

# ZiScraper Project

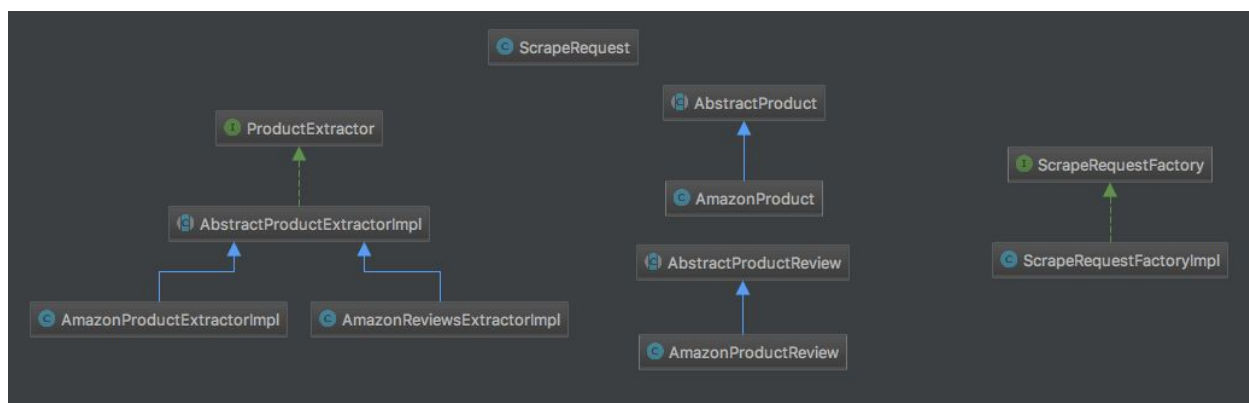
<https://github.com/zivkaziv/ZiScraper>

הסבר על המערכת:

הרעיון במימוש הוא שנוכל להרחיב את הקוד ולהוסיף עוד Extractors בלי לפגוע בקוד הקיים ולכן ישנן שתי מחלקות abstracts בשביל זה:

- **AbstractProductExtractorImpl** - שם יושבת הלוגיקה של איך לחלץ את השדות מה HTML ולמלא את ה Product. בנוסף היא יודעת לקרוא לעצמה להמשיך עבודה כמו במקרה שלנו, את ה Reviews אנחנו מוציאים בכלל מ URL אחר אבל ממלאים את אותו model
- **AbstractProduct** - זהו ה model שאותו אנחנו ממלאים, כל מימוש והמודל שלו.

בקוד הנוכחי ישנו מימוש ל Amazon אבל אפשר מאוד בקלות להרחיב אותו ל Ebay, או לכתוב Extractors אחרים ל amazon (למשל במקרה של A/B Tests בצד של אמאזון)



ה **Main** נמצא תחת המחלקה **ScraperRunner** שם הוא מקבל את ה URL כ Arguments ל main ונוצרות בקשות Scraping שעוברות לטיפול של **Scraper**

כדי להתחיל את התהליך אנחנו משתמשים ב Factory שעפ"י ה URL מייצר את הבקשה ומצמיד לה Product ו Extractor, כך שהוספת אסטרטגיות נוספות תצריך תוספת גם Factory.

במימוש Scraping עצמו ישנו ניסיון להביא חלק מהמידע על המוצר בצורה אסינכרונית. ניתן להביא או לו בצורה זו באתחול של AmazonProduct (כרגע השליטה היא רק ברמת הקוד). ההבדל המשמעותי בין השניים זו הצורה שבה מביאים את ה Reviews כאשר הקריאות מתבצעות במקביל וממתנות עד לסיום הריצה

המון תודה, זו הייתה אחלה משימה

סוף שבוע נעים

זיו