

Task: Utilizarea serviciilor web pentru extinderea arhivei de filme

Curs: Data Management and Storage with Python

Modul: Bazele schimbului de date - XML, JSON și serviciile web

Scopul sarcinii:

Scrieți un script Python care utilizează un serviciu web pentru preluarea datelor și integrarea lor în setul de date inițial.

Contextul sarcinii:

Sunteți angajat ca analist de date de prestigiosul festival de film *CineVerse*. Festivalul prezintă filme din toată lumea de câteva decenii și are loc o dată pe an timp de 7 zile. Organizatorii festivalului au colectat date despre toate filmele care au fost prezentate vreodată la festivalul lor de film.

Ideea organizatorilor festivalului este de a extinde setul de date colectate cu recenzii de pe platforma IMDB, date despre distribuție și popularitatea filmului (număr de voturi pe IMDB). Astfel de date vor servi drept bază pentru implementarea unei funcționalități pe care o echipă de programatori o va implementa pe site-ul lor.

Sarcină:

Sarcina voastră este să extindeți setul de date inițial cu date suplimentare pe care le veți obține folosind un serviciu web. Setul inițial constă din date despre filme și poate fi descărcat de aici:

movies.7z

În cadrul fișierului movies.csv se află date despre filme, care includ:

- titlul;
- anul lansării;
- genul;
- numele regizorului;
- țara de producție;
- durata.

La astfel de date trebuie să se adauge:

- evaluarea de pe IMDB;
- numele actorilor principali;
- numărul de voturi de pe platforma IMDB.

Veți obține aceste date prin API-ul OMDb (Open Movie Database API), un serviciu web gratuit care oferă acces la o bază de date mare de filme și seriale. O scurtă familiarizare cu un astfel de serviciu web este prezentată în următorul tutorial video:

Tutorial video: Ce este și cum se folosește API-ul?

Trebuie să accesați datele solicitate prin serviciul web OMDb API. Trebuie să integrați astfel de date în setul de date inițial și apoi să le exportați pe toate într-un fișier XML separat. În cele din urmă, la ieșire trebuie să afișați cele 10 filme cu cele mai bune evaluări IMDB.

Trimiterea sarcinii

După finalizarea sarcinii, trebuie să-i trimiteți coordonatorului cursului codul sursă al proiectului complet sau folderul rădăcină al proiectului, pe care trebuie mai întâi să îl arhivați într-o arhivă (.rar, .zip, .7z).

Cerințe de evaluare:

Coordonatorul va evalua sarcina pe baza următoarelor criterii:

1. Citirea corectă a datelor dintr-un fișier CSV și importarea datelor în program **(10%)**.
2. Contactarea unui serviciu web și citirea răspunsului serviciului **(20%)**.
3. Parsarea răspunsului serviciului și izolarea datelor necesare **(20%)**.
4. Integrarea datelor obținute în setul inițial **(10%)**.
5. Tratarea adecvată a posibilelor erori în comunicarea cu serviciul **(10%)**.
6. Salvarea cu succes a setului de date extins într-un fișier XML **(10%)**.
7. Afișarea celor 10 filme cu cele mai mari evaluări IMDB **(10%)**.
8. Calitatea (reutilizabilitatea, lizibilitatea, structura) codului **(10%)**.

Ghid pentru rezolvarea sarcinii:

Sarcina poate fi rezolvată într-un număr mare de modalități. Pașii tipici care trebuie luați pentru finalizarea sarcinii sunt:

- **înregistrarea** pe site-ul <https://www.omdbapi.com/> și obținerea unei chei API, fără de care nu puteți contacta serviciul web;
- **citirea** filmelor din setul de date inițial (din fișierul `movies.csv`);
- **iterarea** prin filmele pe care le-ați citit și citirea titlurilor și a anului lansării;
- **formarea** unei adrese URL pentru contactarea serviciului;
- **trimiterea** de cereri către serviciu și analizarea datelor JSON primite, ținând cont de eventualele erori;
- **găsirea** datelor necesare în răspunsul serviciului;
- **integrarea** datelor obținute în structura inițială;
- **salvarea** structurii de date modificate într-un nou fișier XML;
- **sortarea** înregistrărilor după coloană cu evaluări IMDB în ordine descrescătoare;
- **afișarea** primelor 10 filme din structura de date tocmai sortată.