

# Barack Obamaren, edo beste norbaiten, tweet-etatik mamia ateraz



GIZARTE-SAREAK

**Igor Ruiz-Agundez**

**DeustoTech  
Computing, Deusto  
Institute of Technology,  
Deustuko  
Unibertsitatea**

**igor.ira@deusto.es**

**Izaskun Canga-  
Sanchez**

**Sapienza Università  
di Roma**

**izaskun1982@yahoo.  
com**

**Marco Guidi**

**Department of  
Pedagogical,  
Psychological and  
Didactical Sciences,  
Università del Salento**

**marcoguidi73@gmail.com**

## Sarrera

Sare sozialak Interneteko gizakion arteko erlazioen isla dira, komunikatzeko eta erlazionatzeko bide berriak irekitzen dituzte. Sare hauek datuz beterik daude eta datu-meatzaritzako tekniken bidez jakintza eskuratu daiteke. Lan honek Twitter sare sozial ospetsuaren erabiltzaile esanguratsu batek duen portaera aztertzen du, Barack Obamaren Twitterren erabilera, alegia. Kontua sortu zuenetik sareratu dituen mezuak aztertzen dira, jakintza handiagoa eskuratzeko asmoz. Planteatzen dugun metodologiari esker, esanahi-sortze bat gauzatu dugu. Hau da, sarrera-datuetatik jakintza zabalago bat lortzen dugu hasierako datuei esanahi berri eta orokorragoa emanez. Prozesu honetan, kluster analisi bat burutzen dugu Barack Obamak Twitterren sareratutako mezuetan. Emaitzak ikertuz, klusterizazio edo multzokatzearen bidez sorturiko mezu taldeek Obamak izan duen programa politikoan nolabaiteko isla izan dutela nabarmendu daiteke.

## Esperimentua

Bururiko esperimentuaren corpus edo testu-datuak Barack Obamak Twitterrera igorritako 895 post, mezu edo tweetez osaturik daude. Datuak Obama Estatu Batuetako senataria zenean hasi, 2007ko apirilaren 29tik, eta 2010eko irailaren 6 arte bildu dira. Guztira, 16.000 karaktere baino gehiago eta 3.000 unitate lexikal inguruko corpusa lortu dugu.

Tweet bakoitzak hiru atributu nagusi ditu. Lehenengoa Twitter modu bakarrean identifikatzen duen identifikadorea da. Bigarrena mezu bera osatzen duen testua da. Eta hirugarrenak tweeta sortu den ordua eta data adierazten ditu. Beste atributu askok ere osatzen dute tweeta: koordenatu geografikoak, beste mezuekiko erlazioak, sorrera denbora absolutuak, etab. Aparteko datu horiek ez ditugu gure esperimenturako erabiliko.

Gure esperimentuan, tweeten testuaren edukia eta sorrera-data eta ordua bakarrik hartzen ditugu kontuan. Zehatzago esanda, mezu bakoitzak sorrera-momentu bat duenez, datak denbora-tarte jakinetan bildu ditugu, datuen errepresentazioa handitzeko.

Datu hauei testu-analisirako T-Lab software-tresna aplikatu diegu. Software horrek testu-edukien analisia eta testu-meatzaritzaz ahalbidetzen ditu, gure esperimenturako beharrak betez. Programak eskaintzen dituen gaitasunen artean ondorengoak aipa ditzakegu:

- Agerkideztza-analisia (hitzen asoziazioa, hitz bikoteen konparaketa, kontzeptu-mapak, sekuentzia eta komunztadura-analisia).
- Gaien analisia (gaien sorkuntza; gaien analisia, oinarritzko testuinguruen bidez; gaien sekuentzia; testuinguru erabakigarria; dokumentuen sailkapena gaiaren arabera).

- Analisi konparatiboa (espezifikazio-analisia, korrespondentzia-analisia, korrespondentzia anizkoitzen analisia, multzokatze analisia eta kontingentzia-etaulak).

Ezaugarri horiek edozein eratako dokumentuetan aplikatu daitezke (diskurtsoak, egunkariko artikulak, inkestak, liburuak, etc.). Hala ere, guk dakigunetik, inork ez ditu, orain arte, sare sozialetako testuak modu horretan ikertu.

Egin daitezkeen analisi guztien artean, datu horien oinarritzko testuinguruak aztertzea erabaki dugu. Analisi horrek bilduriko tweetak multzokatuko ditu esanahi-sortze berri bat emanez.