Euskal musika-taldeak eta albumak Wikidatan

Koldo Etxaniz*, Lorea Loinaz*, Ane Paniagua*, Kepa Sarasola* **, Ana Zelaia* eta Galder Gonzalez**

* Informatika Fakultatea, UPV/EHU

** Euskal Wikilarien Kultur Elkartea (EWKE)

Galdera hauen erantzunak lor daitezke Wikidatako datuekin?

- Euskal musika-talde eta kantarien artean, zeintzuk dira album gehien plazaratu dituztenak?
- Zein urtetan plazaratu dira euskarazko musika-album gehien?
- Nola eboluzionatu du urtez urte euskarazko albumen kopuruak genero artistikoaren arabera? Eta sexu edo generoaren arabera?
- Mikel Laboa, Benito Lertxundi eta Gatiburen albumak ikus ditzakegu urtez urteko ardatz batean tartekatuta?
- Nongoak dira euskal musikariak? Zein herritan jaio dira?

2018an antzeko galderak erantzuteko lan ikusgarri bat aurkeztu zen blog honetan bertan, ("<u>Euskal literaturaren armiarma sarea Wikidatan aztergai</u>"), kasu hartan galderak musikari buruz ez, literaturako liburuei buruz izan ziren. Orain, berdin egin daiteke musika-albumekin?

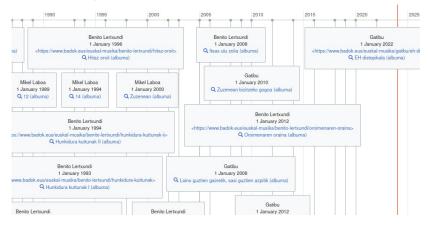
Erantzuna erraza da: "Bai, jakina, datuak edukiz gero egin daiteke". Wikidata oso tresna baliagarria da; kontsulta, azterketa eta grafika oso interesgarriak sor daitezke, baina horrelakorik ezin da lortu daturik ez badago.

Bada, horixe izan da 2023an Donostiako Informatika Fakultateko ikasle batzuek hartu zuten erronka: erantzun horiek lortu ahal izateko Wikidatako datuak osatzea. Aurretik Wikidatan baziren euskal musikari buruzko datu batzuk, hainbat wikilarik azken 10 urteetan modu soltean sartuak. Horiez gain, Musikasten (https://www.musikasten.eus) eta Badok (https://www.badok.eus) atariek horrelako datu batzuk badituztenez, azkenean atari bietan, batean eta bestean, daudenekin lortu da hainbat kantari, musika-talde eta albumen datuekin multzo minimo bat sortzea, eta horrela goiko galderei erantzun itxuroso bat eman ahal zaie. Ez da erantzun osoa, datu guztiak ez baitaude, baina ideia bat sortzeko balio du eta datu gehiago sartuz gero erantzun zehatzagoak lortu ahal izango dira etorkizunean. Bitartean, gainera, lortutako erantzun horiek beste hizkuntza batzuekin ere kontsulta daitezke, eta horrela hizkuntzen arteko konparazio batzuk ere egin daitezke.

Lan hau teknikoki nola egin dugun azaltzeko xehetasunak bukaeran jarri ditugu, atal hori ez da irakurle guztientzat, antzeko lanak egin nahi dituenarentzat baizik. Hor azaldu dugu konputagailuan zer programatu behar izan dugun eta Wikidatan musika-datu horien errepresentazio estandarra nolakoa den. Bestalde, badok.eus ataria askoz konpletoagoa denez, Wikidatako datuetan sistematikoki gehitu ditugu estekak Badokera bideratzeko. Horrela, klik bakar batean joan ahal izango da erabiltzailea informazio sakonagoaren bila.

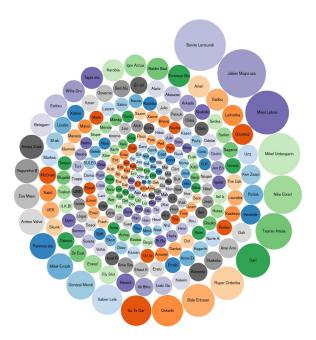
Egindako lana eginda, orain Wikidatan datu multzo minimo bat dugula, galderak egin ditzakegu Wikidata kontsultatzeko webgunearen bitartez (<u>query.wikidata.org</u>). Adibidez:

• Laboa, Lertxundi eta Gatiburen albumak urteen-ardatzean Badokeko erreferentziarekin.

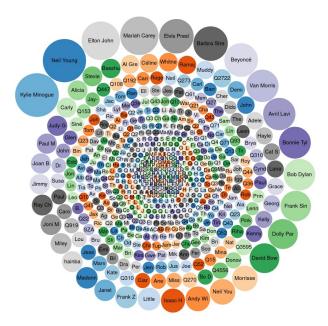


Antzeko galderak dira beste hauek ere:

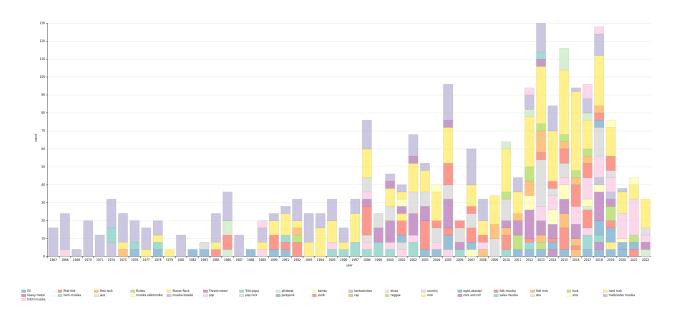
- Exkixu, Alex Sardui eta Gatiburen albumak urteen-ardatzean Badokeko erreferentziarekin
- Berri Txarrak taldearen abesti eta albumak (kontsultan diskografiako propietaterik erabili gabe)
- Gatiburen albumak urtez urteko ardatz batean Badokeko erreferentziarekin (diskografiako propietateak erabiliz)
- Euskarazko musika interpretatzaileak album kopuruaren arabera ordenatuta.(burbuila-diagrama)



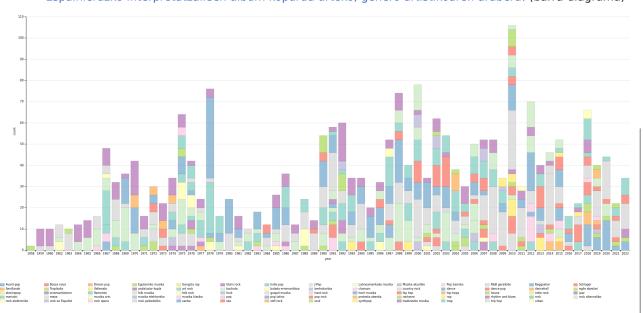
• Ingelesezko musika interpretatzaileak album kopuruaren arabera ordenatuta. (burbuila-diagrama, exekuzioak denbora behar du, asko dira eta)



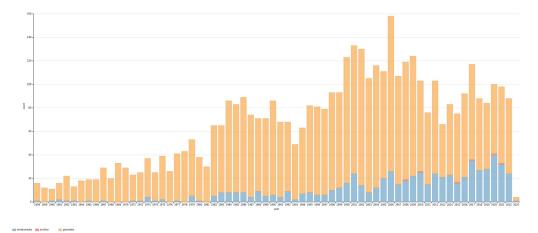
• Euskarazko interpretatzaileen album kopurua urteko, genero artistikoaren arabera. (barra-diagrama)



• Espainierazko interpretatzaileen album kopurua urteko, genero artistikoaren arabera. (barra-diagrama)



• Euskarazko bakarlarien album kopurua urteko, bakarlariaren sexu edo generoaren arabera. (barradiagrama)



Euskal musikarien herriak



• Euskarazko abestiak ezaugarri hauekin: genero artistikoa, musikagilea, noren hitzak eta tonalitatea (gako-armadura).

Agertzen den taulan 156 kantu baino ez dira agertzen, Wikidatan ez baitago informazio asko kantuei buruz.

Egindako ekarpena

Euskal musika-taldeen eta haien albumen zerrenda luzea da. Une honetan badok.eus webgunean 1.350 talde inguru daude. Horietatik 300 talde besterik ez ditugu landu, Wikidatan musikastenID identifikadorea dutenak. Horrela gure lana pisu handieneko interpretatzaileekin bakarrik egin dugu, 300 talde horien Wikidata itemak eta haien albumak (garrantzitsuak diren batzuk, gutxienez) modu koherentean sortzen saiatu gara.

Proiektua amaitu dugun unean, guztira 300 interpretatzaileren eta 1144 albumen erregistroak daude Wikidatan. Denek ez dute informazio bera gordetzen, batzuek informazio gehiago dute, baina garrantzitsuena nabarmentzearren, honakoa esan dezakegu:

- 1. Egitura estandar bati jarraituz sortuak izan dira. Horri esker, euskal musikarako sortu ditugun SPARQL kontsultak erraz molda daitezke beste hizkuntzekin ere erabiltzeko, frantsesezko albumak bilatzeko, esaterako.
- 2. Album guztiek dute informazio minimo bat, garrantzitsuena dena (interpretatzailea, diskoaren izenburua, genero artistikoa, urtea, hizkuntza eta badok.eus-erako esteka).

3. Denek dute badok.eus webgunerako esteka eta horietako batzuek Musikasten webgunerakoa ere. Wikidatan zeudenei eta esteka ez zutenei gehitu zaie. Horri esker, edozein erabiltzaile klik bakar batean joan ahal izango da zuzenean Badok atarira informazio sakonagoaren bila.

Frabilitako tresna informatikoak

Artikulu honetan aurkezten dugun lana aurrera eramateko erabili behar izan ditugun tresna guztiak aipatzen ditugu atal honetan. Izan ere, jakin badakigu proiektu honek jarraipena izan dezakeela, eta hobekuntzarekin aurrera egiteko prest ager daitezkeen ikasleei ondo etorriko zaiela erabilitako tresnen informazioa izatea.

Pywikibot:

https://www.mediawiki.org/wiki/Manual:Pywikibot

Pywikibot/Wikidata:

https://www.mediawiki.org/wiki/Manual:Pywikibot/Wikidata

Python eta pywikipedia online erabiltzeko aukera:

https://hub-paws.wmcloud.org

OpenRefine:

https://openrefine.org/

SPAROL kontsulta-zerbitzua

https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/Wikidata_Query_Help/eu Wikidata Query Service

https://query.wikidata.org/

Tresna informatiko horiek berriak ziren Koldo, Lorea eta Ane ikasleentzat proiektuan lanean hasi ziren unean. Sekula ez zuten Python programazio-lengoaiarekin lan egiteko aukerarik izan, ez zekiten Wikidata nola antolatuta zegoen, ezta Wikipediaren eta Wikidataren arteko lotura nola egiten zen. Proiektu honi esker ikasi dute hori guztia.

Egindako programak eta horien dokumentazioa Aneren, Koldoren eta Lorearen Github orrian daude: https://github.com/koldosaurio/WikiBadok

Lan honi jarraipena emateko aukera batzuk

- 1. Esan bezala, Euskal musikan erreferentzia nagusi den badok.eus atarian 1.350 talde inguru daude guztira, eta horietatik 300 talde besterik ez ditugu landu. Musika-talde eta album gehiago gehitu litezke. Badok atarikoak interesatuta baleude, beren informazio aberatsera kontsultak bideratzeko lagungarri ikusiko balute, eurekin lankidetzan egin liteke.
- 2. Euskal musikarientzat sortu ditugun Wikidata erregistroak Wikipediako artikuluekin lotu litezke masiboki, esate baterako, <u>Hertzainak</u> taldekoa dagoen bezala.
- 3. Musika-taldeak eta album berriak sortuko dira etorkizunean. Horien datuak gehitu egin beharko dira Wikidatan. Wikilari boluntarioek egin dezakete lan hori banaka-banaka eskuz sartuz, edo lan hori errazteko eta datuak eguneratuta edukitzeko laguntza informatikoak sor litezke.

DESKRIBAPEN TEKNIKOA

<u>Donostiako Informatika Fakultateko</u> hiru ikasle aritu dira lanean, irakasleen eta wikilarien laguntzarekin, 2023ko martxotik abendura bitartean.

Wikidatan euskal musikako oso album gutxi zeuden. Erregistro berriak sortu behar izan dira, formatu estandarizatuari jarraituz. Igotako informazioari esker, bilaketa eta analisi sakonagoak egiteko aukera zabaldu da, Wikipediako musika-albumen artikuluetan (oraindik gutxi daude) infotaulen bidez edo grafikoen bidez erakutsi daitekeena.

Txosten honen lehenengo atalean Wikidatan aldez aurretik zegoena azaltzen da. Bigarrenean datu berriak lortzeko iturriak deskribatzen dira. Hirugarrenean, informazioa Wikidatara igotzeko prozeduraren xehetasunak ematen dira. Eta, azken atalean, erabili ditugun tresna informatikoak eta egiteke geratu zaiguna azaltzen da.

Wikidatan euskal musika-taldeak eta albumak bilatzea

Proiektuarekin hasi ginen unean, euskal musika-talde eta album batzuk bazeuden Wikidatan, adibidez, <u>Hertzainak</u> taldea eta <u>Hertzainak (albuma)</u> 2013tik daude. Gure lehenengo lana aldez aurretik bazeuden talde eta album horiek aurkitu eta zerrenda bat osatzea izan da, Wikidatan haien erregistroak ez bikoizteko. Bilaketa egiteko modu desberdinak aplikatu ditugu. "Musikasten ID" edo "Badok datubaseko kode" propietateak bazeuden definituta euskal musikako hainbat taldetarako, atari horietara heldu ahal izateko identifikadore batekin.

Informazioa lortzeko iturriak

Euskal musika-taldeen izenak eta albumak lortzeko bi webgunetara jo dugu: Badok eta Musikasten.

Atari horietatik datu gutxi batzuk besterik ez ditugu hartu Wikidatara eramateko: musika-taldearen izena, albumaren izenburua, urtea, musika-generoa eta herria. Informazio hori .csv formatuan gorde dugu fitxategi batean, gerora horrekin lan egiteko. Hona hemen adibide batzuk (musika-taldea, diskoaren izenburua, badok.eus-erako esteka):

Gailu, "Begiak, egiak ba ote?", https://www.badok.eus/euskal-musika/gailu
121 Krew, "Mahatsen denbora / Azken afaria", https://www.badok.eus/euskal-musika/121-krew
2 Kate, "Birziklatu", https://www.badok.eus/euskal-musika/2-kate

2 Princeses Barbudes,"Enaren kontuak", https://www.badok.eus/euskal-musika/2 Princeses
Barbudes

Izan ere, gure proiektuaren helburua Wikidatara musika taldeen eta albumen informazio minimoa eramatea izan da, datu horiekin oinarrizko estatistikak eta grafikoak egin ahal izateko. Gainera, webgune horietarako estekak gehitu ditugu, musika-taldeei edo albumei buruzko informazio gehiago nahi duen oro euskal musikari buruzko webgune horietara bideratzeko.

<u>Badok</u> (https://www.badok.eus/) Berria egunkariaren musikaren ataria da. Plataforma 2009an sortu zen, eta euskal kantagintzari buruzko informazio aberatsa eskaintzen du: biografiak, diskografiaren fitxak, kantu-hitzen bilduma bat, eta kantu askoren audioarekin.". Wikidatan <a href="https://example.google.com/Badok

Musikasten (https://www.musikasten.eus/) euskal musikaren beste datu-base bat da.

Wikidatan <u>Musikasten ID</u> identifikatzailea sortuta dago eta identifikatzaile horren bidez lotura bat eraikitzen da Wikidata eta datu-basearen artean, adibidez, <u>A Tuti Plain</u> musika-taldea. Talde horrek Wikidatan ez zeukan baliorik Badok datubaseko propietatean eta horrelakoetan gehitu egin diogu, bietara lotuta gera dadin. Talde horrek Wikipedia artikulua badauka, motza bada ere, eta lotura egina dago.

Talde eta album berriak Wikidatara igotzea

Album berriak ez ditugu banaka-banaka eskuz igo Wikidatara, programa batzuk garatu ditugu erregistro berriak modu automatikoan sortzeko. Proiektua hasi dugun unean musika-talde eta albumak baziren Wikidatan. Horiek informazio gehiagorekin osatu ditugu.

Probak egin ondoren, Wikidatan erregistro berriak sortzeko programa martxan jarri dugu eta horrela sortu dugu, adibidez, <u>Ispiluaren aurrean (albuma)</u> item-a (Q117319566). Izenean "(albuma)" bukaera gehitzea erabaki dugu, anbiguotasunak ekiditeko. Erregistroaren deskribapenean Afrika Bibang-en albuma dela adierazi dugu hainbat hizkuntzatan (euskaraz, ingelesez eta gaztelaniaz).



Ondoren, honako adierazpenak gehitu ditugu: album bat dela adierazteko "honako hau da" ezaugarria, Afrika Bibang-en albumetako bat dela adierazteko "honen parte da", soul generoko musika dela adierazteko "genero artistikoa", kantaria zein den adierazteko "interpretatzailea", euskarazko sorkuntza-lana dela adierazteko "lanaren edo izenaren hizkuntza" eta "argitaratzedata". Amaitzeko, identifikatzaileen atalean "badok datubaseko kode" ezaugarrian albumaren badok-eko URLa jarri dugu. Gehitu dugun adierazpen horietako bakoitzean, ezaugarrien erreferentzia moduan, badok.eus webgunerako URL-a gehitu dugu, bai kantariarena eta baita albumarena ere. Izan ere, trafikoa badok.eus webgunera bideratzea da proiektuaren helburu nagusietako bat.

Album solteen item-ak sortzeaz gain, interpretatzaile batek sortutako album guztien bilduma jasotzen duten item-ak ere sortu ditugu, adibidez, <u>Hertzainak-en albumak</u>, <u>Hertzainak-en diskografia</u> erregistroarekin lotuta dagoena.

Adierazpenak



Musika-taldeen erregistroak honako ezaugarriekin sortu ditugu. <u>AKts musika-taldeak</u>, adibidez, deskribapena du hainbat hizkuntzatan.



Ondoren, albumak bezala, "<u>honako hau da</u>", "<u>genero artistikoa</u>" eta "<u>badok datubaseko kode</u>" ezaugarrien bidez adierazten dira musika-taldea, musika-generoa eta badok.eus webgunerako esteka. Horiez gain, "<u>diskografia</u>" ezaugarria ere badauka.

Adierazpenak



Wikidatako musika-talde batetik abiatuta, haren albumak "diskografia" propietatearen bidez lotzen dira.

Balioa <u>Wikimedia artist discography</u> motakoa izan behar da, eta hor barruan hainbat lista defini daitezke irizpide desberdinen arabera. Adibidez Michael Jacksonen diskografian sei lista daude:

Michael Jackson's albums in chronological order

Michael Jackson compilation albums discography

Michael Jackson singles discography

Michael Jackson music sales certifications

List of songs recorded by Michael Jackson

Michael Jackson videography

Gure kasuan, eta hasieran batez ere, gehienetan lista bakarra jarriko dugu hor: album-lista (<u>Wikimedia albums discography</u> motakoa). Batzuetan bi lista izan daitezke: albumak eta singleak.

Adibidez: 121 Krew taldearen (Q12252420) diskografian (121 Krew(r)en diskografia) lista bakarra azaltzen da. Baina egon litezke beste lista batzuk: singleak edo bideoak, adibidez.

121 Krew(r)en diskografia (Q123358964)



Eta, album-lista bakar horrek (Q123358970), <u>Wikimedia albums discography</u> motakoa 121 Krew taldearen lau albumak hartzen ditu: <u>Mahatsen denbora / Azken afaria (albuma)</u>, <u>Hiriaren taupadak (albuma)</u>, <u>Guztiok bat... euskal rap! Mixtape Vol.1 (albuma)</u>, eta <u>El surko del tiempo / Denboraren ildoa (albuma)</u>.

+ gehitu adierazpena

Adierazpenak

honako hau da	 ₩ikimedia albums discography ingelesa v 0 erreferentzia 	→ aldatu+ Erreferentzia gehitu
		+ balioa gehitu
honen parte da	121 Krew(r)en diskografia • 0 erreferentzia	∕ aldatu
	, o erreterentzia	+ Erreferentzia gehitu
		+ balioa gehitu
osatuta	Mahatsen denbora / Azken afaria (albuma)	♪ aldatu
	→ Erreferentzia bat	
	Hiriaren taupadak (albuma)	
	• Erreferentzia bat	
	Guztiok bat euskal rap! Mixtape Vol.1 (albuma)	∕ aldatu
	→ Erreferentzia bat	
	El surko del tiempo / Denboraren ildoa (albuma)	∕ aldatu
	→ Erreferentzia bat	
		+ balioa gehitu
honen zerrenda	album	
	▼ 0 erreferentzia	+ Erreferentzia gehitu

Zenbait kasutan loturak egitea ez da erraza izan, album batzuk hainbat talderen artean sortuak izan direlako, edo badirelako hainbat taldetan ibili diren musikariak edo talde batekoa izanik bakarlari moduan album batean kolaboratu dutenak.

Etorkizunean, Wikipedian artikulua duten musika-taldeen Wikidata item-a artikuluarekin lotu daitezke, <u>Hertzainak</u> taldean bezala. Hala ere, egin dugun lan honetan Wikidatako erregistroak sortzera mugatu gara.

Euskal musika-taldeen eta haien albumen zerrenda luzea da. Une honetan badok.eus webgunean 1.350 talde inguru daude. Horietatik 300 talde besterik ez dugu landu, Wikidatan musikastenID identifikadorea dutenak. Horrela gure lana pisu handieneko interpretatzaileekin bakarrik egin dugu, 300 talde horien Wikidata itemak eta haien albumak (garrantzitsuak diren batzuk, gutxienez) modu koherentean sortzen saiatu gara.

Proiektua amaitu dugun unean , guztira 300 interpretatzaileren eta 1144 albumen erregistroak daude Wikidatan. Denek ez dute informazio bera gordetzen, batzuek informazio gehiago dute, baina garrantzitsuena nabarmentzearren, honakoa esan dezakegu:

- 4. Egitura estandar bati jarraituz sortuak izan dira. Horri esker, euskal musikarako sortu ditugun SPARQL kontsultek erraz moldatu daitezke beste hizkuntzekin ere erabiltzeko, frantsesezko albumak bilatzeko, esaterako.
- 5. Album guztiek dute informazio minimo bat, garrantzitsuena dena (musika-taldea, diskoaren izenburua, badok.eus-erako esteka, eta urtea).
- 6. Denek dute badok.eus webgunerako esteka eta horietakok batzuk musikasten webgunerakoa ere. Wikidatan zeudenei eta esteka ez zutenei gehitu zaie. Horri esker, edozein erabiltzaile klik bakar batean joan ahal izango da erabiltzailea informazio sakonagoaren bila.

Grafikoak eta analisiak

Behin item-ak Wikidatan sortu ditugunean, informazio horretan oinarrituz hainbat analisi mota egin ditugu. Wikidataren SPARQL kontsulta-zerbitzua erabiliz, oso ikusgarriak diren grafikoak egin daitezke kontsultan lortutako emaitzekin. Gainera, kontuan izan behar da emaitzak dinamikoak direla, hau da, aldatuz doazela unean Wikidatan dagoen informazioaren arabera. Horrek esan nahi du, euskal musika-taldeak eta albumak Wikidatan gehituz joango diren neurrian, uneoro informazio eguneratua jasoko dela egindako kontsultetatik.

Kontsulta horien emaitzak Wikipediako <u>Euskal Herriko musika</u> artikuluan txerta daitezke grafiko moduan, adibidez. Horrela, modu bisual eta ikusgarrian erakutsi ahal izango genuke Wikidatan txertatutako informazioa.