

Regăsirea evenimentelor pe web

Proiectul își propune implementarea mai multor module care să obțină date ce contribuie apoi la prezentarea completă a unor informații de interes pentru utilizatori, mai exact prezentarea evenimentelor organizate de instituțiile culturale sau de divertisment din cadrul unui oraș. Astfel, se dorește agregarea informațiilor obținute din surse diferite în cadrul unei singure aplicații.

Aplicația este formată din patru module independente care comunică folosind modelul pipeline: un modul cu rol de web crawler, modulul de parsare, modulul de comunicare cu API-ul oferit de Facebook și modulul care implementează interfața cu utilizatorul. Un modul suplimentar este cel de servicii web prin care se face accesul la baza de date.

Web crawler-ul este cel care primește adresa unui site web ca URL de start și obține toate URL-urile relevante, respectând restricțiile de parcurgere impuse de acel site. În cazul de față, a fost folosit site-ul turistic www.tripadvisor.com pentru a obține pagini web ce conțin informații de bază despre restaurante, hoteluri, muzee și alte localități de interes din orașe și țări din toată lumea. Toate acestea sunt reținute în baza de date și vor constitui intrarea pentru următorul modul.

Modulul de parsare are rolul de a parcurge paginile obținute de web crawler și de a extrage informațiile necesare dintre care principalele sunt: numele, adresa, numărul de telefon, scurtă descriere, precum și adresa site-ului web al locației în cazul în care acesta este pagina Facebook de prezentare. Similar, toate datele obținute sunt salvate în baza de date.

Următorul modul este cel care folosește funcțiile furnizate de API-ul Facebook. Astfel, sunt utilizate paginile Facebook ale locațiilor obținute anterior pentru a accesa evenimentele ce urmează a fi organizate de acestea. Totodată, sunt reținute informații specifice ale evenimentelor cum ar fi: data de desfășurare, descrierea, numărul persoanelor care sunt interesate, care participă sau nu.

Ultimul modul reprezintă interfața cu utilizatorul și expune informațiile obținute de implementările anterioare într-un mod unitar. Astfel, utilizatorul poate filtra căutarea selectând de exemplu orașul și data calendaristică pentru care dorește să vizualizeze evenimentele, iar rezultatele constau în date extrase atât de modulul de parsare, cât și prin interogarea API-ului Facebook. Lista evenimentelor returnate va cuprinde informații despre locația care le organizează, inclusiv localizarea geografică afișată cu ajutorul Google Maps, precum și detalii despre eveniment. Aplicația permite crearea unui cont care îi oferă utilizatorului acces la funcționalități suplimentare: salvarea unor locații ca favorite, salvarea evenimentelor la care utilizatorul dorește să participe, vizualizarea istoricului acestora, precum și afișarea de notificări atunci când locațiile favorite organizează un eveniment nou.

Serviciile web implementate sunt de tip REST, bazate pe protocolul HTTP și impun constrângeri la un set cunoscut de operații. Prin intermediul lor, este permis accesul la baza de date SQL și se face schimb de date între modulele prezentate. Reprezentarea resurselor este decuplată de acestea, fiind accesibilă în mai multe formate (HTML, JSON, XML).

Implementarea modulelor de pe partea de server s-a făcut folosind tehnologia Java în mediul de dezvoltare Eclipse, iar partea de client este reprezentată de o aplicație bazată pe arhitectura MVC.

Se urmărește astfel o modularizare cât mai bună a aplicației atât pe partea de server, cât și pe partea de client în vederea unor eventuale oportunități de extindere. Pe de o parte, se poate

opta pentru implementarea și altor module care să contribuie cu noi informații și pe de altă parte, se pot dezvolta aplicații desktop sau mobile pentru interacțiunea cu utilizatorul. Comunicarea cu baza de date a oricărui viitor modul se va face tot prin intermediul serviciilor web.

Profesor îndrumător,
Sef lucr.dr.ing. Adrian Alexandrescu

Student,
Cobelea Diana