

Europako hizkuntzen baliabide teknologikoak

Definizioa

Hizkuntza bakoitzeko Digital Language Equality (DLE) metrikaren arabeko puntuazioa kalkulatzeko da automatikoki **faktore teknologikoak** aztertuta.

Hizkuntzen Europako Sarean (European Language Grid, ELG) katalogatutako hizkuntza-baliabide eta -teknologiak hartu dira aintzat kalkulu hori egiteko.

Baliabide hauek aztertu dira: tresnak eta zerbitzuak, corpusak, hizkuntza- eta gramatika-eredu konputazionalak, eta baliabide lexikal eta kontzeptualak.

Puntuazioa kalkulatzeko metodoa baliabide moten eta ezaugarrien maiztasun haztatuetan oinarritzen da (adibidez: anatatuta dauden ala ez, espezifikoa edo orokorra diren, erabilera librekoak edo jabetunak diren, eta beste hainbat).

Honela kalkulatu da DLE teknologikoaren puntuazioa hizkuntza bakoitzeko:

1. ELG katalogoko baliabide bakoitzak puntuazio bat lortzen du, bere ezaugarri nagusien pisuen batura, hain zuzen ere. "Ohar-mota" eta "Domeinua" ezaugarriak ponderazio-faktore batekin biderkatu ondoren batzen dira.

Adibidez, demagun ELG katalogoko baliabide batek ezaugarri hauek dituela: testua, corpusa, anotatua, elebakarra, hiru anotazio desberdinekin (morfologia, sintaxia eta semantika), domeinu bati dagokiona (esate baterako, finantza), eta "ikerketarako baimendua" erabilera baldintzekin.

Erabiltzen den metodoak honako puntuazioa esleituko lioke:

$$\text{Score}_{LR1} = 5 + 1 + 2.5 + (3 * 0.25) + 1 + (1 * 0.3) + 3.5 = 14.05$$

2. X hizkuntza jakin baterako puntuazioa kalkulatzeko (TechDLELangX), X hizkuntzaren euskarri diren baliabide guztientzat (LR1, LR2, ... LRN) kalkulatu da beren puntuazioa, eta ondoren guztiak batu. Hau da:

$$\text{TechDLE}_{\text{LangX}} = \text{Score}_{LR1} + \text{Score}_{LR2} + \dots + \text{Score}_{LRN}$$

Hedadura

Europa

Aldagaiak

Hizkuntza:

- Ingelesa
- Alemana
- Gaztelania
- Frantsesa
- Italiara
- Suomiera

- Portugesa
- Suediera
- Errumaniera
- Katalana
- Irlandako gaelera
- Norvegiera
- Euskara
- Galegoa
- Galesa
- Eskoziako gaelikoa
- Britainiera
- Okzitaniera
- Asturiera
- Aragoiera
- Sardiniera

Iturria(k)

European Language Grid

<https://live.european-language-grid.eu/catalogue/dashboard>

Iturriaren ezaugarriak

Europako Batasunak finantzaturako proiektu batetik eratorritako webgune bat da.

Lehenengo datua eta maiztasuna

Lehenengo datua 2023ko abendukoa da.

Datuak sei hilabetero berritzen dira.

Oharrak

Informazio osagarria

Argitalpenak eta txostenak

Egitasmoaren txostenak honako helbidean daude: <https://european-language-equality.eu/deliverables/>

Bereziki kontuan hartzekoa da euskarari buruzko txostena (*D1.4 Report on the Basque Language*): https://european-language-equality.eu/wp-content/uploads/2022/03/ELE_Deliverable_D1_4_Language_Report_Basque_.pdf