

umap.eu

Josu Azpillaga, Aitzol Naberan
CodeSyntax
jazpillaga@codesyntax.com, anaberan@codesyntax.com
www.codesyntax.com



GIZARTE-SAREAK

Sarrera

2010eko azaroan, CodeSyntax-ek UMAP proiektua aurkeztu zuen. Umap, Twitterretik mezuak jaso, prozesatu eta informazioa ateratzeko produktu bat da.

Umap-ek instantzia diferenteak izan ditzake, aztertu nahi den komunitatearen arabera. Aztertu dira, orain arte, hizkuntza komunitateak (euskara, katalana eta galesa martxan daude) baina, aurrera begira, bestelako ezaugarriren bat komunean duten erabiltzaileen tuitak filtratu eta handik informazioa jasotzeko helburua ere badu.

eu.umap.eu

Euskaldunok eta euskaraz, Twitterren, zertaz ari gara? Galdera horri erantzuteko nahia du Umap-en euskarazko bertsioak, eu.umap.eu-k hain zuzen ere.

Honako funtzionamendua du Umap-ek:

- Erabiltzaile euskaldun guztien tuitak denbora errealean jaso eta artxibatzeko aplikazioa, Twitterren APIa erabilita.
- Twitterreko erabiltzaile euskaldunak automatikoki detektatzeko metodo bat.
- Erabiltzaile horien mezuetan euskara detektatu eta haren arabera diskriminatzeko metodoa.
- Jasotako mezuetan, termino eta gako/trend esanguratsuenak konputatzeko metodoa.
- Jasotako mezuetako loturak aztertuz, interes gehien duten loturak detektatu eta lotura hoietatik informazioa ekartzeko metodoa.

Hitzaldi honetan puntu horien gaineko informazioa emango da, deskribapen teknikoak eginez eta kode zatiak erakutsiaz.

Neurri apalago batean, umap-ek urte erdian bildu dituen datuei ere begiratu bat emango diegu. Erabiltzaile kopurua, euskararen presentzia, bilakaera... Horretarako ere balio baitezake Umap-ek.

Umap, Python lengoaiari egina dago, Django garapen ingurunea erabilita eta MySql datubasea darabil.