

Тема доклада:
"Поиск объекта по фото в интернет-магазинах через
Telegram"

Выполнил:

Студент 3 курса факультета математики и компьютерных наук
Группа ЗПМ, Никитенко Владислав

Научный руководитель:

Кандидат физико-математических наук, Лобанов Андрей Григорьевич

Майкоп
2022г.

ВВЕДЕНИЕ

Курсовая работа посвящена созданию Telegram-бота, который по полученной фотографии от пользователя определит объект и вернёт ссылки на наиболее похожий вариант в интернет-магазинах.

ВВЕДЕНИЕ

Задачи:

- Разработать алгоритмы определения объекта по фото и поиска его в интернет - магазинах;
- Получить информацию с сайтов;
- Очистить, проанализировать и отсортировать данные;
- Создать Telegram - бота и интегрировать в него написанные алгоритмы.

ОПРЕДЕЛЕНИЕ ОБЪЕКТА

Изображение загружается на сервис "Яндекс. Картинки", после чего получаем ссылку на результаты поиска.

Далее все данные с сайта парсятся библиотекой Beautiful Soup 4.

Пример поиска

ОПРЕДЕЛЕНИЕ ОБЪЕКТА

Яндекс Будьте в Плюсе

Поиск **Картинки** Видео Карты Товары Новости Переводчик Все



Размер изображения: 1280x720

Кажется, на изображении

мегаминкс некубический мегаминкс
кубик рубика мегаминкс мегаминкс кубик сборки
головоломка мегаминкс megaminx qiheng

Товары на изображении

Всё изображение

 1 510 Р Shengshou Megaminx... cccstore.ru	 640 Р QIYI MoFangGe Megaminx... cccstore.ru	 10 944 Р ShengShou Teraminx... cccstore.ru	 Товаров гораздо больше <input type="button" value="Смотреть все"/>
---	---	--	--

Похожие изображения



Другие размеры изображения

Таких же изображений не найдено

Сайты с информацией про изображение

-  **Стоит ли покупать Головоломка Megaminx QiHeng / Кубик Рубика мегаминкс / Логическая игра, Кубик.? Отзывы на...**
market.yandex.ru
Головоломка Megaminx QiHeng / Кубик Рубика мегаминкс / Логическая игра, Кубик.: важная информация о товаре на Яндекс.Маркете.
-  **Характеристики модели Головоломка Megaminx QiHeng / Кубик Рубика мегаминкс / Логическая игра, Кубик. на...**
market.yandex.ru
Подробные характеристики модели Головоломка Megaminx QiHeng / Кубик Рубика мегаминкс / Логическая игра, Кубик. — с описанием всех особенностей.

ПОИСК ОБЪЕКТА

Получим данные из каталога с помощью библиотеки Selenium.

Если название удалось определить точно, также будем искать объект по названию через поисковик Яндекс.

ПОИСК ОБЪЕКТА



1 200 ₺
LanLan Master skewb
cccstore.ru



БЕЗ КОРОБКИ
895 ₺
MoFangGe 3x3
Thunderclap V3...
cubemarket.ru



550 ₺
LanLan Skewb
cubemarket.ru



1 060 ₺
Кубик Рубика Cyclone
Boys 3x3x3 Feiku...
cccstore.ru



от 650 ₺
Мегаминкс Shengshou
(грань) черный...
Яндекс.Маркет



1 250 ₺
Qiyi MoFangGe Clover
cube
gamecuber.ru



1 890 ₺
Купить megaminx
"Galaxy V2 Magnetic"...
gamestil.ru



2 969 ₺
2020 Moyu AoHun
WRM 3x3x3 Megamin...
m.aliexpress.ru



583 ₺
Оригинальный
скоростной куб QIYI...
altaieko.ru



10 \$
Qiyi кубики 3x3 кубик
рубика головоломка...
momoyo.ru



1 088 ₺
MoYu 3x3x3 AoLong v2
cccstore.ru



382 ₺
3x3x3 наклейка из
углеродного волокн...
aliexpress.ru



440 ₺
Головоломка
многогранник...
bayplus.ru



3 240 ₺
WitEden Super
3x3x3:00 купить...
cccstore.ru



2 415 ₺
Оригинальный
ShengShou...
m.aliexpress.ru



5 \$
COLOURFUL tTETRIS
DE STRESS FUN...
tvc-mail.com



2 174 ₺
Образовательная
головоломка Lanlan...
aliexpress.ru



1 830 ₺
Обзор - Для
мозговитых 1
cccstore.ru

Получаем HTML код каталога

ПОИСК ОБЪЕКТА

```
21 # Создает HTML файл элементов каталога, возвращает soup переменную
22 def get_catalog(link, html_path):
23     options = Options()
24     options.headless = True
25     driver = webdriver.Firefox(options=options)
26
27     try:
28         driver = webdriver.Firefox(options=options)
29         driver.get(link)
30         driver.maximize_window()
31
32         element = driver.find_element(By.CLASS_NAME, "CbirMarketProducts-MoreButton")
33         element.click()
34         time.sleep(2)
35
36         scroll_height = 1000
37         for i in range(5):
38             driver.execute_script(f"document.querySelector('.CbirMarketProducts-ModalScroller').scrollTop = "
39                                 f"{scroll_height}")
40             time.sleep(0.5)
41             scroll_height += 1000
42
43         time.sleep(2)
44
45         items = driver.find_element(By.CLASS_NAME, "MasonryLayout-Container")
46
47         directory = html_path.rsplit('/', 1)[0]
48
49         if not os.path.exists(directory):
50             os.makedirs(directory)
51
52         with open(html_path, 'w+', encoding="utf-8") as file:
53             file.write(items.get_attribute('innerHTML'))
54
55     except Exception as ex:
56         print(ex)
57     finally:
58         driver.close()
59         driver.quit()
```

РАБОТА С ДАННЫМИ

Все полученные данные очищаются от HTML, отсеиваются и сортируются по схожести с искомым объектом.

РАБОТА С ДАННЫМИ

```
302 # Объединение всех функций по очистке
303 def cleaning_items(items, main_text, flag, ali_flag):
304     # Очистка элементов от html
305     items = get_info_from_html(items)
306
307     # Добавляем элементы из поиска
308     if flag:
309         items = items + get_search_yandex(main_text)
310
311     # Удаляю лишние магазины
312     new_items = checking_item_link(items, ali_flag)
313
314     # Проверяю товары по названиям, если общих слов слишком мало - отбрасываю
315     if flag:
316         # Обработка товаров по точному названию
317         new_items = checking_item_name(new_items, main_text)
318     else:
319         # Обработка товаров по обобщенному результату
320         checking_item_name_generically(new_items, main_text)
321
322     # Находим среднее арифм. цен
323     middle_price = get_average_price(new_items)
324
325     # Удаляю товары с сильно отличающейся ценой
326     new_items = checking_item_price(middle_price, new_items)
327
328     return new_items
```

Интерфейс бота

TELEGRAM БОТ

Привет! Отправь мне фото предмета и я верну тебе ссылки на его покупку.

! Немного полезной информации:

Качество фотографии и отсутствие лишних объектов повышают точность распознавания!

Если на объекте есть важный для идентификации текст (название, фирма и т.п.), постарайтесь сделать его максимально читабельным.

Если объект определен верно, но модель не та, то есть большой шанс, что искомый предмет все равно будет в результатах поиска.

0:48

Вы можете настроить список сайтов, на которых будем искать объект или сразу начать поиск.

Сайты по-умолчанию: Яндекс Маркет, Ozon, Wildberries

0:48



Поиск

Настройки

Меню настроек списка сайтов

НАСТРОЙКИ ПОИСКА

В текущий момент ищем на следующих сайтах:
Яндекс.Маркет, Wildberries, Aliexpress 0:52

Выберите сайт, который хотите добавить / убрать 0:52

Яндекс Маркет

Ozon

Wildberries

Aliexpress

Любые сайты

Отмена



НАСТРОЙКИ ПОИСКА

Изменение списка сайтов

Ozon добавлен в список. 0:55

К главному меню

Показать текущий список

Wildberries удален из списка. 0:55

К главному меню



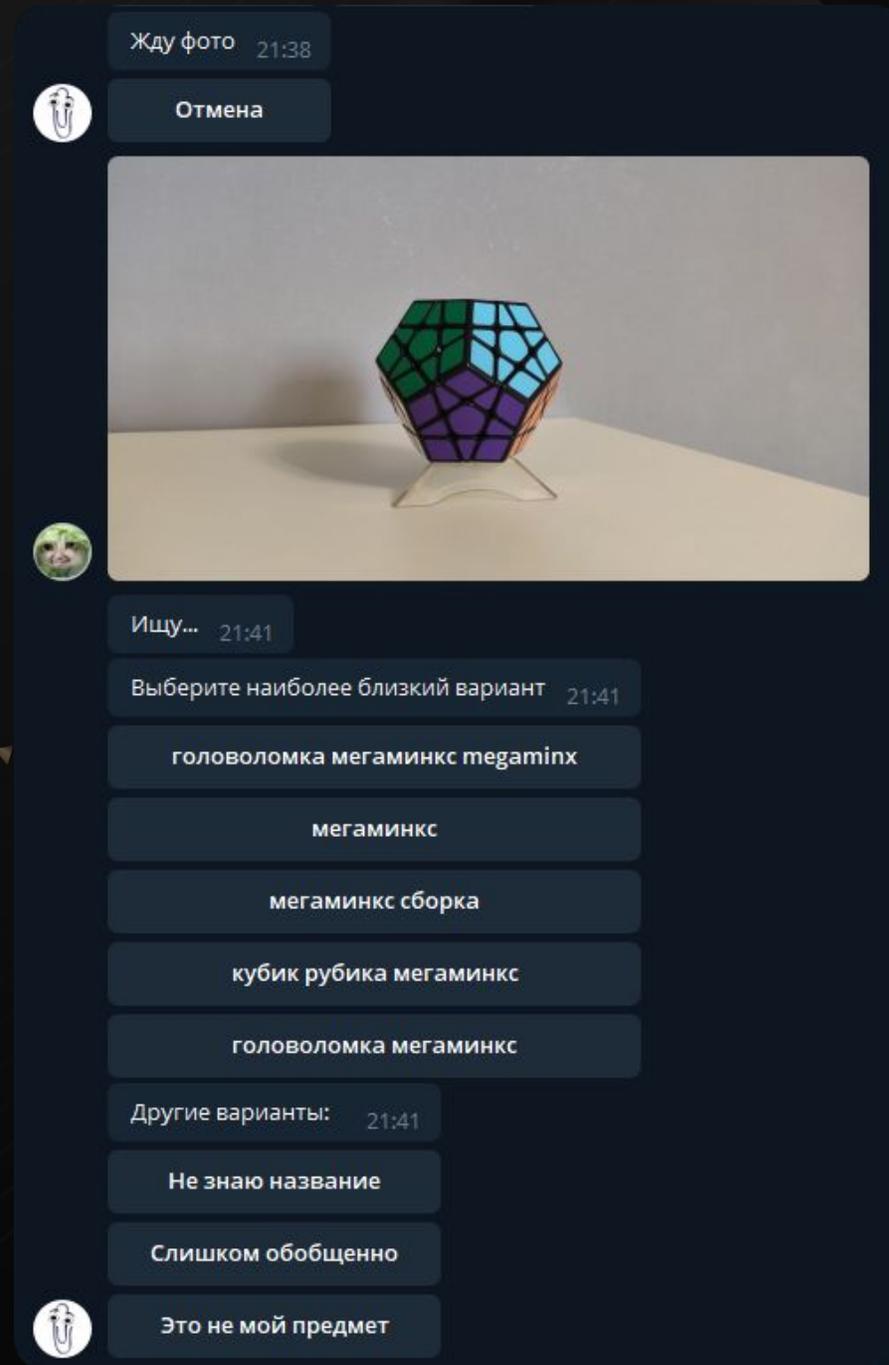
Показать текущий список

База данных

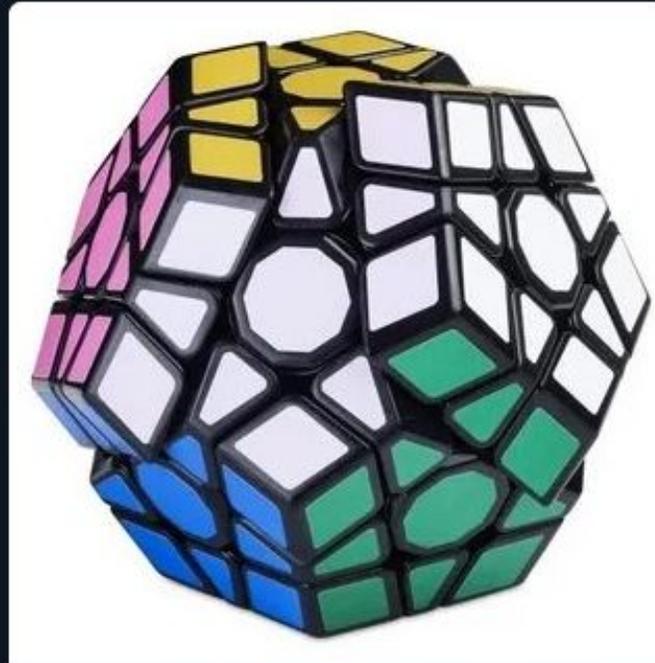
НАСТРОЙКИ ПОИСКА

	id	sites_list
	Фильтр	Фильтр
1	1139342069	Яндекс.Маркет Wildberries
2	1703040346	Wildberries Aliexpress
3	584356423	Яндекс.Маркет Ozon Wildberries

ПРИМЕР ВЫПОЛНЕНИЯ



ПРИМЕР ВЫПОЛНЕНИЯ



Головоломка Meffert's Мегаминкс (M5040)

Цена: 1 850 ₽

Магазин: market.yandex.ru

22:04



Ссылка на магазин



Вы нашли свой предмет? 22:04

Нашел

Показать еще



Выбрать другой вариант

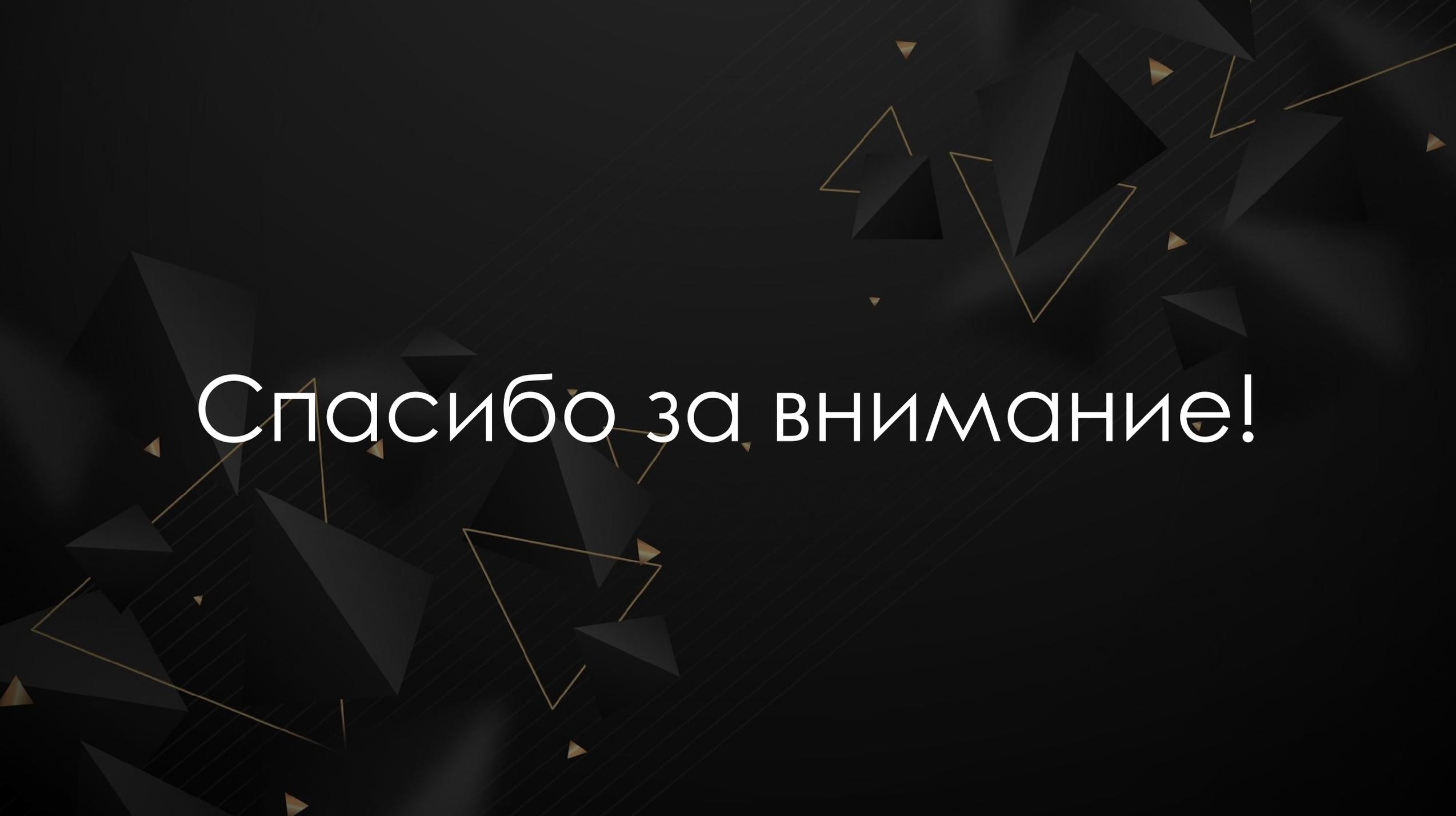
РЕАЛИЗАЦИЯ ПОИСКА

```
132 # Начинаем поиск
133 @dp.callback_query_handler(text='main_menu_search', state=None)
134 async def inline_item_name_call(callback: types.CallbackQuery):...
140
141
142 # Реакция на кнопку отмены
143 @dp.callback_query_handler(text='inline_bnt_cancel', state="*")
144 async def btn_cancel(callback: types.CallbackQuery, state: FSMContext):...
150
151
152 # Получаем фото
153 @dp.message_handler(state=Search.photo, content_types=["photo"])
154 async def photo_analysis(message: types.Message, state: FSMContext):...
196
197
198 # Найден не тот предмет
199 @dp.callback_query_handler(text='item_not_found', state=Search.item_get_name)
200 async def not_found(callback: types.CallbackQuery, state: FSMContext):...
206
207
208 # Возвращает клавиатуру вариантов после первых трех элементов
209 @dp.callback_query_handler(text='inline_another_item', state=Search.item_get_name)
210 async def another_item(callback: types.CallbackQuery, state: FSMContext):...
230
231
232 # Получаю ответ пользователя ( название предмета / другие варианты )
233 @dp.callback_query_handler(lambda x: x.data and x.data.startswith('inline_item '), state=Search.item_get_name)
234 async def inline_item_name_call(callback: types.CallbackQuery, state: FSMContext):...
348
349
350 # Закончить
351 @dp.callback_query_handler(text='inline_btn_end_search', state=Search.item_get_name)
352 async def end_search(callback: types.CallbackQuery, state: FSMContext):...
358
```

ЗАКЛЮЧЕНИЕ

Бот часто успешно определяет и находит искомый объект, что может быть полезно в повседневной жизни.

Развитие программы может быть продолжено улучшением алгоритмов поиска и сортировки, а также добавлением новых функций.



Спасибо за внимание!