

ECIA 2: Processament del llenguatge natural

Description

El projecte El Català i la IA (ECIA) té com objectiu generar els recursos lingüístics necessaris per al català, que permetin i facilitin desenvolupar aplicacions en llenguatge natural com chat bots, traductors automàtics, aplicacions d'extracció d'informació o resum automàtics. La tecnologia de la llengua ha entrat de ple en el món de l'aprenentatge profund i el big data, per tan cal posar a l'abast de la comunitat científica i la indústria models de llengua en català, pre-entrenats sobre grans quantitats de dades lingüístiques de qualitat.

Es produirà també un bon nombre de dades anotades que serviran per entrenar i, posteriorment, avaluar models per a tasques específiques (com ara sistemes de pregunta resposta, de classificació semàntica i d'altres tasques que impliquen comprensió del llenguatge). La generació de dades manualment anotades és un procés crític i molt costós que aquest projecte pretén resoldre.

Finalment, es desenvoluparà una aplicació de benchmarking que permetrà l'avaluació contínua de models i sistemes utilitzant dades anotades. L'objectiu final de l'aplicació és impulsar la investigació en el desenvolupament de sistemes generals i robusts de comprensió del llenguatge natural.

Esperem que el projecte tingui un gran impacte en la indústria de la llengua en català i en els ciutadans ja que, per primera vegada, el català disposarà de recursos suficients per a ser considerat llengua de primera en el món digital i els ciutadans podran fer vida digital en català, sense haver de passar a l'anglès o al castellà com passa actualment. Els objectius del següent projecte són el desenvolupament d'un sistema d'informació d'accés públic que serveixi com a eina d'assistència a la presa de decisions en la gestió de nous brots de COVID-19, tant en l'àmbit d'autoritats sanitàries com de la ciutadania. Tal eina consta de quatre grans components que inclou:

- i) un sistema d'informació geogràfic que integri entre altres fonts, informació de casos reportats, patrons de mobilitat, disponibilitat hospitalària i altres;
- ii) una eina interactiva de visualització i anàlisi, que resumeixi en forma clara l'estat sanitari dels diferents municipis / zones hospitalàries i que identifiqui tant tendències com patrons de contagi significatius;
- iii) implementació de nous sistemes d'anàlisi de dades, incloent-hi un nou sistema de representació i anàlisi de les dades en forma de xarxa, la implementació de models d'anàlisi d'anomalies i tendències en les dades per estudiar la incidència de les mesures de restricció de moviments;
- iv) calibratge de models epidemiològics per a realitzar prediccions a curt termini i per poder simular diversos escenaris i l'efecte de diferents tipus d'intervencions. En conjunt, l'objectiu final és proveir d'un recurs integrat i de lliure accés, que faciliti la gestió de nous futurs brots de COVID-19. Addicionalment, el sistema s'associarà al Portal europeu de COVID organitzat per l'EMBL-EBI i ELIXIR mitjançant el seu sistema de federació de sistemes nacionals, el que contribuirà al seu manteniment i disseminació a mitjà i llarg termini.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación

Source URL (retrieved on 10 Ene 2025 - 21:30): <https://www.bsc.es/es/research-and-development/projects/ecia-2-processament-del-llenguatge-natural>