 🔗 Субтитры Ютуб v1.7 | xmlz | 225 555 | 23.02.2020 13:11 | -a-- |

Обновленная логика. Все служебные файлы убраны для избежания недоразумений

| | | | | |
|-----------------------|--|--|--|------------------------------|
| 📁 [..] | | | | <Папка> 23.02.2020 21:29 --- |
| 📁 [глупые истории] | | | | <Папка> 23.02.2020 21:29 --- |
| 📁 [дурацкий праздник] | | | | <Папка> 23.02.2020 21:29 --- |
| 📁 [женские глупости] | | | | <Папка> 23.02.2020 21:29 --- |
| 📁 [мужские глупости] | | | | <Папка> 23.02.2020 21:29 --- |
| 📁 [неудачник] | | | | <Папка> 23.02.2020 21:29 --- |
| 📁 [обзор перфоратора] | | | | <Папка> 23.02.2020 21:29 --- |

В папке «**Тексты**» лежат тексты, разложенные по ключевому слову.

| | | | | |
|----------------|-----|----|------------------|------------------------------|
| 📁 [..] | | | | <Папка> 23.02.2020 21:29 --- |
| 📄 key_tmp | txt | 33 | 23.02.2020 17:53 | -a- |
| 📄 link_youtube | txt | 0 | 23.02.2020 18:28 | -a- |

В папке «**Systems**» находятся два файла. В «**key_tmp.txt**» находится ключ, по которому парсятся тексты, и он же создает папку по названию ключевого слова.

Здесь есть тонкость – если вы хотите парсить по своим ссылкам, вы должны ввести название в файле «**key_tmp.txt**» (например, мои ролики), которое будет использоваться как директория, куда будут складываться тексты. Пример – вы кладете в файл «**Y_link.txt**» ссылки формата <https://www.youtube.com/watch?v=UQSh4oUyvBU>, во входящих настройках должно быть включен пункт «**Использовать свои ссылки**». В других случаях ничего трогать не нужно.

В файле «**link_youtube.txt**» находятся файлы, которые подходят по параметрам (от 3 до 10 минут), именно они будут проверяться. Как только они закончатся, возьмётся следующее ключевое слово на проверку. Здесь тоже ничего трогать не нужно.

| | | | | |
|-------------|-----|--------|------------------|--------------------------|
| 📁 [..] | | | | <Папка> 23.02.2020 21:29 |
| 📄 BlackList | txt | 19 395 | 23.02.2020 18:28 | |

Черный список - Здесь находятся ссылки на ролики, которые уже парсились

Файл «**key.txt**» - сюда складываете ключевые слова. Пример:

Психология
Женская психология
Основы психологии

Видео: <https://youtu.be/-T6bBO6zAY> По вопросам – footashes@gmail.com