

Postup řešení

Jelikož jediný programovací jazyk, který momentálně ovládám, je Python, byla jeho volba jasná. Bohužel, po přibližně třech měsících kurzu mé znalosti stále nejsou dostatečné, a proto jsem se velmi často radil s AI. Z dřívějších rozhovorů s bratrem, který mi kdysi popisoval svou práci junior testera webových aplikací, jsem měl základní představu, jak postupovat. Hned mi bylo jasné, že pro tuto úlohu bude dostačující kombinace Python, Selenium a ChromeDriver. Sice toto řešení pravděpodobně není nejrychlejší, protože simuluje chování uživatele na webové stránce, ale pro mé potřeby bylo dostačující.

Jak program funguje

Po proklikání jednoduchého menu se spustí proces zachytávání komentářů. Tento proces může běžet na pozadí nebo ve viditelném okně, kde lze sledovat, jak Python pomocí Selenia ovládá webový prohlížeč. Pomocí zobrazení zdrojového kódu jsem dohledal selektory, které jsou jedinečné pro:

- komentář,
- autora,
- odkaz na článek, kde byl komentář zveřejněn.

Tyto zachycené údaje se následně ukládají do databáze, kde lze vyhledávat fráze a exportovat komentáře do HTML souboru.

Snažil jsem se kód optimalizovat pro rychlejší scrapování, ale ukázalo se, že bez hlubších znalostí optimalizace začne zachytávání komentářů padat. Aplikace si pak myslí, že získala všechny komentáře, ve skutečnosti ale může obsahovat například jen 15 % skutečného obsahu.

V budoucnu bych rád přišel na způsob, jak tento proces urychlit a zpřesnit.