

گزارش پروژه

سیستم جمع‌آوری و تحلیل آگهی‌های مسکن

اعضای تیم:

- محمد صالح شمس‌زاده سينا فاطمی سجاد نوری امیرحسین ساجدی

شرح پروژه:

هدف این پروژه، استخراج آگهی‌های مسکن از دو سایت و انتقال آن‌ها از طریق RabbitMQ به پایگاه داده MySQL بود. این سیستم شامل بخش‌های Web Scraping، تشخیص آگهی‌های مشابه، انتقال داده و ذخیره‌سازی است.

تقسیم وظایف:

- محمد صالح شمس‌زاده: طراحی و پیاده‌سازی Scraper برای سایت "maskan-file"
- سجاد نوری: طراحی الگوریتم تشخیص آگهی‌های مشابه
- سينا فاطمی و امیرحسین ساجدی: توسعه ماژول Scraper و Detector برای سایت melkradar و بخش پایگاه داده

مشکلات توسعه:

دسترسی به API یکی از سایت‌ها دشوار بود که باعث شد از Selenium برای استخراج داده استفاده شود. همچنین پیاده‌سازی و هماهنگی بین ماژول‌ها با RabbitMQ فرآیند توسعه را پیچیده‌تر کرد.

بررسی ساختار و معماری:

سیستم شامل Scraper برای جمع‌آوری داده، ماژول تشخیص مشابهت، ماژول انتقال‌دهنده RabbitMQ و دیتابیس MySQL برای ذخیره اطلاعات است. ارتباط بین اجزای سیستم از طریق صف پیام پیاده‌سازی شد.

پیشنهاد بهبود:

در بخش Scraper سایت maskan-file می‌توان به جای Selenium از متد POST برای استخراج اطلاعات سایت maska-file و سپس کدگذاری استفاده کرد تا کارایی افزایش یابد.

قابلیت توسعه:

سیستم قابلیت توسعه دارد و می‌تواند به کمک یک رابط کاربری فرانت‌اند و بک‌اند کامل، به یک پلتفرم واقعی خرید، فروش و اجاره مسکن تبدیل شود.

کاربرد واقعی:

در دنیای واقعی این سیستم مشابهایی دارد و می تواند در بازار املاک به کار رود. چالش اصلی در این مسیر، اجرای دائمی و به روز نگه داشتن آگهی ها است.

الگوریتم ها و پکیج ها:

- BeautifulSoup و Regex برای پردازش HTML
- Requests برای دریافت صفحات
- Selenium برای استخراج آگهی در شرایط خاص
- Pymysql برای ارتباط با دیتابیس
- Pika برای اتصال به RabbitMQ

این ساختار به تیم اجازه داد تا داده های مورد نیاز را به صورت مؤثر جمع آوری، تحلیل و ذخیره کند.