

TripPlanner

Aplicarea tehnicilor de regăsire a informațiilor și a algoritmilor genetici în planificarea călătoriilor

Andreea-Roxana Coșmănescu

Rezumat

Lucrarea își propune aplicarea metodelor de regăsire a informațiilor și a algoritmilor genetici pentru obținerea itinerarului unei călătorii.

Aplicația a fost proiectată în două module: primul modul se ocupă cu colectarea datelor necesare aplicației și stocarea lor în baza de date, iar cel de-al doilea modul implementează algoritmi genetici și interfața aplicației.

Sistemul din primul modul al aplicației are la bază o arhitectură ale cărei componente sunt:

- entități de tip *crawler* - care extrag de pe web link-uri ce ar putea conține informații utile aplicației
- cozi de date - ce conțin URL-urile extrase de entități de tip *crawler*
- entitate de tip *load balancer* - care gestionează cozile de date
- entități de procesare - care parcurg paginile web localizate prin URL-urile extrase de roboți și extrag informații referitoare la obiective turistice, hoteluri, restaurante, etc
- entități de tip *worker machine* – sunt entități care știu să pornească anumite procese atunci când primesc o comandă
- entitate de tip *deployer* – conține o evidență a mașinilor de tip worker și trimite comenzi acestora când este necesar

Toate aceste componente sunt independente unele de celelalte și comunică pentru a asigura funcționarea corectă a sistemului. Parsarea paginilor web se face în urma unei analize a elementelor din pagină și a stabilirii unui anumit format pentru fiecare site de interes.

Sistemul astfel construit este ușor de configurat și este folosit la extragerea datelor și la salvarea lor în baza de date după aplicarea unor filtre care elimină datele considerate invalide. Un astfel de sistem se pretează atunci când există un volum mare de date ce trebuie procesate și se dorește reducerea timpului de execuție prin distribuirea pe mașini de lucru diferite.

Cel de-al doilea modul cuprinde implementarea algoritmilor care vor construi traseul optim al orașelor, precum și cel al obiectivelor turistice.

Problema se reduce la găsirea unui algoritm care să determine un traseu optim, având ca date un set de orașe, un punct de start și un punct final și faptul că se dorește un traseu cu o distanță minimă. De asemenea, utilizatorul poate să-și exprime preferințele în ceea ce privește categoria de atracții turistice de care este interesat precum și numărul de obiective pe care dorește să le viziteze într-un anumit oraș. Aceste preferințe ale utilizatorului sunt luate în considerare atunci când se determină itinerarul călătoriei.

Luând în considerare specificațiile menționate, un algoritm genetic este potrivit pentru o astfel de problemă, pentru care spațiul de căutare este foarte mare și pentru care se dorește un timp de răspuns cât mai mic.