EXTRAGEREA DATELOR DESPRE PRODUSELE UNUI

MAGAZIN ONLINE FOLOSIND WEB SCRAPING

Nume şi prenume profesor laborator: ANTON LICZ

Nume şi prenume student: ASAVOAEI ANDREEA-DENISA

Specializarea, anul, grupa, subgrupa: INFORMATICĂ ROMÂNĂ, 2, 1, 1

Limbaj de programare folosit: PYTHON

Acest proiect reprezintă o aplicaţie creată cu ajutorul limbajului de programare Python şi al tehnicii de Web Scraping. Web Scraping-ul este o practică care presupune obţinerea de informaţii de pe un site web prin interogarea serverului, crearea unui *request data* care preia codul HTML al site-ului şi apoi parsarea acestui cod pentru a extrage datele dorite.

Programul implementat extrage date cu privire la ofertele de notebook-uri puse la dispoziţie de site-ul web al cunoscutului magazin Altex. Informaţiile se referă la titlul laptopului, preţul actual, preţul iniţial(dinaintea aplicării reducerii), nota acordată de clienţii magazinului şi eventual o imagine cu produsul şi/sau specificaţiile acestuia. Datele vor fi stocate într-un tabel sau într-un fişier cu extensia .csv creat pentru parcurgerea mai facilă a acestora.

Utilitatea proiectului este aceea de a accesa rapid informaţiile şi preţurile produselor respective fără a folosi un browser care poate îngreuna utilizarea computerului personal prin ocuparea unui segment mare de memorie.

Modulele(bibliotecile) Python folosite la implementarea iniţială a codului sunt *urllib* şi *bs4*(BeautifulSoup).

Menţiuni:

1. Magazinul sau categoria de produse se poate modifica în funcţie de dorinţa utilizatorului prin adaptarea codului sursă în funcţie de codul HTML al site-ului dorit astfel încât să se evite erorile produse de diferenţele dintre codurile HTML ale paginilor.

2. Deoarece produsele sunt afişate pe site încadrate în mai multe pagini şi, implicit, în mai multe coduri HTML, este posibil să fie necesară separarea codului în funcţie de pagini.