

PUBLICIDAD
GAUR8  

07 OCT. 2022

Euskararako eta gaztelaniarako eredu neuronal berriak sortu dituzte, hizkuntza-teknologiei begira

Euskararako eta gaztelaniarako azken belaunaldiko hizkuntza-eredu neuronalak sortu ditu euskal ikerlari talde batek. Euskara bezalako baliabide urriko hizkuntzaren kasuan, «inoiz izan den corpus handiena» osatu dela nabarmendu dute.

GAUR8





Deep Text proiektuan parte hartzen duten ikerlariak, irudian. (UPV/EHU)

Euskararako eta gaztelaniarako azken belaunaldiko hizkuntza-eredu neuronalak sortu dituzte UPV/EHUko HiTZ zentroak eta Oraik eta Vicomtech proiektuek.

Euskararen kasuan, sortutakoak lehen hizkuntza-eredu neuronalak dira. Horiez gain, gaztelaniazkoak eta eredu eleaniztunak ere badaude, euskara, gaztelania, frantsesa eta ingelesa biltzen dituztenak, hain zuzen ere.

HiTZ zentroa buru duen DeepText proiektuaren barruan burutu da lana, eta Euskal Herriko industriaren hizkuntza-teknologiak eraldatzea izan da egitasmoaren beste helburuetako bat.

Zenbait ataza egiteko ahalmena izateko, makinak gure hizkuntza ulertzeko eta sortzeko gai izatea da hizkuntza naturalen prozesamenduaren helburua, eta horretarako, testu-corpus erraldoiak behar dituzte, azaldu dute eredu berrien funtzionamenduz.

Bi hizkuntzen aukeraketaren inguruan, zehaztu dute ekoizpen zientifiko eta teknologiko handiena ingelesaren inguruan murgiltzen dela kontuan hartuta, euskara bezalako baliabide urriko hizkuntzaren garapenean jarri dela fokua.

350 milioi hitzeko euskal corpusa

«Baliabide urriko hizkuntzek arazoak dituzte halako corpus handiak osatzeko, baina proiektu honetan euskarazko inoiz izan den izan den corpus handiena osatu da: 350 milioi hitzeko corpusa», nabarmendu dute.

«Munduan 7.000 hizkuntza inguru daude –azaldu dute ikertzaileek–; gehienak, baliabide urrikoak. Corpus eta material digital gutxi dutenez, zailtasunak dituzte entrenamendu-adibideak sortzeko. Euskara ere multzo horretan sar dezakegu», esplikatu dute hizkuntza-teknologiaren garapenean funtsezkoa den tresnaz.

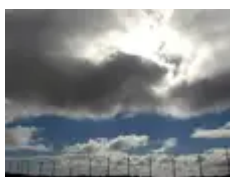
Erantsi dutenez, kasu horietan hizkuntza-eredu eleaniztunak erabiltzea alternatiba eraginkorra da, gai direlako hizkuntza desberdinetako testuak ulertzeko, baita baliabide urrikoetakoak ere.

«Oinarri hori hizkuntza handi bateko adibideekin entrenatzen da (ingeleza, adibidez), eta gero euskarazko datuekin probatzen da ea zer emaitza ematen dituen ikusteko», argitu dute 'transfer learning' deritzon teknikaz.

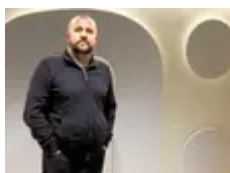
Onartzen dute teknika horrek ez dituela emaitza «perfektuak» ematen, baina «oso emaitza interesgarriak» ematen dituela planteatu dute, «adibidez, galdera-erantzun bidezko bilaketak egiteko».

Ikerlanaren barruan ebaluazio-ingurune bat ere sortu dute, eredu neuronalek hizkuntza ulertzeko duten gaitasuna neurtzeko. Atal horretan, lehena ere sortu dute euskararako, 'BasqueGLUE' izenekoa.

Agian hau ere zure interesekoa izan daiteke



[¿Qué dice Aritz Otxandiano sobre las renovables? Principio de realidad y debate abierto](#)



[Gogora quiere meter a ETA también en el Memorial 3 de Marzo de Gasteiz](#)



PUBLICIDAD



Lo +



Maurice Ravel, konpositore euskalduna, jaioterriaz eta Boleroez harago
