



Ezkerretik eskuinera, German Bordel, Arantza Casillas, Mikel Peñagarikano eta Luis Javier Rodríguez EHUko GTTSko kideak. / SANTOS CIRILO

Hitz-ehiztari informatikoa

Euskarazko, ingelesezko eta gaztelaniar azken-ahots-erregistroak bilatzeko garatu du EHUko Zientzia Fakultateko ikertzaile talde batek

JUAN. A. HERNÁNDEZ
Bilbo

Interneten edo norberaren artxibo digitaletan bilaketak egitea ohikoa da dagoeneko milioika lagunengatik eguneroko bizi-moduaren. Nahikoa da termino egokia tekleatzea bilatzaile komertzial batean, instant batean aurkitzeko aisialdirako zein lanerako behar dugun hori. Jo dezagun, ordea, bilaketa horiek ez daudela mugatuta idatzizko artxiboetara eta giza-ahotsa oinarri duten soinu artxiboak ere arakatu ditzakegun erreminta informatiko baten bidez.

Hain zuzen ere, horrelako lanabes bat sortu dute EHUko Zientzia Fakultateko GTTSko (Software Teknologietako Lan Taldea) ikertzaileek. Euskaraz, ingelesez zein gaztelaniar egin ditzake bilaketak EHzitari izen adierazkorrekin bataiatu duten tresnak.

Teknologia berriak bi motako bilaketak ahalbidetuko di-

tu: Sarean eta norberaren artxibategietan, GTTSko kideek azalzen dutenez. "Hau da, telebista-kate batek erabil dezake, adibidez, bere soinu-artxibategiak arakatzeko eta termino jakin batzuk dituztenak identifikatzeko", azaldu du Mikel Peñagarikano Bilaketa eta Elektrozitate fakultateko irakasle eta EHzitari garatu duen ikertzaile taldeko kideak. Harekin batera German Bordel (taldeko burua), Arantza Casillas, Luis Javier Rodríguez, Amparo Varona eta Maider Zamalloa irakasleek parte hartu dute 2005ean abiatu ziren lanetan.

Urte horretan Industria Sailaren dirulaguntza jaso zuten "erabat jorrratu gabe" zegoen esparruan jarduteko, Bordelek gogoratzen duenez. Gaur egun, ordea, Estatu Batuetan badago antzeko erreminta merkaturatzen ari den enpresa bat; besteak beste *The Boston Globe* egunkariak eskaintzen die tresna hori internautei argitalpena-

ren edizio digitalean. Funtsean, honako funtzionamendua du bilatzaileak: Lehenik eta behin, artxiboak analizatzen ditu audioren bat dutenak identifikatzeko. Behin artxibo horiek zati edo segmentu txikiagotan ba-

Interneten bilaketak egiteko balio dezake erreminta berriak

natuta, ahotsak dituztenak aukeratzeko. Ondoren, ahotsen hizkuntza identifikatzen du eta, jarraian, artxiboa transkribitzen du dagokion hizkuntzaren fonetika eta arau gramatikalak aintzat hartuta. "Hortik aurrera, Google edo Yahoo bezalako bilatzaileen ohiko erremintak erabiltzen dituzte idatzizko testuetan bilaketak egiteko", argitu du Bordelek.

Sistemak logika matematikoa erabiltzen du prozesu horiek guztiak aurrera eramateko, automatikoki ikasitako eredu matematikoak oinarri hartuta. Azken hauek artxiboek dituzten soinuak identifikatzeko balio dute, fonema eta hitz bakoitzak sorrarazten duen uhinarekin frekuentzia formula matematiko baten bidez adieraz baitaitezke. "Guk ez diogu sartu sistemari euskararen edo gaztelaniaren gramatika", argitu du Bordelek. "Prozesuak lagin bidezko ikasketa deritzana du oinarri. Sistemak seinaleak lokalizatzen ditu eta zuk emandako informazioaren arabera identifikatzen ditu. Behin hori ikasita, beste lagin batzutan aplikatu dezake", gaineratu du.

EL PAÍS.COM

► **Lotura interesgarriak:**

gts.ehu.es
(Bilatzailea garatu duen EHUko taldearen ataria)

Ehiztari-erregistroak

► **Erabilpenak:** Sarean edo hedabideetako soinu-artxibategietan bilaketak egiteko balio dezake.

► **Kode irekia:** Software libreko erreminta da, beste ikertzaile batzuek garatzeko modukoa.

► **Euskaraz, zailagoa:** Hizkuntzaren egitura gramatikalak ahalegin teknologiko handiagoa eskatzen du.

Gramatikak eragina du teknologian

Makinek arazo gehigarria dute euskaraz 'ikasteko'. Hizkuntza honen egitura gramatikalak —deklinabideetan oinarritua— zaildu egiten du hitzen identifikazioa, eta horrek eragina du mota honetako erremintak garatzeko behar diren eredu matematikoetan. "Gaztelania eta ingelesa, berriz, errazagoak dira, hitzetan oinarriturik baitaude", dio Peñagarikano.

"Deklinabideek konplikazio tekniko gehigarriak sorrarazten dituzte honelako teknolo-

gietan. Zentzu horretan euskaraz ingelesa baino konplexuagoa da eta nolabaiteko parekotasuna du alemanarekin eta bereziki eskandinaviar hizkuntzekin", gaineratu du ikertzaileak. Antzeko arazoak sortzen dira, besteak beste, itzulpen automatizatorako sistemetan.

Ildo horretan, Europako iparraldeko herrialdeetako zenbait erakunderekin lankidetzan aritzea aztertzen ari da EHUko taldea.

Ehiztari dagokionez, Sau-

trera izeneko teknologia espezi-fikoa garatu du GTTSk. Java hizkuntzan garatua, kode irekiko erreminta da eta hizketa prozesatzeko sistemaren garapenerako sortu dute.

"Gero aplikazioak fabrikatzeko erabiltzen diren programa-multzoak dituzte oinarri teknologia hauek", azaldu du Peñagarikano. Munduan koibete libreak beste bi sistema nagusi daude gaur egun. "Hau hirugarrena izan dadila nahi dugun", diote EHUko ikertzaileek.

Liburuak

Ogibidea eta ofizioa

JON KORTAZAR

Lutxo Egiaren (Santander, 1969) *Zubigilea* nobelak Bilboren aldatuta urbanistikoa hartu du gaitz, Zorrotzaurren egin behar den aldatetaren inguruan sortuz nobelaren matata. Arkitzeko bat dugu narratzaile, baina gaurko eguneko ekintza sakondu egiten da 1964. urtean kokatzen den fikziozko saboteaje. Kontalari horren ahotsean ezagutuko dugu lana. Nobela badira hiru lau ezaugarri begira bat-batean agertzen direnean.

Konstruktibismoa litzateke lehen, hau da egitura eraikitze-ko agertzen den joera nagusia. Kapituluaren izenetik hasi, ibaiaren norabidearen kontran edo zenbakien markak, kapituluak tartekatuz. Izenak eta zenbakiak erabiliz ere maila bitan osatzen den gaia kontatzen du egileak. Batetik, ofizioz zubigile denaren lan mundua eta Bilbo berri baten ikuspegia; bestetik, pertsonaia batzuen bilaketaren inguruan eraikitako matata. Nobela, azken batean, zubia bezala eraikitzen da, eta gainera badira han eta hemen batasun hariak esanahia azpimarratuz.



Bigarrena, *glamour*-aren munduaren deskribapena da. Aspaldiko batean izan zen eskal kontagintzan errealisismoaren kutsua emateko markak aipatzeko eta erabiltzeko joera nagusia. Hemen ere horrela jokatzen da, pertsonaien luxuzko mundua iradokitzeko, eta ustelkeria salatzen.

Hirugarren, bizkaieraren erabilera erraz eta gaiari lotuz ematen da, baina erabilpen hori ez da soilik ukitu dialektologiko berezia. Egiak askotan mantentzen du esanahiaren haria eta deskribapenak moldatzen ditu metafora eta hizkuntza hari bakar bati eutsiz, horrela aipagarriak dira itsasoko hiztegiaren inguruan eratuakotasunak.

Nobelak deskribapen mailan funtzionatzen badu ere, eta era berean esan dezakegu ongi moldatzen dela bere mezu sinbolikoa, ez da gauza bera gertatzen kontaketa mailan. Astiro doaz ekintzak eta sekuentzien arteko lotura ere ez da garbi azaltzen. Futbol atearen azpigaia behin eta berriro agertzen da lehen pasartean eta ez da garbi gertatzen zein den bere lotura ekintza nagusiarekin.

Lutxo Egi: *Zubigilea*. Susa. Zarautz. 2007. 169 or. 14,50 Euro.