

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

Mihaela-Adriana Ciobanu

Rezumat

Scopul principal al acestui proiect este de a informa utilizatorii cu privire la preponderența interesului populației pentru diferite evenimente curente. Acest lucru a fost realizat prin monitorizarea în timp real a subiectelor de interes general cu ajutorul Streaming API-urilor de la Twitter, preluate după o anumită interogare, și prin extragerea celor mai trending tweet-uri ce sunt discutate pe glob.

Structura aplicației cuprinde două părți importante:

- Un server ce are scopul de a realiza conexiunea la Twitter și de a procesa rezultatele primite, cât și trimiterea unui email cu cel mai popular tweet din ziua respectivă tuturor utilizatorilor abonați la newsletter,
- Interfața cu utilizatorul ce permite accesul la resursele furnizate de serverul de procesare.

Aplicația permite utilizatorului posibilitatea explorării distribuției geografice a tweet-urilor prin geo-localizarea lor în timp real, preluate prin conexiunea Streaming de la Twitter, în funcție de coordonatele geografice ale acestora. Poziționarea lor pe harta lumii este realizată cu ajutorul API-ului de la Google Maps, cu scopul de a se observa impactul asupra populației pe glob a subiectului introdus de utilizator în inputul de search. Pentru o vizualizare mai bună, a fost realizată clusterizarea în timp real a tweet-urilor prin adaptarea și implementarea algoritmului de clusterizare DBSCAN (engl. “Density-Based Spatial Clustering of Applications with Noise”), realizându-se în acest mod evitarea suprapunerii marcajelor pe hartă datorate volumului mare de date provenite din streamuri.

Un dezavantaj al afișării pe hartă a tweet-urilor este faptul ca numai cele care au activată opțiunea de geo-tagging pot fi reprezentate, 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, inclusiv cele care nu au coordonate geografice și s-a realizat analiza conținutului acestora cu ajutorul modulului Sentiment instalat din managerul de pachete NPM al Node.js. 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.

De asemenea este oferită posibilitatea utilizatorului de a vizualiza printr-un Pie Chart cele mai trending tweet-uri ce sunt discutate pe glob și de a se abona la newsletter-ul paginii pentru a primi pe mail cel mai popular tweet din fiecare zi.

Prin combinarea tuturor acestor trăsături, aplicația oferă rezultate semnificative în contextul obținerii opiniei generale cu privire la un anumit subiect, oferind totodată posibilitatea unor dezvoltări ulterioare prin explorarea API-urilor de la Twitter și de la Google Maps.