

Опис на проект – Web Scrapping Апликација

Краток опис

Овој проект вклучува развој на веб скрепинг апликација што автоматски ги презема и ажурира финансиските податоци за различни издавачи на македонската берза во период од 10 години. Податоците се преземени од веб-страница на берзата, која обезбедува историски статистички податоци за различни финансиски анализи. Апликацијата презема информации како што се кодот на издавачот, цената, обемот и промената, осигурувајќи дека нема да се додаваат дупликати во складираната база на податоци. Со користење на датотека за складирање на податоците, се следат преземените податоци, овозможувајќи записите што недостасуваат или застарени се ажурираат кога е потребно.

Функционални барања:

- **Преземање кодови на издавач:** Системот ќе превзема список со кодови на издавач од дадена URL-адреса, ги складира во датотека исклучува кодовите со нумерички знаци.
- **Преземање историски податоци:** за секој код на издавачот, системот ќе презема финансиски податоци кои се водат во последните 10 години, подредени по годишни периоди, без да се вметнуваат дупликат записи во датотеката.
- **Форматирање и валидација на податоците:** Системот ќе ги внесува датумите во конзистентен формат и цените имаат соодветно форматирање (запирка за илјадиници и точка за децимали).
- **Ажурирање и ракување со податоци:** ако податоците на издавачот недостасуваат или се застарени, Системот ги враќа најновите податоци за тој издавач, ажурирајќи го документот без дуплирање на постоечките записи.
- **Интегритет на податоците:** Системот ќе проверува дали сите преземени податоци се валидни и дека неважечките или празните записи се прескокнати.

Нефункционални барања:

- **Перформанси:** Системот мора ефикасно да ракува со повеќе кодови на издавачи, притоа преземајќи и обработувајќи податоци без значителни одолжувања.
- **Сигурност:** Системот мора да биде робустен и флексибилен, способен да се опорави од неуспешни барања или прекини и да не обезбеди оштетување на податоците.
- **Употребливост:** Системот треба да биде лесен за употреба, со јасни пораки за евиденција што укажуваат на напредок и грешки за време на процесот на скрепинг и ажурирање.
- **Безбедност на податоците:** Системот мора да се погрижи да се одржува приватноста на податоците и да не се изложуваат чувствителни информации за време на процесот на преземање податоци.
- **Приспособливост:** Системот треба да биде скалабилен за да прими повеќе кодови на издавачи или да го продолжи периодот на преземање податоци во иднина доколку е потребно.

Кориснички сценарија:

- **Финансиски аналитичар:** Финансиски аналитичар, има потреба од податоци за различни финансиски инструменти за да ги процени трендовите на пазарот. Тој го користи системот за автоматско преземање и складирање на податоците за анализа, ја конфигурира апликацијата да влече податоци за изминатите 10 години, осигурувајќи дека ги добива најсовремените достапни информации.
- **Системски администратор:** Системски администратор треба да се погрижи податоците зачувани во датотеката да се актуелни и да не се дуплираат. Редовно го користи системот за да провери дали има исчезнати податоци или застарени записи, осигурувајќи дека датотеката секогаш се ажурира без вишоци.

Персони:

Сара - финансиски аналитичар: Сара е професионалец кој работи во финансиска фирма која користи финансиски податоци за да донесува информирани одлуки. Таа е запознаена со алатките за анализа на податоци и бара систем кој автоматски презема точни податоци за да заштеди време при рачно собирање податоци.

Петар – Системски администратор: Петар е задолжен за одржување на системот што обработува финансиски податоци. Тој се грижи за дата фечинг и апдејтинг процесот дека процесот ќе работи непречено без грешки или дупликат записи.

Описен наратив:

Системот започнува со преземање кодови на издавачот од одредена URL адреса, осигурувајќи дека собира само важечки кодови. Откако ќе се генерира списокот со кодови на издавачот, системот започнува да презема податоци за изминатите 10 години. Податоците се преземаат во парчиња врз основа на годишни интервали, а системот го проверува секој запис за да се осигура дека нема да се додаваат дупликации. Сите нови податоци или ажурирања за издавачот се додаваат во база на податоци. Овој процес е автоматизиран за да се осигура дека финансискиот аналитичар или администратор нема потреба рачно да презема или ажурира податоци, заштедувајќи време. Системот проверува и за исчезнати или застарени записи, со што осигурува дека податоците се секогаш ажурирани и точни.

Со автоматизирање на процесот на преземање и ажурирање финансиски податоци, системот им овозможува на финансиските експерти да се фокусираат на анализа наместо на собирање податоци, подобрувајќи ја ефикасноста на работниот тек.