

dimarts, 05 de març de 2024

La UdL, en un projecte europeu que aplica IA al processament de llengües

Transcripció, traducció automàtica neuronal i síntesi de veu del català, occità, basc i aragonès

Desenvolupar coneixement en intel·ligència artificial (IA) sobre nous models de llenguatge generatiu aplicables a les llengües amb pocs recursos dels dos costats dels Pirineus (català, occità, aragonès i basc) és l'objectiu del projecte europeu LINGUATEC IA, on participen investigadors de la Universitat de Lleida (UdL).

Liderat per [Elhuyar](https://www.elhuyar.eus/es) [<https://www.elhuyar.eus/es>] (a través del seu centre d'IA Orai), aquest projecte europeu cofinançat per fons FEDER a través de la convocatòria [Interreg-POCTEFA](https://www.poctefa.eu/) [<https://www.poctefa.eu/>], pretén avançar en la digitalització d'aquestes llengües i en la construcció d'una infraestructura lingüística i intel·ligent transfronterera que faciliti la comunicació entre els seus parlants i l'accés multilingüe a la informació.



El processament del llenguatge és una poderosa eina per a les comunitats lingüístiques amb escassos recursos que ajuda a revitalitzar la llengua i a promoure'n l'ús, per tant és important treballar en intel·ligència artificial, per no quedar-se enrere, ja que la qualitat aconseguida en el processament del llenguatge natural no és a l'abast de totes les llengües, expliquen els investigadors.

És per això que totes les entitats que formen part del projecte (Elhuyar, UdL, Lo Congrès Permanent de la Lenga Occitana, HITZ zentroa (Universitat del País Basc/Euskal Herriko Unibertsitatea), Universitat Jean Jaurès de Tolosa, Universitat de Perpinyà, IKER-CNRS de Baiona i Govern d'Aragó) ja han començat a treballar en diferents àmbits relacionats amb el processament del llenguatge, com ara el desenvolupament de nous algorismes i arquitectures neuronals adaptats a situacions de còmput i recursos lingüístics limitats.

També en la millora dels sistemes de transcripció, traducció automàtica neuronal i síntesi de veu del basc, català, occità, aragonès i les seues variants dialectals que combinin amb el francès i l'espanyol, el desenvolupament d'una plataforma lingüística multilingüe de subtitulació i doblatge automàtic. "A més hem previst crear una plataforma o repositori en línia amb tots els recursos, tecnologies i aplicacions que desenvolupem per a les llengües dels Pirineus", explica Josu Aztiria, coordinador de LINGUATEC.

Pel que fa a la UdL, l'equip integrat per Jordi Suïls, Glòria Vázquez i Helena Torres, s'ocuparà de millorar les eines de traducció automàtica català-aranès de crear un corpus digitalitzat per a l'aranès, tant de llengua escrita com oral, per tal d'ajudar a construir processadors i sintetitzadors de llenguatge oral.

El projecte LINGUATEC IA recull el testimoni de [LINGUATEC](https://linguatec-poctefa.eu/) [<https://linguatec-poctefa.eu/>], un projecte previ cofinançat amb fons POCTEFA, ja finalitzat, en què després de tres anys de camí, van consolidar una xarxa d'excel·lència en intel·ligència artificial per a la construcció d'una infraestructura lingüística transfronterera.

TEXT: Elhuyar/Prensa UdL