

europapress / país vasco / euskera

5 por  
sorpresa

MIÉRCOLES 21:45 H

**DKISS**

También disponible en

dplay

Ver ahora ▶

[Privacidad](#)

noiznahi, doan. Informazio horrek "berebiziko garrantzia du erakunde, enpresa, eragile eta abarrentzat, beti izan baitute interesa jakiteko herritarrek, erabiltzaileek edo bezeroek zer iritzi duten haiei buruz. Tradizionalki, kostu handia duten inkesta edo galdera-sorta bidez eskuratu izan dute informazioa, baina laginak beti txikiak izan dira", azaldu du EHUko IXA Taldean zuzendutako lanaren egile nagusi Iñaki San Vicentek.

Orain, berriz, datu kantitate erraldoiak sortzen dira sare sozialetan, eta, oro har, Interneten. Erronka, ordea, hortik interesa duen informazioa behar bezala erauzi eta sailkatzea da. 'Sentimenduen Análisis' deritzon ikerketa arloak edonolako testu batean iritzi edo sentimendu positibo edo negatiboren bat adierazten ote den ebazteko metodo automatikoak bilatzen ditu. "Gu 2011. urtean hasi ginen euskararako lantzen sentimenduen analisisa egiteko teknikak", adierazi du.

EHUko Informatika Fakultateko IXA Taldeak eta Elhuyar Fundazioko Hizkuntza eta Teknologia Unitateak aspalditik duten lankidetzaren emaitza izan da garatutako sistema. Bereziki, Twitter sare sozialean euskara hutsean idatzitako txioak izan dituzte langai, edo tartean euskara dutenak.

Euskarazko testuetan sentimenduen analisisa egiteko sistema sortzeko lehengo pausoa polaritate lexikoak sortzea izan zen, hau da, berez kutsu positiboa edo negatiboa duten hitzen zerrendak osatzea. Adibidez, 'txarra' edo 'gaiztoa' hitzak beti negatiboak dira, eta 'ona' eta 'maitagarria', berriz, positiboak.

"Zerrenda horiek sortzean, ordea, oso kontuan izan behar da zer gai edo testuingurutarako ari zaren lanean, hitz batzuek kontrako polaritatea izan dezaketelako testuinguruaren arabera. Esate baterako, 'gora egin' edo 'behera egin' aditzek ez dute kutsu bera eskailerak igo edo jaisteaz ari bagara, burtsako akzioez ari bagara edo langabeziak ari bagara", azaldu du San Vicentek.

## FENOMENO LINGUISTIKOAK

Lexikoaz gain, hizkuntza orotan gertatzen diren fenomeno linguistikoak ere kontuan hartu behar izan zituzten lanaren egileek: "Esate baterako, ezezko esaldiek kontrako zentzua ematen diete hitzei ('hau ez da batere ona'), edo ironiaz esandako adierazpenak antzematea ere garrantzitsua da. Sor daitezkeen arazoak ebazteko, programa informatiko bat garatu genuen, testuetako informazioa erauzi, hitzak lematizatu eta bestelako azterketak egiteko".

Halaber, Twitterren berezitasunak ere kontuan hartu behar izan zituzten ikertzaileek

Privacidad

hizkera berezia baitago sare sozialetan, ahozko hizkerara asko hurbiltzen dena.

"Gramatika ez estandarra izaten da askotan, eta horrek lana zailtzen die hizkuntza azterketarako tresnei. Gainera, oso maiz gertatzen dira hizkuntzen arteko nahasketak ('Kaixo. Acabo de hacer la azterketa de gizarte. Fatal atera zait!'), eta halako fenomenoak tratatu egin behar dira", esan du San Vicentek.

Informazio guztia ikasketa automatikoko sistemak entrenatzeko erabili zuten ondoren, San Vicentek azaldu duenez: "Milaka adibide sortu genituen, behar bezala sailkatuta, eta sistemari eman genizkion, baita zer ezaugarri erreparatu behar dion erakutsi ere, hortik aurrera bere kabuz egin zezan adierazpenen azterketa".

## BEHAGUNEA

Behagunea izan zen garatutako sistemaren lehenengo aplikazioa, eta 2016. urteko Donostia Europako Kultur Hiriburutzari lotuta Twitterren esan zirenen sentimenduen analisia egin zuten. "Oso ongi funtzionatu zuen; une jakinetan ekitaldi polemikoak egon ziren, eta horiek denak jaso zituen sistemak. Hiriburutzaren barruan gauzatutako proiektuen ebaluazioetan ere erabili ziren bildutako datuak", azaldu du.

Beste proiektu bat Berria egunkariarekin egin zuten, 2016. urteko Euskal Autonomia Erkidegoko (EAE) hauteskunde autonomikoen kanpainaren segimendua egiteko, "eta hor ere ez ziren falta izan polemikak".

Elhuyarreko ikertzaileak EHUko Kriminologiaren Euskal Institutuarekin ere aritu dira lanean, sare sozialetan terrorismoaren biktimei buruzko pertzepzioa nolakoa den aztertzeko. "Guretzat oso garrantzitsuak dira proiektu horiek, tesian oinarritzko ikerketatik industria-aplikazio errealerara bitarteko ibilbidea osorik egin dela erakusten dutelako. Egungo sistema gai da euskarazko testuak ez ezik, gaztelania, frantsesa eta ingelesa ere aztertzeko, hankamotz gelditzen baita euskaraz egindako adierazpenen analisia bakarrik egitea", adierazi du.

Garatutako sistemen emaitzaren balorazioa ona bada ere, oraindik badago hobetzeko tarterik, eta ikertzaileek lanean dihardute: "Ohiko eredu estatistikoetatik algoritmo neuronalekin lan egitera pasatu gara, eta oso emaitza onak ematen ari da. Helburua da adierazpenen sailkapenean arrakasta edo asmatze tasa hobetzea, une honetan %75 ingurukoa baita".

Privacidad