



2КиноПоиск

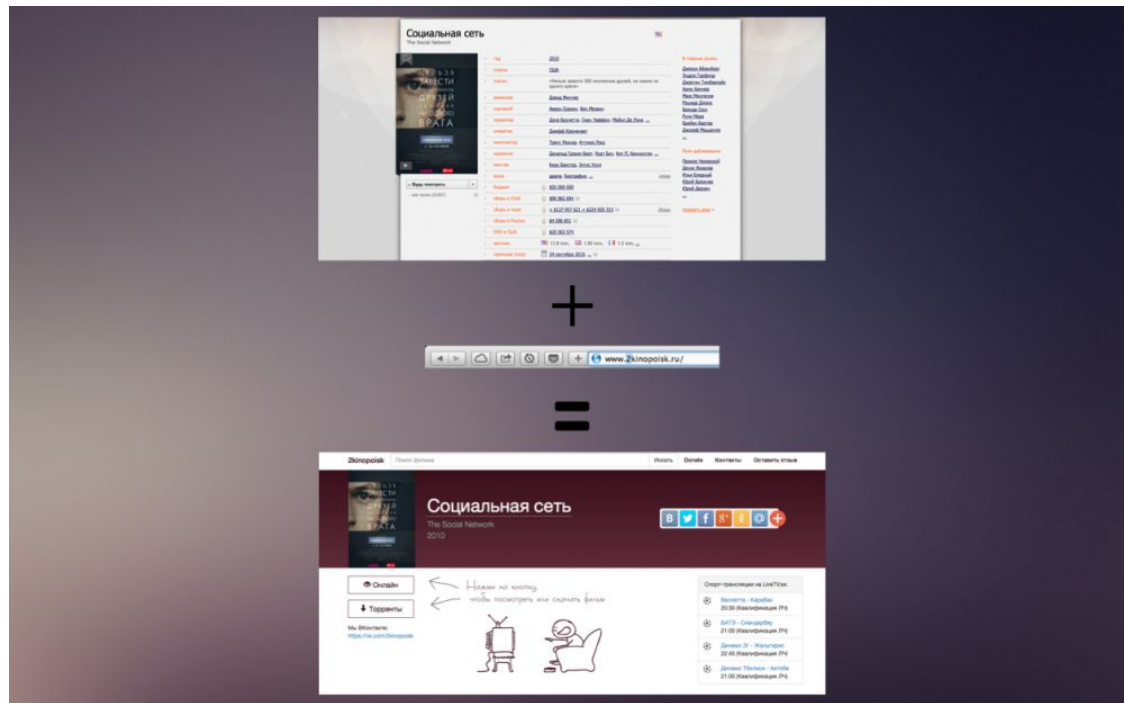
Проект и технологии



План

1. Идея. Запуск. Хайп
2. Архитектура. Технологии
3. Подробнее про компоненты
4. Заметки

Идея



Запуск. Хайп

4681

LifeHack [мод]

Пользователи кинопоиска, просто добавляете "2" перед ссылкой на фильм - и можете посмотреть его онлайн в HD или скачать с торрент-трекеров.



465 Комментариев



HappyReindeer 1319 дней назад

LifeHack

Лайфхак

кинопоиск

Фильмы

HD

скачать фильм

торрент

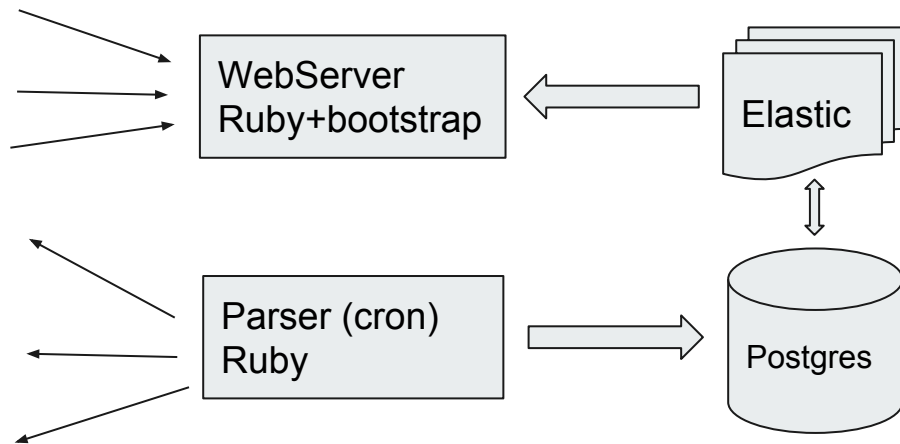


▲▼ свернуть

250+ ТЫСЯЧ

Уникальных пользователей в месяц

Архитектура. Технологии



WebServer

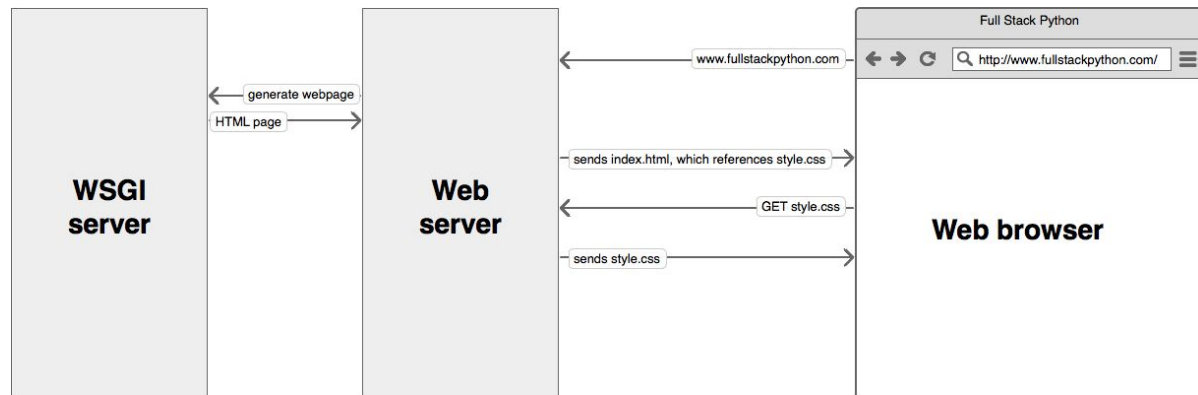
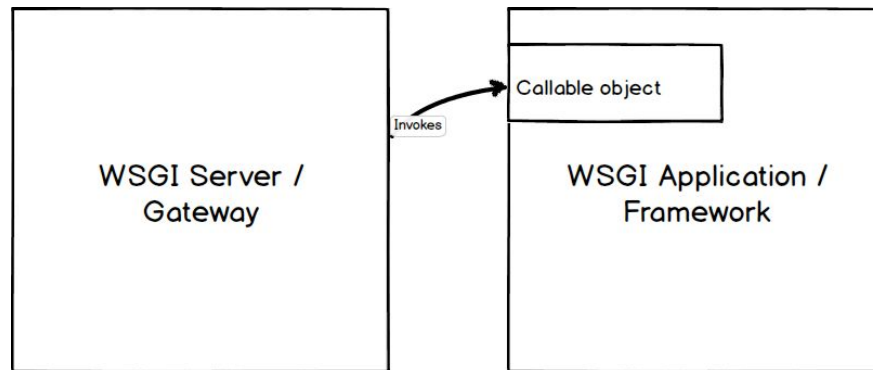
```
from flask import Flask
app = Flask(__name__)
```

```
@app.route("/")
def hello():
    return "Hello World!"
```

```
if __name__ == "__main__":
    application.run(host='0.0.0.0')
```

```
---
```

```
python app.py
```

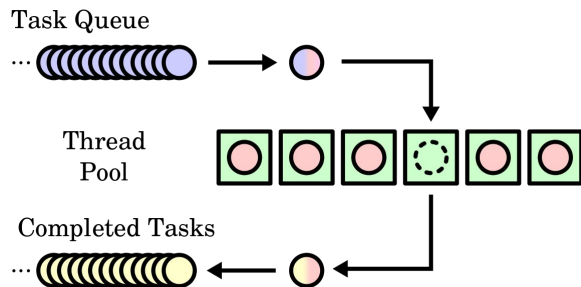




Parser. Сложности

1. Нужно парсить: кинопоиск, сайты с онлайн-просмотром, трекеры
2. Все они защищаются от большого количества запросов - прокси
3. Retry политика + надёжные прокси + фильтрация их
4. Обобщённый парсер для каждого ресурса

Parser. Cxema



```
class TvcoParser
  def initialize(parser_config, continue_lambda, end_lambda)
    @continue_lambda = continue_lambda
    @end_lambda = end_lambda
  end

  def valid_result(res)
    res.body.match(/class='top-logo'/) || res.body.match(/movie_name/)
  end

  def start
    ["animation", "video"].each do |name|
      request = Typhoeus::Request.new("http://www.tvco.ru/#{name}", @options)
      request.original_options = {:category => name}
      @continue_lambda.call(request, :process_category)
    end
  end

  def process_player(res)
    # Regexes
    @end_lambda.call( Player.new(...) )
  end
end
```



Parser. Общий

<https://docs.google.com/document/d/1eeoNtpBZJmeTKi-Q0xjp4JMIdaekyIevonRRHezz8S4/edit?usp=sharing>



Заметки

1. 2к строк могут решать крутую задачу
2. 33к запросов в первый день на “ruby app.rb” на сервере за 10\$
3. Пока вы можете работать в рамках одной машины не надо городить что-то распределённое. Вряд ли вам потребуется работать больше, чем на одной машине
4. Скриптовых языков достаточно почти для всего веба. Уберите ваши golang’и.



Thanks!

Contact me:
@makrusak (tg, vk, fb)
maxim@sourcerer.io

