

## EUSKALTERM ETA TERMINOLOGIAKO DATU-BASEAK.

Datu-base terminologikoak eta datu-banku terminologikoak: ezaugarriak eta desberdintasunak.

Miel Loinaz Beristain

UZEI

### SARRERA

Hitzaldiaren titulua bera da kontu hauei buruz dauzkagun zenbait iluntasun eta zalantzaren adibide. Beraz, gaiari hasieratik heltzeko, bi gauza izango dira gaur nire mintzagai:

- a) datu-base terminologikoak
- b) datu-banku terminologikoak

Biak dira terminologiako informazioa antolatzekeo egiturak, baina oso desberdinak dira beren helburuan, antolamenduan eta tamainan. Egia esan, gaur datu-bankuei buruz aritu naiz gehienbat, eta datu-baseak haien konparazioko aitzakia besterik ez dira izango.

6.5

DATABASE

collection of data organized according to preestablished structure.

6.6

TERMINOLOGICAL DATABASE

database containing terminological data.

6.7

DATA BANK

collection of databases, including the organizational framework for recording, processing, disseminating data.

6.8

TERM BANK

TERMINOLOGICAL DATA BANK

data bank containing terminological data.

*ISO 1087 Terminology work. Vocabulary (2000-03-01)*

## 1. DATU-BASE ETA DATU-BANKU TERMINOLOGIKOEN EZAUGARRI KOMUNAK

Informazio terminologikoa kudeatzeko datu-baseek eta datu-bankuek ezaugarri komunak dituzte, tratatzen dituzten datu-motei eta datu horiek antolatzeke erabiltzen diren fitxen egiturari dagokionez.

Irizpide nagusia: Fitxa terminologiko berean lantzen dira kontzeptu bati dagozkion informazio guztiak.

### 1.1. DATU-KATEGORIAK

Fitxa terminologikoan erabiltzen diren datu-motak ISO 12620 arauan daude zehaztuta, eta ondoren azaltzen dira garrantzizkoenak.

#### TERMINOA

- Termino-sarrerak.

#### TERMINOARI BURUZKO INFORMAZIOA

- Termino-tipoa: besteak beste, forma nagusia, sinonimoa, forma laburra, forma luzea.
- Oharra: esanahia argitzeko azalpen laburra.
- Informazio gramatikala: kategoria gramatikala, generoa, numeroa.
- Erabilera geografikoa.
- Iturria: terminoa aurkitu den dokumentuaren erreferentzia.
- Testuingurua: terminoa agertzen den esaldia, nonbaitetik hartua.
- Ponderazioa: terminoaren normalizazio-egoerari edo erabilgarritasunari dagokion marka.

#### KONTZEPTUAREN DESKRIBAPENARI BURUZKO INFORMAZIOA

- Sailkapena: gaiaren eremu-zuhaitzaren arabera.
- Definizioa: esanahiaren adierazpen formala.
- Azalpena: kontzeptuaren adierazpen informala.

#### INFORMAZIO ADMINISTRATIBOA

- Kontzeptu-zenbakia.
- Fitxaren arduraduna.
- Azken aldaketaren data.

## 1.2. FITXA-EGITURA

Informazio terminologikoa ezin da behar bezala kudeatu Access bezalako datu-base klasikoekin taula soil baten bidez. Kalkulu-orriak ere ez dira egokiak. Horrela landutako hiztegieta, maiz, sinonimoak hainbat erregistro desberdinetan banatzen dira.

Fitxa terminologikoan, datuen arteko egitura zuhaizkara da, eta, beraz, hura irudikatzeko datu-base hierarkikoa edo erlazionala behar da. Posible da, esate baterako, Access-ekin egitura erlazionalako datu-base bat eraikitzea, baina programazio-lan dezentea eskatuko luke. Horregatik, bereziki terminologiarako pentsatutako datu-base kudeatzaileak gomendatzen dira.

Hona hemen terminologia kudeatzeko tresna informatiko bati eskatu behar zaizkion ezaugarri nagusiak:

- Fitxa terminologikoa norberak definitzeko aukera ematea. Horretarako, ISO 12620 datu-kategoriak kudeatzeko gai izan behar du aplikazioak. Datu-egitura hori hierarkikoa da neurri batean.
- Eremuen luzerari mugarik ez jartzea.
- Eremuak *errepikagarriak* izatea. Esate baterako, terminoaren eremua errepikagarria izatea, sinonimia kudeatu ahal izateko. Terminoaren iturriaren eremuak ere errepikagarria izan behar du, iturri bat baino gehiago jaso ahal izateko.

## 2. DESBERDINTASUNAK

### 2.1. HELBURUA

Helburu desberdinak dituzte bi egiturek:

- Datu-base terminologikoen helburua hiztegi tekniko bat lantzea izaten da. Hau da, arlo konkretu bateko terminoak, hizkuntza batean edo gehiagotan, bildu, sistematizatu eta txukun-txukun ordenatuta prestatu.

Hiztegi terminologiko bat lantzeko prozesuan eraikitzen dira, eta hiztegia amaitu ondoren, normalean, datu-base hauek ez dira sistematikoki eguneratzen.

- Datu-banku terminologikoez helburu zabalagoak dituzte:
  - Arlo askotako terminoak bildu eta sistematizatzea.
  - Termino horiek garaian garaiko normalizazio-irizpideen arabera eguneratuta mantentzea.
  - Zientzia edo teknikako arlo landu berrietako terminoak bankuan gehitzea.
  - Bankuaren edukia kontsulta-sistema irekien bidez kanpora eskaintzea.

Alde honetatik, iraunkortasuna eta publikotasuna haien ezaugarri garrantzizkoak dira.

Gainera, badaude xede orokorreko termino-bankuak, Euskalterm edo Termium bezalakoak, eta xede partikularragoa dutenak, arlo konkretu bateko terminologia lantzen dutenak. Adibidez, HAEE/IVAPen termino-bankua, Zuzenbideko eta Administrazioako terminologiaz arduratzen dena.

## 2.2. AZPIEGITURA ANTOLATUA

Ezaugarri hau da, ISO definizioaren arabera, termino-bankuen bereizgarria. Terminologiako datu-bankuak, goian aipatutako helburuak betetzeko, azpiegitura antolatu iraunkor bat behar du: terminologoak, dokumentazio-arduraduna, informatikaria...

Antolaketa horri nahi den izena eman diezaiokegu, terminologia zentroa, edo beste bat. Publikoa edo juridikoki pribatua izan daiteke. Edozein modutan ere, euskararen kasuan argi dago diru publikorik gabe ezinezkoa dela honelako antolaketa bat mantentzea (geroago, EUSKALTERMi buruz aritzean, helduko diot berriro puntu honi).

## 2.3. TAMAINA ETA EUSKARRIA

Bolumena da bi egituren arteko beste desberdintasun nabaria.

- Terminologiako datu-baseetan lantzen den fitxa-kopuruari dagokionez, alde handiak egon daitezke batetik bestera, baina, oro har, esan daiteke, 50 fitxako hiztegi txiki batetik hasi eta 2.000 kontzeptuko hiztegi mardul baterainokoa izan daitekeela. Kasu berezietan, izan daitezke handiagoak ere, 5.000 edo 10.000koak. Honelako bolumenak kudeatzeko mahai gaineko ordenagailuetan korritzen duten aplikazioak erabiltzen dira, datu-baseak kudeatzeko sistemak, terminologiarako bereziak normalean.
- Badira zenbait aplikazio merkatuan, honelako datu-baseak kudeatzeko bereziki eginak. Gure artean ezagunena Trados enpresaren MultiTerm dela uste dut.

Donostiako Nazioarteko Terminologia Biltzarrean aurkeztu ziren beste batzuk ere. Nire iritzirako, Multiterm, nahiz eta oso mugatua izan, bera da erabiltzeko errazena, eta malguena.

- Terminologiako datu-bankuak askoz handiagoak izaten dira. Txikiek, 100.000 erregistro izan ditzakete, eta milioi bat baino gehiagokoak ere badira. Fitxa-egitura konplikatuagoa da eta ordenagailu eta datu-baseak kudeatzeko programa ahaltzuagoak behar izaten dira. Ordenagailu *profesionalagoak*, Unix edo gisa bereko sistema eragilea dutenak, eta datu-base kudeatzaile azkarrak, adibidez, Oracle edo antzeko maila dutenak.

#### 2.4. DOKUMENTAZIO-EUSKARRIA

Seguruena hau izango da egitura bien arteko diferentziarik handiena ezartzen duen ezaugarria.

- Terminologiako datu-base bat eraikitzeke ez da behar dokumentazio aldetik baliabide handiegirik. Proiektuan erabiltzen diren obren bibliografia kontrolatu behar da, eta argitalpena prestatzean, han txertatu...
- Terminologiako datu-bankuan ezinbestekoa da datu-base dokumental bat ongi antolatuta izatea. Bertan egoten dira jasota terminoen egiaztagarritasuna posible egiten duten erreferentzia bibliografikoak. Fitxa terminologikoan, termino baten, edo testuinguru baten, jatorria adierazten duen kode bat ezartzen da, eta datu-base dokumentalean kode horren esanahia zehazten da erreferentzia bibliografiko baten bitartez: autoreak, titulua, argitaletxea, urtea...

Terminoen iturriak argitaratzeari buruz, arazoak sor daitezke obra argitaratuen autore-eskubideekin eta barne-dokumentuen konfidentziasunarekin.

#### 2.5. EREMUEK BALIDAGARRITASUNA

Zenbait eremutako edukia balidatu beharra izaten da, eduki posible guztiak aurrez zerrenda batean zehaztuta egon behar dutelako. Balio posibleen zerrenda hori txikia denean, posible da egiaztapena MultiTerm bezalako tresna batekin egitea, atributu izeneko eremuen bidez.

Balidazio-zerrendak, gutxienez bi eremuko lerroak izaten ditu: kodea eta azalpena.

Termino-bankuetan balidazio-zerrenda horiek oso luzeak izaten dira batzuetan. Eta beste batzuetan, balidazioa baldintzazkoa da, hizkuntzen arabera egin beharrekoa. Ondoren aipatzen dira balidatu beharreko eremu batzuk.

#### 2.5.1. ITURRI-KODEAK

Hainbat milaka dokumenturen erreferentziak aurki daitezke edozein termino-bankutan. Erreferentzia horien kudeaketa datu-base dokumental baten bidez burutzen da, eta ezin zaio termino bati dokumentu baten erreferentzia ezarri, aurrez datu-base dokumentalean dokumentu horri alta eman ez bazaio.

#### 2.5.2. GAIAREN SAILKAPENA

Fitxa terminologiko guztiek izan behar dute adierazita nahitaez gaiaren sailkapena. Hemen ere kodea eta azalpena adierazi behar dira. Ehunka batzuetatik hasi eta milatik gora izan daitezke.

#### 2.5.3. INFORMAZIO GRAMATIKALA. ERABILERA GEOGRAFIKOA

Zerrenda laburrak dira, baina hizkuntza bakoitzean balio desberdinak dituzte, eta, beraz, balidazioa hizkuntzaka egin behar da. Esate baterako, erabilera geografikoa izeneko eremuan, euskarari dagokionez, forma bat *hegoaldekoa* edo *iparraldekoa* dela esan daiteke, baina frantsesari buruz, *Kanadakoa* edo *Frantziakoa* den zehazten da.

#### 2.6. KONTZEPTU-SISTEMAKO ERLAZIOAK EZARRI AHAL IZATEA

Ezagutzaren ingeniartzan, edo hizkuntza naturalaren tratamenduan (adibidez, WordNet bezalako proiektuen gisako ontologiak eraikitzen laguntzeko) ezinbestekoa da kontzeptuen arteko erlazioak fitxa terminologikoetan adierazi ahal izatea. Erlazio hauek hainbat eratakoak izan daitezke:

- a) Erlazio hierarkikoak. Kontzeptu superordinatuak eta subordinatuak. Beste modu batera esanda, generikoa-espezifikoa da seguruenik kontzeptuak sailkatzeko gehien erabiltzen den erlazioa.

Zatia-osoan erlazioan, kontzeptu konprehentsiboa (osoa) da superordinatua, eta partitiboa (zattia) da subordinatua. Kontzeptuak koordinatuak direla esaten da, hierarkia maila berekoak direnean.

- b) Erlazio ez-hierarkikoen artean hainbat daude: kausa-efektua, prozesua-produktua, objektua-operazioa...

## 2.7. TERMINOEN ARTEAN *THESAURUS* ERAKO ERLAZIO SEMANTIKOAK EZARRI AHAL IZATEA

Dokumentazio-sistemetan, informazioa berreskuratzeko egiten diren indexazio eta sailkapenetan, termino normalizatuen zerrendak baliatzen dira. Terminoak semantikoki elkarrekin lotzen dira, ondorengo erlazioen bitartez:

- a) baliokidetasuna: *sarrera nagusia, sinonimoa*
- b) subordinazioa: *termino generikoa, termino espezifikoa*
- c) asoziazioa: *termino erlazionatua*

## 3. EUSKALTERM DATU-BANKU TERMINOLOGIKOA ETA HURA MANTENTZEA

### 3.1. AURREKARIAK

EUSKALTERM ez da banku handi bat, ez oso txikia ere. Gutxi gorabehera, ia 100.000 fitxa ditu, eta 400.000 termino pasa. Datu-bankua 1987an sortu zen, UZEIk egindako hiztegi guztiak fusionatuz. Ondorengo urte guztietan handituz eta eguneratuz joan da, erakunde publikoen diru-laguntzari esker.

Beste kontsulta-sistema batzuk saiatu ondoren, 1999an Internetetik eskaini zen, UZEIren web gunean, doako kontsulta-sistema baten bidez. Zerbitzu honek arrakasta ederra izan du, hilean 30.000 termino-kontsulta egiten zirelako 2000-2001 urteetan.

2001ean, UZEIk Eusko Jaurlaritzari egindako lagapenaren bidez, Terminologia Banku publiko bilakatu da, eta haren kontsulta [euskadi.net](http://euskadi.net) gunetik eskaintzen da.

### 3.2. EGUNERATZEA

Termino-bankuaren eguneratzea hiru modutara egiten da: a) termino berriak gehituz; b) lehengo terminoen formak eguneratuz; c) une bakoitzeko normalizazio-egoerak eskatzen duen termino egokia proposatuz.

- a) Termino berriak gehitzea

Erreferentzia-puntu bat izateko, 1996an Hiztegi terminologikoak 70.000 fitxa zituen.

Azken 6 urteetan 22.000 fitxa gehitu dira. Ondoren zehazten dira, 1998tik hona bitartean EUSKALTERM bankuari egin zaizkion ekarpen garrantzitsuenak.



1998

- *Arrantza hiztegia* (990 fitxa).
- *Gizarte-lanaren hiztegia* (argitaratu gabea, 483 fitxa).
- *Bazarra, ehunkiak, oinetakoak eta drogeria hiztegia* (1.182 fitxa).
- Biokimikako 510 fitxa.
- Psikologia hiztegiako 1.250 fitxa eguneratu dira (Psikologia Fakultatearekin lankidetzan adostuak).
- Aseguruen arloko 698 fitxa.
- *Txirrindularitza hiztegia* (351 fitxa).

1999

- *Administrazio hiztegia* (HAEE/IVAP), Funtzio Publikoari dagokion behin-behineko bertsioa integratu da, lehendik zeuden fitxak eguneratuz.
- *Ekonomia eta Finantzak* hiztegi berria integratu da, lehendik zeuden fitxak eguneratuz.
- Justizia Saileko Errotulazioa (Justizia Saileko idazki-ereduak normalizatzeko Batzordearen zerrenda).
- *Lan Harremanen hiztegia* (808 fitxa berri).
- *Etxetresna Elektrikoen hiztegia* (437 fitxa).
- Arau-hausteen zerrenda (Justizia Saileko idazki-ereduak normalizatzeko Batzordearena).

2000

- HAEE/IVAPek egindako *Zirkulazio Hiztegia* integratu da, UZEIko Zirkulazio hiztegi zaharrarekin bateratuz.
- Fisikan Elektrizitateko arloa berrikusi da.
- *Administrazio Sanitarioko hiztegia* (1.210 fitxa).
- Informatikan, 'Sarrera-irteera eta komunikazioak' arloko 250 fitxa.

2001

- *Zuzenbide Prozesaleko hiztegia* (852 fitxa).
- *Internet hiztegia* (314 fitxa).
- *Geografia-entitateen hiztegia* (380 fitxa).
- *Hezkuntza Administrazioa* (Hezkuntza Saila)
- *Administrazio eta Prozedura hiztegia* (HAEE-IVAP)
- *Europako Itunen hiztegia* (HAEE-IVAP)
- *Bide Seinaleak*.

2002

- *Ardoaren hiztegia* (480 fitxa)
- *Terminologiari buruzko hiztegia* (179).

b) Lehendik bankuan dauden terminoetako formak berrikustea

Etengabe egin beharreko zeregina da. Azken urte hauetan, Hiztegi Batua gorputz hartzen joan den neurrian, forma asko egokitu behar izan dira terminoetan. Bestalde, hiztegi zaharretako terminoen iturriak ere pixkanaka egiaztatzen dira. Hala ere, esan beharra dago, terminoen berrikuspena ez dela egiten guk nahi genukeen edo komenigarria litzatekeen erritmoan. Baina, eguneratze-tasa handitzeak, baliabide gehiago eskatuko luke.

c) Termino egokia proposatzea

Terminoen eguneratzea ez da forma-egokitzearekin amaitzen. Badago beste eguneratze bat, une bakoitzean normalizazioaren ikuspegitik egokia den terminoa aukeratzearena, alegia. Adibidez, garai batean *adin txikiko* esaten zen lekuan gaur egun badirudi *adingabe* edo *adingabeko* esatea egokiagoa dela. Edota, lehen *ingurugiro* idazten zen lekuan, orain *ingurumen* idaztea agintzen du Euskaltzaindiak. Alde honetatik, bereziki nabarmena izan da urte hauetan egindako eguneratzea, zuzenbideko eta administrazioko arloetan. Baina badaude beste zenbait arlo, esate baterako zientzietakoak, sakonki berrikusi beharrekoak.

Eguneratzeari buruzko atal hau amaitzeko, ez nuke utzi nahi aipatu gabe Jaurlaritzak joan den uztailen kaleratu zuen dekretua, Terminologia Batzordea sortzen duena. Izan ere, aurrerantzean garrantzi berezia izango dute batzorde horren proposamenek bankuaren eguneratzean.

#### 4. EUSKALTERM MANTENTZEKO BEHARREZKOA DEN AZPIEGITURA MINIMOA

Goian adierazi dudan bezala, datu-banku terminologiko bat eguneratuta mantentzeko, ezinbestekoa da azpiegitura iraunkor sendo bat izatea. Bestela, euskararen egungo egoera (edo ibilera) ikusita, honelako banku baten balioa urtetik urtera galduz joango litzateke. Orain dela bost urteko bertsioak ez luke asko balioko gaur egun.

##### TERMINOLOGIA-ZENTROAREN OSAERA:

Zehaztasunetan sartzeak dituen arriskuak kontuan izanda ere, ausartzen naiz esatera hau izan daitekeela gaur egun EUSKALTERMen egiten den eguneratze-tasa mantentzeko

*terminologia-zentro* minimo baten osaera:

- 3 terminologo
- Informatikari bat

- Terminalista bat
- Dokumentalista bat, lanaldi erdiz
- Administrari-kudeatzaile bat, lanaldi erdiz

Lehenik, argi utzi nahi dut, honelako terminologia-zentro bat ez dela termino berriak asmatzeko fabrika bat. Aitzitik, informazio terminologikoa bildu, antolatu eta gizarteratzea da bere egiteko nagusia, babes ofiziala emanez eta ezinbesteko erreferentzia bihurtuz arlo honetan.

#### TERMINOLOGIA-ZENTROAREN EGITEKOAK:

- Teknika eta zientzietako arloetan adituek erabiltzen dituzten terminoak etengabe bildu.
- Bildutako terminoak sistematizatu, eta datu-baseetan behar bezala antolatu.
- Bankuan lehendik dauden terminoak berrikusi, hizkuntzaren normalizazio-arau berriak argitaratu ahala.
- Erabiltzen diren terminoen artean lehen iragazketa bat egin.
- Terminologia Batzorde edo antzeko normalizazio-erakundeentzat terminoei buruzko informazio-txostenak prestatu.
- Terminologian lan egin nahi dutenei laguntza metodologikoa eskaini.
- Hiztegi berriak landu eta argitaratzeko prestatu.
- Termino-bankuen argitalpenerako sistema egokiak prestatu.
- Kontsultaren erabiltzaileek dituzten zalantza puntualak argitu.