

Analiza știrilor online pentru detecția titlurilor false

Constantin Croitoru

Rezumat

Prelucrarea informațiilor în domeniul jurnalistic, în mass-media contemporană creează un mediu favorabil pentru distribuirea știrilor false. Studiul unor metode prin care se pot identifica automat astfel de știri face obiectul mai multor dezbateri și a unor încercări de dezvoltare de programe informatice care, însă deocamdată nu răspund tuturor necesităților. Un caz particular al acestui fenomen îl reprezintă titlurile atractive care induc cititorului o idee greșită despre conținutul articolului. Acest fenomen poartă denumirea de clickbait.

În lucrarea de față se propun și se analizează metode prin care se pot identifica și semnala cititorului posibilele știri din mediul online în care discrepanța dintre titlu și conținut este considerabilă. Dintre algoritmi specifici inteligenței artificiale a fost studiat și folosit în special algoritmul k-Nearest Neighbors(kNN) cu rolul de a identifica dacă titlul este relevant în raport cu conținutul articolului. Modul de utilizare și implementare a algoritmului și metricile folosite alături de alte metode sunt explicate pe larg în această lucrare.

S-au utilizat limbaje de nivel înalte precum Java, prin intermediul cărui s-a facilitat lucrul cu procesările de text, pentru analiza cuvintelor din titlu și a corespondenței acestora cu frecvența apariției lor în conținutul știrii. Analiza rezultatelor sugerează că aceste metode pot fi utile, eventual adăugându-le diverse îmbunătățiri.

În capitolul I s-au luat în considerare mai multe aspecte teoretice legate de problema relevanței titlului în raport cu analiza conținutului știrilor online și sunt prezentate diverse tehnologii ce pot ajuta la realizarea acestui obiectiv. În capitolul II sunt explicate pe larg modalitățile de proiectare a proiectului, apoi, în capitolul III implementarea acestuia este explicată. La final sunt discutate rezultatele(capitolul IV) și sunt expuse concluziile finale.