

Webscraper/Metawyszukiwarka

- implementacja web crawlera automatycznie odwiedzającego różne serwisy i zbierającego z nich relewantne informacje do wskazanych słów lub innych danych wejściowych (np. zdjęcia) (opcjonalne)
- jednolity sposób wyświetlania wyników z deduplikacją oraz informacją o źródle danych
- podstawową właściwością projektu ma być jego **rozszerzalność o nowe źródła danych** np. poprzez implementację parsera interpretującego strony poprzez wcześniej dostarczone szablony
 - można to również zrealizować poprzez automatyczną analizę strony i identyfikację jej kluczowych elementów (searchbox, seria wyników, odnośnik do następnej strony)
- uwzględnić potrzebę odwiedzenia kolejnych stron wyników z pojedynczego źródła oraz ich odpowiedniego sortowania

Technologie i materiały powiązane:

- Webkit
- <https://stackoverflow.com/questions/18539491/headless-browser-and-scraping-solutions>
- https://en.wikipedia.org/wiki/Website_Parse_Template

Temat:

Metawyszukiwarka cen kryptowalut.

Dlaczego wybrano dany temat?

- zdobycie doświadczenia w tworzeniu narzędzi tego typu pozwalający uzyskiwać dowolne dane
- stworzone narzędzie może zostać wykorzystane w ułatwieniu śledzenia kursów walut
- pozwoli ułatwić analizę ceny wybranej waluty
- w przyszłości może być rozbudowane o boota handlującego

Zbierane wyniki:

- Nazwa waluty
- Cena
- Zmiana ceny w przedziale 24h
- Dzienny wolumen
- Market Cap waluty
- Market Cap giełdy

Terminarz:

- zbieranie podstawowych wyników z giełdy
- zbieranie wyników z uwzględnieniem podstron giełdy
- selekcja danych
- stworzenie bazy danych
- moduł dodawania nowych giełd
- prezentacja wyników na stronie

Narzędzia:

- język Java
- biblioteka jsoup
- język HTML CSS
- framework bootstrap
- Hibernate/JPA