

Wikidata, ezagutzarako datu-base libre kolaboratiboa

Denok ezagutzen dugu Wikipedia, mundu osoko erabiltzaileen artean osatuz doan eta Wikimedia fundazioak kudeatu eta sustatzen duen entziklopedia librea. Baina fundazio horrek baditu hain ezagunak ez diren beste proiektu interesgarri asko ere, elkarlanean ezagutza librea osatzea helburu dutenak denak: Commons irudizko eta ikus-entzunezko materialarentzat, Wiktionary hiztegiarentzat, Wikibooks, Wikisource, Wikiversity... Denetan berriena Wikidata da, ezagutzarako datu-base librea. 2012tik existitzen bada ere, azken urteotan loratu da, eta gauza interesgarri asko ekarri ditu.

[Wikidata](#) ezagutzarako datu-base libre eta kolaboratibo bat da. Baina, Wikipedia ez bezala, zeina testuz eta bestelako baliabide grafiko osatutako artikuluen bilduma baita, Wikidata informazio egituratuaren bilduma bat da, eremu gutxi eta labur batzuk dituzten erregistroz osatua. Datu-base horretan, pertsonen jaiotza- eta heriotza-datak eta tokiak gordetzen dira, baita hirien biztanle-kopuruak eta beste datu batzuk ere. Eta erlazioak ere gordetzen dira; adibidez, pertsonen arteko senidetasunak, hiriak zein probintziatakoak diren eta haiek zein estatutakoak, animalia-espezieen erlazio taxonomikoak...

Wikipediarekiko beste ezberdintasun nagusi bat da ez dagoela hizkuntza bakoitzerako bat. Datuak besterik ez direnez, Wikidata bakarra dago, eleaniztuna. Gero, datu-unitate bakoitzak izena eta deskribapena nahi beste hizkuntzatan izan dezake.

Wikidataren egitura

Wikidatan edozein motatako datuak eta haien arteko loturak gordetzen dira. Baina, izatez, hiru datu-mota baino ez daude: elementuak, propietateak eta adierazpenak.

Elementu motako datuek pertsonak, hiriak, abestiak, animalia-espezieak, kontzeptu abstraktuak eta horrelakoak adierazteko balio dute. Horieta

Igor Leturia Azkarate
Informatikaria eta ikertzailea



bakoitzak identifikadore bat dauka Wikidatan, “Q” karaktereak eta zenbaki batek osatua. Adibidez, Q1 elementuak unibertsoa adierazten du, eta <https://www.wikidata.org/wiki/Q1> helbidean atzi daiteke; [Q12256717](#) elementuak Elhuyar anaiei egiten die erreferentzia; [Q47588](#) elementuak Euskal Herriari... Horrez gain, elementu bakoitzak izen edo etiketa bat, deskribapen bat eta hainbat alias edo bestelako izendapen izan ditzake hizkuntza bakoitzeko.

Propietateek, aldiz, elementuek izan ditzaketen informazio- eta erlazio-motak adierazten dituzte. Esaterako, [P31](#) propietatea elementu bat zer motatakoa den adierazteko erabiltzen da, eta propietateak daude jaioteguna adierazteko ([P569](#)), zerbaiten parte dela adierazteko ([P361](#)), egilezarentzat ([P51](#))...

Azkenik, adierazpenek informazioa eransten diete elementuei, propietate baten bidez lotuz balio bati edo beste elementu bati. Esate baterako, ia elementu guztiek daukate adierazpen bat P31 propietatearekin (mota), dagokien motarekin lotzen dituen; pertsona ia guztiek P569 (jaioteguna) adierazpena... Adibidez, adierazpen bat izan daiteke [Q937](#) (Einstein) – P31 (mota) – [Q5](#) (pertsona), edo [Q937](#) (Einstein) – P569 (jaioteguna) – 1879/03/14; hurrenez hurren, adierazten dute Einstein pertsona bat dela eta data horretan jaio zela.

Hiru datu-mota horien konbinazioz, edozeri buruzko informazio guztia jaso daiteke. Gaur egun, Wikidata 7.000 propietate inguru ditu, ia 100 milioi elementu eta 1.400 milioi adierazpen inguru.



WIKIDATA

Informazio lexikografikoa ere bai

Wikidata jatorrian hori besterik ez bazen ere (elementuak, propietateak eta adierazpenak), 2018an datu-mota berri batzuk gehitu zitzaizkion informazio lexikografikoa ere gordetzeko. “L” karakterearekin hasten dira horien identifikadoreak, eta hizkuntzak, hitzak eta kategoriak definitzen dituzte (adibidez, euskarazko “bederatzi” hitza, “izen” kategoriakoa, [L74178](#) da). Lexema batek hainbat forma har ditzake, eta formen datu-mota bat dago, lexemarenari “F”-rekin hasten den beste identifikadore bat gehituta identifikatzen dena. Formaz beraz gain, ezaugarri gramatikalak eta nahi beste adierazpen gorde ditzake. Azkenik, lexemek ere hainbat adiera izan ditzakete, eta, horiek gordetzeko, adieren datu-mota dago.

Egitura hori baliatuta, edozein hizkuntzatarako hiztegiak osa daitezke. Gainera, adierak Wikidatako kontzeptuetara lotuz gero, hizkuntza arteko

erlazioak ere egin daitezke, eta, beraz, edozein hizkuntza-paringen arteko hiztegi elebidunak osa daitezke.

Erabilerak, milaka

Eta zertarako balio lezake horrelako datu-base batek? Zertarako ez! Milaka erabilera-posible eta aukera eskaintzen ditu. Edozein erabiltzailek jaitsi dezake Wikidata, eta nahi duenerako erabili. Web-interfazean bilaketa arruntak egin daitezke, baina, bilaketa ohikoez gain, [SPARQL](#) lengoian [kontsultak egiteko aukera](#) ere ematen du, eta, horren bidez, “ministro baten seme edo alaba diren ministroen kopurua, herrialdeka” eta horren pareko bestelako galdera konplexu eta interesgarriak egin daitezke.

“Egunean Behin joko arrakastatsuko galderak osatzeko ere Wikidata erabiltzen dute”

Eta, APIa baliatuz zein jaitsita, programak egin daitezke informazioa baliatzeko. Esaterako, Wikipedian bertan, gaur egun infotaulak (artikulu batzuen hasieran eskuinean agertzen diren informazio-taulak) ez dira eskuz editatzen, zenbait programa daude idatzita horretarako Wikipediako artikuluetan baliatu daitezkeenak lerro bakarra jarrita. Programak Wikidatatik informazioa hartu, eta taula osatuko du, eta, Wikidatan informazioa aldatu edo eguneratzen bada, artikuluko infotaulan automatikoki agertuko da, eguneraketa aldatzen

ibili beharrik gabe. Infotaulen sistema berri hori Kataluniako [Amical Wikimediaren](#) eta [EWKE Euskal Wikilarien Kultur Elkartearen](#) artean [garatu zuten](#). [Egunean Behin](#) galdera-erantzunen joko arrakastatsuko [galderak osatzeko ere Wikidata erabiltzen](#) dute [CodeSyntax](#) enpresakoek.

Esan dugunez, Wikidata datu-base bakarra dago, eta bertako datuetan hizkuntza guztien informazioa sar daiteke. Hala, euskararentzat euskarazko Wikipedia ahalik eta garatuen egotea ezinbestekoa den gisara, Wikidatan ere euskarazko izen eta deskribapenak eta informazio lexikografikoa egotea oso-oso garrantzitsua da. Elhuyarren, EWKEren enkarguz eta haiekin elkarlanean, bi proiektu eraman ditugu aurrera horretarako. Batetik, 2018an, Wikidatako zientzia eta teknologiako 6.500 elementutan [Elhuyarren Zientzia eta Teknologiaren Hiztegi](#) Entziklopedikoko [definizioak idatzi genituen](#). Bestetik, 2019an [Elhuyarren Ikaslearen Hiztegi](#)ko 10.000 izen erabilienak, bakoitzaren 65 forma eta haien adiera eta definizioak [gehitu genituen](#). Lan horrekin, euskara seigarren hizkuntza bihurtu zen lexema edo hitz-erroen kopuruan, bigarrena hitz-formen kopuruan eta lehena adieren kopuruan. Geroztik egin diren gehikuntzekin, orain lexemetan bederatzigarren gara, 23.000 ingururekin, baina jarraitzen dugu izaten bigarren formetan, 1.250.000 ingururekin, eta lehenenak gara adieretan, ia 31.000ekin.

Proiektu oso interesgarria da, beraz, Wikidata, zeina oso baliagarria baita jada, eta etorkizunean, handitzen joan ahala, are baliagarriagoa izango baita. ●