

SCSS 2017: Secțiunea **Aplicații software și tehnologii web** - rezumatele lucrărilor -

1. Alocarea eficientă a task-urilor într-un sistem distribuit

Cristian-Ilie Vasluianu

Aplicația dezvoltată are drept scop eficientizarea execuției task-urilor într-un sistem distribuit, prin utilizarea resurselor într-un mod cât mai dinamic, în funcție de complexitatea task-urilor. Apariția sistemelor de calcul tot mai performante cu resurse hardware impresionante impune nevoia de dezvoltare a aplicațiilor care să utilizeze aceste cerințe hardware la maximum, prin rularea rapidă și fluentă a task-urilor. Alocarea resurselor dinamic se va face cu ajutorul tehnicii de „load balancing” [1], tehnică care are rezultate foarte bune atunci când un număr foarte mare de task-uri trebuie procesate concurrent. Task-urile vor ajunge la un dispatcher ulterior fiind împărțite către mai mulți workeri unde se va realiza prelucrarea efectivă a task-urilor. Pentru a observa cât mai bine modul în care se gestionează resursele de către server, s-a optat pentru o aplicație practică care necesită timpi foarte variați de calcul și anume generarea unor fractali de diferite modele, culori și alte opțiuni de personalizare.

2. Soluții de securizare pentru sisteme informatice medicale

Vlad Micu

Pacienții unui spital au nevoie de o monitorizare a parametrilor vitali, atât în perioada de spitalizare, cât și în perioada imediat următoare, la domiciliu. Sistemele de monitorizare la distanță oferă date precise și în timp util, însă au nevoie de o securitate robustă. În absența unui modul de securitate, datele pot fi ușor interceptate sau accesate de persoane neautorizate. Cazul cel mai sensibil îl reprezintă monitorizarea pacientului la domiciliu. Acest proces implică o comunicare prin rețele nesigure, caz în care crește nevoia unei soluții de securitate. Documentul de față prezintă o astfel de soluție, ce cuprinde metode sigure de autentificare, autorizare și gestionare a identităților. Modulul de autentificare se bazează pe protocolul de autentificare Kerberos și protocolul criptografic SSL. Autorizarea este realizată atât prin intermediul sistemului de gestionare al identităților – LDAP – cât și prin intermediul operațiilor de genul „cerere – aprobare”. Comunicația este realizată sub o arhitectură REST, ce oferă o mapare intuitivă a resurselor și o flexibilitate în ceea ce privește comunicarea între tipuri diferite de dispozitive.

3. Geo-localizarea și clusterizarea pe hartă a tweet-urilor în timp real

Mihaela-Adriana Ciobanu

Acest proiect are scopul de a monitoriza în timp real evenimente de interes general prin preluarea după o anumită interogare a Streaming API-urilor de la Twitter, pentru a se putea observa impactul lor asupra populației pe glob. Localizarea se realizează în funcție de coordonatele geografice ale tweet-ului, urmând poziționarea acestora pe harta lumii cu ajutorul

API-ului de la Google Maps. Pentru o vizualizare mai bună, a fost implementat algoritmul DBSCAN (engl. “Density-Based Spatial Clustering of Applications with Noise”) realizându-se în acest mod evitarea suprapunerii marcărilor pe hartă. Doar tweet-urile care au activată opțiunea de geo-tagging pot fi afișate pe hartă, ceea ce reprezintă doar 20% din totalul de tweet-uri trimise de pe glob. Din acest motiv, pentru a veni în ajutorul utilizatorului de a afla rezultate semnificative apropiate de realitate, s-au preluat toate tweet-urile filtrate după un anumit subiect și s-a realizat analiza conținutului cu ajutorul modulului Sentiment. Această bibliotecă de analiză a sentimentului are rolul de a verifica dacă subiectul de interes este menționat în tweet-uri în mod pozitiv sau negativ. Aplicația mai are rolul și de a ține la curent utilizatorul cu cele mai trending tweet-uri ce sunt discutate de populația de pe glob.

4. Studiul expunerii unui sistem MapReduce sub formă de servicii REST

Flavius Condrea

Scopul proiectului este studiul MapReduce și al posibilității de a expune această funcționalitate sub formă de servicii REST. În cadrul acestui proiect s-a dezvoltat suportul pentru putea încărca, prin intermediul unei interfețe web, fișiere executabile, iar acest suport oferă posibilitatea de a executa aplicațiile încărcate conform algoritmului MapReduce. Comunicarea din cadrul componentelor sistemului MapReduce s-a realizat prin intermediul web socket-urilor, iar acest sistem este expus prin intermediul unor servicii de tip REST. Avantajele unei astfel de expuneri ar putea fi un caracter mai general și scăderea costurilor hardware.

5. Sisteme de regăsire a informațiilor. Colectarea datelor web

Ioan-Adrian Duliba

Un motor de căutare este un ansamblu de programe care au ca scop colectarea datelor expuse pe web, pregătirea acestora pentru diferite modele de căutare și realizarea efectivă a funcției de căutare. Pentru crearea unor astfel de motoare de căutare este necesar un “robot” Web (bot, spider). Activitatea unui robot Web este cea de a colecta, într-un mod automat, resurse web. Dezideratele pentru crawlerele Web se încadrează în două categorii: caracteristicile pe care trebuie să le furnizeze (robustețe și respectarea pseudo-protocolului REP), urmate de cele pe care ar trebui să le furnizeze (scalabilitate, extensibilitate, distributivitate, utilizarea eficientă a resurselor de sistem și calitatea paginilor colectate). Scopul acestei lucrări este de a studia principalele componente ale unui robot Web și de a propune soluții de implementare pentru toate aceste componente.

6. Sisteme de regăsire a informațiilor - Metode de indexare și căutare a datelor

Anca Farcaș

Tema abordată în proiectul de diplomă este procesul de indexare și căutare în motoarele de căutare. Indexarea poate fi privită ca un proces ce oferă o relație de ordine asupra datelor pe care le analizăm. În această lucrare sunt studiate două tipuri de indexare: directă și inversă. În studiul indexării pre-procesarea este considerată componenta de analiză formei canonice a cuvintelor. Pentru a aduce cuvântul la forma sa canonică au fost construiți doi arbori care comunică între ei,

unul din aceștia fiind populați cu toate cuvintele limbii engleze, iar al doilea cu formele neregulate ale substantivelor, verbelor etc., fiecare cuvânt din arborele neregulat având legătură cu acel cuvânt care îi corespunde în arborele dicționar. Scopul acestei abordări este de a soluționa acele cazuri în care algoritmi de stemming (Porter, Lovins și Paice) nu oferă rezultate satisfăcătoare. Această formă de preprocesare este aplicată atât în faza de construire efectivă a indexului, cât și în component de căutare. Pentru componenta de căutare, accentul s-a pus pe modulul de căutare de tip boolean și cel bazat pe reprezentarea vectorială.

7. Sisteme de regăsire a informațiilor. Metode avansate de analiză a datelor

Roxana Ungureanu

Tehnologia de astăzi ne permite să producem și să stocăm un volum imens de date. Identificarea de tipare, tendințe și anomalii în seturile de date și sintetizarea lor în modele clare este una din provocările erei informațiilor. Clusterizarea reprezintă o metodă avansată de analiză a datelor care are aplicabilitate și în domeniul regăsirii informațiilor. Aceasta se bazează pe învățare nesupervizată, numită și învățare prin observare. În această lucrare am scos în evidență avantajele și dezavantajele a 3 metode de clusterizare a documentelor: clusterizare prin partiționare (K-medoids), clusterizare ierarhică (DIANA) și clusterizare bazată pe densitate (DBSCAN). Rezultatele de performanță pentru fiecare dintre cele 3 metode au fost obținute prin aplicarea unor implementări proprii ale algoritmilor K-medoids, DIANA și, respectiv, DBSCAN, pe un set de documente text. Rezultatele obținute par să favorizeze metodele bazate pe densitate.

8. Modele de căutare pentru documente multimedia

Ionuț-Dragoș Roban

Lucrarea de față presupune o aplicație complexă ce urmărește regăsirea informațiilor, pe web, bazate pe un subiect predefinit de către autor, mai exact implementarea unui robot de căutare ce facilitează munca utilizatorului de a căuta informațiile dorite pe web. Aplicația propusă are prestabilit un set de informații pe baza căruia se extrag de pe web, dintr-un număr restrâns de locații/domenii, date cu privire la tema specificată. Această lucrare constă în restrângerea domeniului de căutare, caracterul documentelor căutate fiind bazat pe subiectul celor de tip multimedia, știri, noutăți, ziare online. În ceea ce privește rezultatul lucrării, comportamentul este similar unui portal web ce reprezintă un site web complex ce realizează schimbul de informații între aplicații/sisteme independente din punct de vedere al limbajului de care sunt descrise, presupunând anumite proceduri unificate de regăsire a informațiilor. Cu alte cuvinte, lucrarea descrie un „Vertical Portal Web (Vortal)”. Un vortal este o particularizare a conceptului de „Portal Web” ce descrie un comportament bazat pe un subiect concret, în care se restricționează scopul regăsirii informațiilor în funcție de tema/semnificația celor dorite, în cazul de față fiind multimedia, știri, articole.