

Web Scraper for real estate website

Last update on

Jan 12, 2026

Khaled Alruwita

١) الفكرة

المشروع عبارة عن سكريبر (**Web Scraper**) يسحب بيانات العقارات من موقع عقار بشكل آلي، ويخزنها في دatabse **SQLite**. تقدر تستخدمها لاحقاً في التحليل أو العرض أو أي مشروع آخر.

ال스크ريبر مصمم إنه:

- يشتغل بشكل مستمر
- يتعامل مع الصفحات الديناميكية (يفضل **Playwright**)
- يتعامل مع الانقطاعات
- ما يكرر البيانات

٢) الأدوات المستخدمة:

Playwright (1)

- يشغل متصفح حقيقي (Chromium)
- يفتح الصفحات كأنك مستخدم فعلي
- يتعامل مع JavaScript (تحميل، scroll، أزرار)
- مناسب لمواقع ما تنفع معها requests العادية

BeautifulSoup (2)

- بعد ما الصفحة تتحمل، نستخدمه لتحليل إلـ HTML

• استخراج:

- السعر
- العنوان
- نوع العقار
- المساحة
- الرابط
- التاريخ

SQLite (3)

- قاعدة بيانات خفيفة كوني بستخدمها على جهاز ضعيف مثل Raspberry pi
- ملف واحد (aqr.sqlite)

3) كيف يمشي السكريير ؟

(1) تشغيل السكريير

أول ما تشغّل الملف:

- يفتح المتصفح عن طريق Playwright
- يروح لصفحة العقارات الرئيسية
- يجهّز الاتصال بقاعدة البيانات

(2) تحميل الصفحة

- ينتظر لين الصفحة تكمل (حق العناصر تظاهر)
 - يسوى scroll عشان يحمل إعلانات أكثر
 - يتأكد إن المحتوى صار موجود فعلًا قبل التحليل

قراءة الإعلانات (3)

لكل إعلان عقار:

- يقرأ البيانات من الـ HTML
 - ينظف النصوص (أرقام عربية → إنجليزية)
 - يتتأكد إن القيم مو فاضية

4) منع التكرار

قِبَلٌ مَا يَحْفَظُ الْإِعْلَانُ:

- يفحص هل الرابط موجود مسبقاً في SQLite
 - إذا موجود → يتجاهله
 - إذا حديد → يكمله

5) التخزين في SQLite

يحفظ الإعلان في جدول مثل:

- | | |
|------------|---|
| id | • |
| title | • |
| price | • |
| area | • |
| city | • |
| url | • |
| created_at | • |

كما ، اعلان حديد بنضاف مباشرة للقاعدة.

GIT 4) رابط المشروع في

<https://github.com/KhaledAlruwita/realestate-scraper>