

**‘TZOS’: TERMINOLOGIA ZERBITZURAKO ON-LINE SISTEMA<sup>1</sup>**

Arregi X. \*, Arruarte A. \*, Artola X. \*, Lersundi M. \*, Santander G. &amp;, Umbelina J. +

\* (Euskal Herriko Unibertsitatea)

&amp; (IKER UMR 5478, Centre National de la Recherche Scientifique, CNRS)

+ (Enalquiler.com, Bartzelona)

[AURKIBIDEA](#)**1. Sarrera**

Terminoak dira unibertsitarioron eguneroko langaia. Terminoen bidez adierazten ditugu, nork bere jakintza-arloan, kontzeptuak. Baina, zer lanabes erabiltzen dugu langai hori maneiatzeko? Zein tresna erabiltzen dugu, adibidez, terminoen glosategiak osatzeko eta termino horiek partekatzeko? Zein, bestek aukeratutako terminoen berri jakiteko? Hain gurea dugun jarduera honetan, zein aplikazio informatiko erabiltzen dugu?

Galdera eta kezka horiek motibatuta ekin zitzaion *Terminologia Zerbitzurako On-line Sistema* (TZOS) garatzeari. Terminologiak jarduera kolektiboa izan behar duelako ustean, TZOSek terminologia-sortzaile potentzialen komunitateari rol aktiboagoa eskaini nahi die. Elkarlaneko terminologia (*collaborative terminology*) deitu izan zaion molde horretara hurbiltzeko saioa da, nolabait.

Sistemaren hasierako prototipoa Ingeniaritza Informatikoko karrera-bukaerako proiektu batean eraiki zen 2006an. Berriki, prototipo hori sendotzeko eta EHUren testuinguruan ezartzeko ekimena jarri da abian, Euskara Zerbitzuaren laguntzaz.

Honako hauek dira sistemaren ezaugarri behinenak:

- Alderdi funtzionala: hiztegiaren kudeaketa erabiltzaile desberdinen eta, nahi bada, askoren eskuetan geratzen da. Horretarako erabiltzaile-rolak definitu dira. Erabiltzaileek, nork bere rola araberak, noiznahi eta nonahi ahal izango dute terminoak kontsultatu, termino berriak proposatu, proposamenei buruz erabaki, terminoak editatu edo terminoei buruz eztabaidatu. Horiek horrela, tresna hau baliagarri izaten ahal da adostasun/desadostasun terminologikoa azaleratzeko, eta, erabaki terminologikoak hartu behar direnean, euskarria izateko.
- Proposamenen kudeaketa: terminoak proposatzeari begira, proposatzailearen eta gainbegiratzzailearen rola bereizten dira, lehendik proposatutako terminoekin bateratzeko/bereizteko zeregina errazten du eta proposamenen eztabaidarako ingurunea du.
- Informazioaren errepresentazio estandarra: informazio terminologikoa errepresentatzeko formatu eta kodeketa estandarra erabiltzen da, datuen elkartrukerako nazioarteko irizpideak eta jarraibideak errespetatuz.
- Datu-base terminologikoko informazioaren inportazioa eta esportazioa: nork bere kasa osatutako termino-zerrendak edo glosategiak datu-basean gordetzea errazteko, kanpoko

<sup>1</sup> Lan honen egileak proiektu hauetako partaide dira: Eusko Jaurlaritza (IE09-262 eta S-PE09UN52); Espainiako Ministerio de Educación y Ciencia (HUM2007-65966-CO2-02); Espainiako Ministerio de Ciencia e Innovación (TIN2007-63173, TIN2009-14715-C04-01 eta TIN2009-14380); UPV-EHU (GIU09/19 eta UE09/09); AECID (A/024790/09); Frantziako ANR: Agence Nationale de la Recherche (ANR-07-CORP-033).

fitxategietako edukia inportatzeko aukera eskaintzen du. Halaber, kontsulteren eta bilaketen ondorioz lortutako termino-zerrendak esportagarriak dira hainbat formatutara.

- Sistema moldagarria eta lokalizagarria: aplikazioaren hainbat elementu testuinguru ezberdinetara egokitu daitezke aplikaziotik bertatik. Esate baterako, jakintza-arloak edo interfaze-hizkuntzak erraz aukera daitezke.
- Sistema eleaniztuna: testuinguru eleaniztunean lan egiteko diseinatu da aplikazioa. Hasierako prototipoan euskarako, ingeleseko, gaztelaniako eta frantseseko terminoak kudeatzeko egitura definitu da.

TZOS ez da sortu testuinguru isolatu batean; terminoak maneiatzeko beharra ez da mundu akademikora mugatzen. Oso bestela, gaur egun, merkatu globalak eta enpresa-jarduera barneantolatze nahiz nazioartekotzeko nahiak terminologia eleaniztunaren erabilera sistematizatzeko beharra ekarri du, baita industria mundura ere. Hortaz, ez da harrizkoa terminologia kontsultatzeko eta kudeatzeko aplikazio informatikoak ugaltzea.

Aplikazio horien errepasso bat egitetik abiatuko gara 2. puntuan, batez ere gure hurbilpenerako baliagarriak izan daitezkeen aspektuetan fokua jarritz. Sistemaren deskribapenari helduko diogu 3. puntuan. Informazio terminologikoa nola errepresentatzen den azalduko dugu 4. puntuan, eta, azkenik, 5.ean, ondorioak eta aurrera begirakoak nabarmenduko ditugu.

## 2. Terminologiarako aplikazio informatikoak

Sarreran esan dugunez, testu tekniko-zientifikoak landu behar dituztenentzat, edozein dela ere jakintza-arloa, baliabiderik erabilienetakoa dira termino-bankuak. Nekez ariko da zientzialaria edo itzultzailea testu bat atontzen nolabaiteko termino-biltegiak lanerako hartu gabe. Hori horrela, ez da harrizkoa eskala txikiko eta handiko datu-base terminologikoak hain arrunt eta ohikoak izatea.

Testuinguru honetan, erakunde eta ekimen ugari sortu dira, batzuk publikoak eta beste batzuk pribatuak, terminologiaren alorreko lankidetzak sustatzeko eta baliabide terminologikoak garatzeko, arautzeko eta hedatzeko. Elkarte horien xede behinena irizpide metodologikoak finkatzea da, eta baliabide estandarrak eta trukagarriak garatzea, nolabait elkarreragingarritasuna bermatzeko. Erakunde horien artean, nolabaiteko sareak edo komunitateak osatu dira, terminologia ezin baitaiteke jardun isolatuztat hartu. Asmo horiekin sortu ziren *the European Association for Terminology (EAF<sup>2</sup>)*, *the International Information Centre for Terminology (Infoterm<sup>3</sup>)*, *the International Network for Terminology (TermNet<sup>4</sup>)*, eta beste hainbat foro.

Erakunde eta zentro horien babesean eta arlokako profesionalen bultzadari esker, datu-base terminologiko asko eta askotarikoak kontsulta daitezke gaur egun. TZOS tresna garatzen hasterako, termino-bankuen mundu zabal horri erreparatu nahi izan diogu, eta errepasso horren laburpena da hona ekarri duguna. Ez da gure asmoa baliabideen eta tresnen inbentario oso bat egitea. Hori baino, TZOSen diseinuan ideia-iturri izan diren bakan batzuetara mugatuko gara. Hedadura zabala duten datu-banku terminologikoen artean, *Inter-Active Terminology for Europe (IATE)* nabarmendu beharrekoa da. IATE da Europar Batasuneko (EB) datu-banku terminologiko erreferentziazkoa. EBko baliabide terminologikoetarako web-interfaze bat

<sup>2</sup> <http://www.eaft-aet.net/en/index/>

<sup>3</sup> <http://www.infoterm.info/>

<sup>4</sup> <http://linux.termnet.org/>

sortzea zuen helburu proiektu honek, informazioa erraz atzitu ahal izateko eta EBko erakundeen artean estandarizazioa bermatzeko. Lehenagoko *EuroDicAutom* baliabideari lekukoa hartu zion, eta 2007an publiko egin zen IATE interfazea<sup>5</sup>. Erakunde-arte datu-base honetan 1,4 milioi sarrera eleaniztun inguru daude, eta hogeita lau hizkuntza hartzen ditu, denak EBko hizkuntza ofizialak. Euskara kanpo geratu da.

Erabiltzaile-interfazeak (ikus 0. irudia) terminoak eta beren itzulpenak bilatzeko aukera ematen du. Bilaketarako iturburu- eta xede-hizkuntzak zehatz daitezke, bai eta jakintza-arloak mugatu ere. Bilaketaren emaitzak zerrenda batean aurkezten dira, non sarrerako terminoaren ordainak ematen diren xede-hizkuntzetan, beren fidagarritasun-mailarekin eta jatorri-erreferentziarekin. Emaitza-zerrendatik abiatuta terminoen deskribapen osoak kontsulta daitezke. Interfazeak badu *feedback* aukera, posta elektronikoko bidezkoa.

Jakintza-arlo batzuek dituzten behar espezifikoak erantzun nahian, terminologia kontsultatzeko baliabide berezituak garatu dira arlo batzuetan, esate baterako, medikuntzan. Medikuntzan, terminologiaren erabilera kritikotzat hartzen baita, ahalegin handia egin da hizkuntza eta terminoak bateratzeko eta normalizatzeko. Horren adierazgarri da *Unified Medical Language System (UMLS)*<sup>6</sup>. UMLSk biomedikuntzaren arloko corpusak, zerbitzuak, ontologiak, sare semantikoak, sailkatze- eta kodetze-estandarrik eta terminologia biltzen ditu. UMLS meta-tesaurusaren barruan *SNOMED Clinical Terms® (SNOMED CT)*<sup>7</sup> terminologia-bankua garatu da, ingelesezko eta gaztelaniazko terminoekin.

## 1. IATE datu-banku terminologikoaren interfaze grafikoa

SNOMED CT (*Systematized Nomenclature of Medicine--Clinical Terms*) estandarizazio-ahalegin baten emaitza da eta medikuntzako dokumentazioa sortzeko eta kontsultatzeko oinarri terminologikoa finkatzen du. SNOMED CTko errepresentazio-ereduan honako osagaiak daude:

<sup>5</sup> <http://iate.europa.eu/>

<sup>6</sup> <http://www.nlm.nih.gov/research/umls/>

<sup>7</sup> [http://www.nlm.nih.gov/research/umls/Snomed/snomed\\_main.html](http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html)

- Kontzeptuak: hierarkikoki antolatuta, beren granularitatearen arabera.
- Deskribapenak: kontzeptuen deskribapenak, terminoen, sinonimoen edo esaldien bidez adieraziak.
- Erlazioak: kontzeptuen arteko erlazio-estekak. Lau motako erlazioak bereizten dira: definizio-erlazioak, kualifikazio-erlazioak, erlazio historikoak eta erlazio osagarriak.

SNOMED CT egitura aberatsa da bai giza kontsultarako bai software-garapenerako ere. Gure errealitate soziolinguistikotik gertuago, Quebec erreferentzia garrantzitsua da. *Office Québécois de la Langue Française* zentroak garatutako *Le grand dictionnaire terminologique*<sup>8</sup> baliabide ezaguna da. Terminoak sarreretan edo definizioetan bilatzeko aukera ematen du interfazeak, hiru hizkuntzatan: frantsesez, ingelesez edo latinez. Terminoek buruzko informazio osatuagoa kontsulta daiteke frantsesez, besteak beste, erabilera-oharrez hornituta. Interfazeak ez du termino berriak sartzeko aukerarik ematen.

Zentro berak badu beste tresna interesgarri bat: *Inverterm*<sup>9</sup>. Tresna honek Internet bera hartzen du terminologia inbentariotzat. Interneteko 3.200 web-gune inguru indexatuta daude, eta guztira 800.000 terminotik gora kontsulta daitezke. Kontsulta-interfazea web-bilatzaile arrunten antzekoa da oso.

Eremu urriko hizkuntzetan terminologia normalizatzeko jarduerak garrantzi handia du. Horren isla da *Minority Languages and Terminology Policies* mintegia<sup>10</sup>, *European Association for Terminology (EAFTE)* erakundeak antolatua 2007an.

Testuinguru honetan, Kataluniako erreferentzia interesgarria da. Han TERMCAT erakunde ari da, 1985ez gero, terminologia normalizatzeko eta hedatzeko bitartekoak garatzen, nola orokorrak hala arlokakoak. *Cercaterm*<sup>11</sup> da tresnarik hedatuena. *Cercaterm*-ek badu on-line kontsultarako interfaze bat, terminoak eta jakintza-arloak zehazteko aukera eman, eta terminoen itzulpenak eta definizioak erakusten dituena.

TZOS lanaren antzera, badira unibertsitate-mailan garatutako baliabide terminologikoak ere. Adibidez, Katalunian, Bartzelonako Unibertsitateko datu-base terminologikoa (*UBTERM*<sup>12</sup>) edo Kataluniako Unibertsitate Politeknikoko terminologia tekniko eleaniztuna (*UPCTERM*<sup>13</sup>) aipa daitezke. Bi tresnok web bidezko kontsulta-interfaze sinplea dute, bilaketa-terminoa eta arloa zehazteko aukera ematen duena.

Euskal Herrian, datu-banku terminologiko ezagunena *EUSKALTERM* dugu<sup>14</sup>. 100.000tik gora fitxa terminologiko biltzen ditu, eta euskarakoez gain, ingeleseko, frantseseko, gaztelaniako eta latineko ordainak biltzen ditu. Web bidezko interfazeak aukera ematen du terminoak (horietako edozein hizkuntzatan) eta jakintza-arloak zehaztu, eta terminoen definizioak eta itzulpenak kontsultatzeko.

Elhuyar fundazioak berriki argitaratu du *Zientzia eta Teknologiaren Hiztegi Entziklopedikoa (ZTH)*<sup>15</sup>. Baliabide honek 23.000 kontzeptu biltzen ditu. Euskarako

<sup>8</sup> <http://www.granddictionnaire.com>

<sup>9</sup> <http://www.inverterm.com/>

<sup>10</sup> <http://www.eaft-aet.net/en/activities/special-eaft-seminar-minority-languages-and-terminology-policies/>

<sup>11</sup> <http://www.termcat.cat/> > cercaterm

<sup>12</sup> [http://www.ub.edu/slc/ubterm/td\\_Arrencada.html](http://www.ub.edu/slc/ubterm/td_Arrencada.html)

<sup>13</sup> <http://www.upc.edu/slt/upcterm/>

<sup>14</sup> <http://www1.euskadi.net/euskalterm/>

<sup>15</sup> <http://zthiztegia.elhuyar.org/>

terminoekin batera, definizioak eta ingeleseko, gaztelaniako eta frantseseko ordainak ematen dira. Gainera, 607 sarrera entziklopediko ere biltzen ditu. ZTH web bidez kontsulta daiteke, eta erabiltzaile-interfazeak termino-sarreretan eta definizioetan bilatzeko aukera ematen du, arloa zehazteko aukerarekin batera. Sarreraren egitura aberatsa eta landua da.

ZTH eta euskarako nahiz ingeleseko *WordNet*<sup>16</sup> ontologiak integratzeko asmoak *WNTerm* proiektuari eman dio bidea (Pociello *et al.*, 2008). Proiektu honen azken helburua zientzia eta teknologiaren arloko ontologia eleaniztuna garatzea da.

Terminologia kontsultatzeko baliabideak erreparatu ditugu orain arte, baina gure sistemari begira, horiek ez ezik, terminologia kudeatzeko tresnak ere gainbegiratu beharrekoak dira.

Aspektu bat nabarmendu behar da lehenik eta behin: enpresak ere jabetu dira zein garrantzitsua den terminologia ondo kudeatzea. IBMk softwarearen globalizazioaz egiten duen hausnarketan honakoa irakur daiteke<sup>17</sup>:

A survey conducted by the Localization Industry Standards Association (LISA) confirmed that actively managing terminology should be part of any global product development strategy. Eighty-eight percent of respondents who actively manage terminology reported an increase in product quality, while 74 percent reported an increase in productivity, 62 percent saved costs, and 56 percent felt it increased their company's competitive edge.

Eta, jakina, interes komertzial horrek eraginda, enpresa asko dira terminologia kudeatzeko tresnak erabili eta sortu dituztenak. Tresna horien xedea terminologia lantzeko zeregin-kate osoa integratzea da: terminoak corpusetatik erauztea, iragaztea, itzultzea, glosategiak edo bildumak osatzea, esportatzeko eta inportatzeko aukerak ematea, informazio terminologikoa datu-base batean antolatzea (eredu estandarren arabera), informazio hori pertsonalizatzea eta proiektu jakinetarako doitztea, eta kontsultak, bilaketak eta eguneraketak egitea. Terminologia kudeatzeko tresnak dira, esate baterako, honako hauek: *evoTerm*<sup>18</sup>, *heartsome*<sup>19</sup>, *Terminology Management Software*<sup>20</sup> edo *Enterprise Terminology Management*<sup>21</sup>. Tresna kontzeptualak dira, esan nahi baita hizkuntza desberdinetako terminoak kontzeptu bati lotzen zaizkiola.

*TermWiki*<sup>22</sup> ere aurrekoen izaera bereko tresna da, baina azpimarratzekoa da terminologia kudeatzeko hurbilpen tradizionaletik aurrerago doala eta lankidetzako ingurune bat eskaintzen duela.

Esparru komertzialetik akademikora etorri, Katalunian garatutako *Terminus*<sup>23</sup> aipatu beharrekoa da. Tresna hau Universitat Pompeu Fabra eta Institut Universitari de Lingüística Aplicada-ren artean garatu da, eta terminologiarako lan-estazioa da. Bi arlo integratzen ditu: corpusen eta terminologiaren kudeaketa. Egitura kontzeptual hierarkikoa hartzen du oinarri, eta hainbat atal biltzen ditu, besteak beste, proiektuak, iturburuak, dokumentuak edota terminoak lantzeko moduluak (ikus 2. irudia).

<sup>16</sup> <http://wordnet.princeton.edu/>

<sup>17</sup> <http://www-01.ibm.com/software/globalization/topics/terminology/introduction.jsp>

<sup>18</sup> <http://www.evoterm.net/>

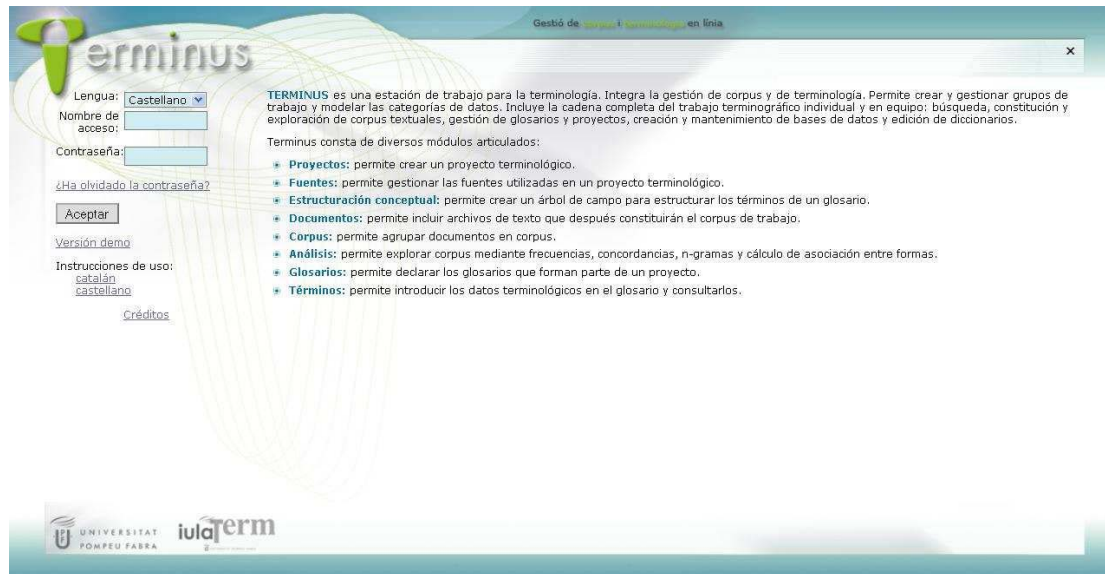
<sup>19</sup> <http://www.heartsome.net>

<sup>20</sup> <http://www.termbases.eu/>

<sup>21</sup> [http://www.acrolinx.com/terminologymanagement\\_en.html](http://www.acrolinx.com/terminologymanagement_en.html)

<sup>22</sup> <http://www.csoftintl.com/termwiki.php>

<sup>23</sup> <http://terminus.iula.upf.edu>



## 2. Terminus lan-estazioa

Ikusi dugu asko direla han eta hemen aurkitu ditzakegun baliabide eta tresna terminologikoak. Ezaugarri aberatsekoak eta tresna sendoak dira horietako asko, beraz, erreferente oso interesgarriak. Esan behar da, hala ere, guztiak direla software jabedunak eta oso bakanak lankidetzako jarduera irekia helburu gisa hartzen dutenak.

Gure ustetan, elkarlanerako ekarpenik interesgarrienetakoa (Désilets *et al.*, 2009)-n deskribatutako *Collaborative Multilingual Terminology (CMT)* ideiatik dator. Autore hauek diotenez,

a massive collaboration process like Wikipedia could have beneficial effects on TDBs (Terminology Databases), by spreading their creation and maintenance costs across a large number of individuals, and by fostering collaboration between terminologists, translators, domain experts, and even members of the general public.

Hain zuzen ere, ideia horri forma emateko saioa da TZOS, unibertsitateko eremuak oso aukera ona eskaintzen duelakoan horretarako.

## 3. TZOS: erabiltzaileak eta funtzionalitatea

TZOS sisteman lau erabiltzaile mota bereizten dira: arruntak, proposatzaileak, gainbegiratzaileak eta administratzaileak.

Erabiltzaile arruntak sistema kontsultatu baino ezin du egin, eta ez du kautotu edo erregistratu beharrik, beste erabiltzaileek bezala. Proposatzaileen eginkizun nagusia terminoak sistemaratzea da. Gainbegiratzaileena, berriz, terminoen hizkuntza aldetiko egokitasuna bermatzea. Aurrerago, funtzionalitatea azaltzen denean, emango dira xehetasun gehiago erabiltzaile mota horiek sisteman duten eginkizunaz.

Azkenik, administratzailea erabiltzaileen eta sistemaren kudeaketaz arduratuko da: erabiltzaileak sortu eta ezabatu, sisteman egin beharreko aldaketak egin (jakintza-arloak aldatu, hizkuntzak gehitu...), foroa administratu, babeskopiak egin eta abar.

Ikus dezagun orain TZOS sistemak funtzionalki eskaintzen dituen aukerak zein diren. Sistemaren hasierako interfazea ikus daiteke 3. irudian.



### 3. TZOSen hasierako interfazea

Hona hemen TZOSen egin daitezkeen eragiketen deskribapen laburra:

- Bilaketak: arrunta eta aurreratua

Erabiltzaile batek kontsulta bat egiten duenean, kontsultatu nahi duen terminoa idatzi beharko du, zein hizkuntzatan bilatu nahi duen zehaztuz. Emaitza pantailan jasoko du. Kontsulta arrunt honetaz gain, bada kontsulta aurreratua egiteko aukera, eta hor zehazten ahalko du jakintza-arloa, terminoaren egoera eta halakoak. Laugarren irudian ikusten ahal duzue "adabegi" terminoari buruz galdetuta jasotako erantzuna.

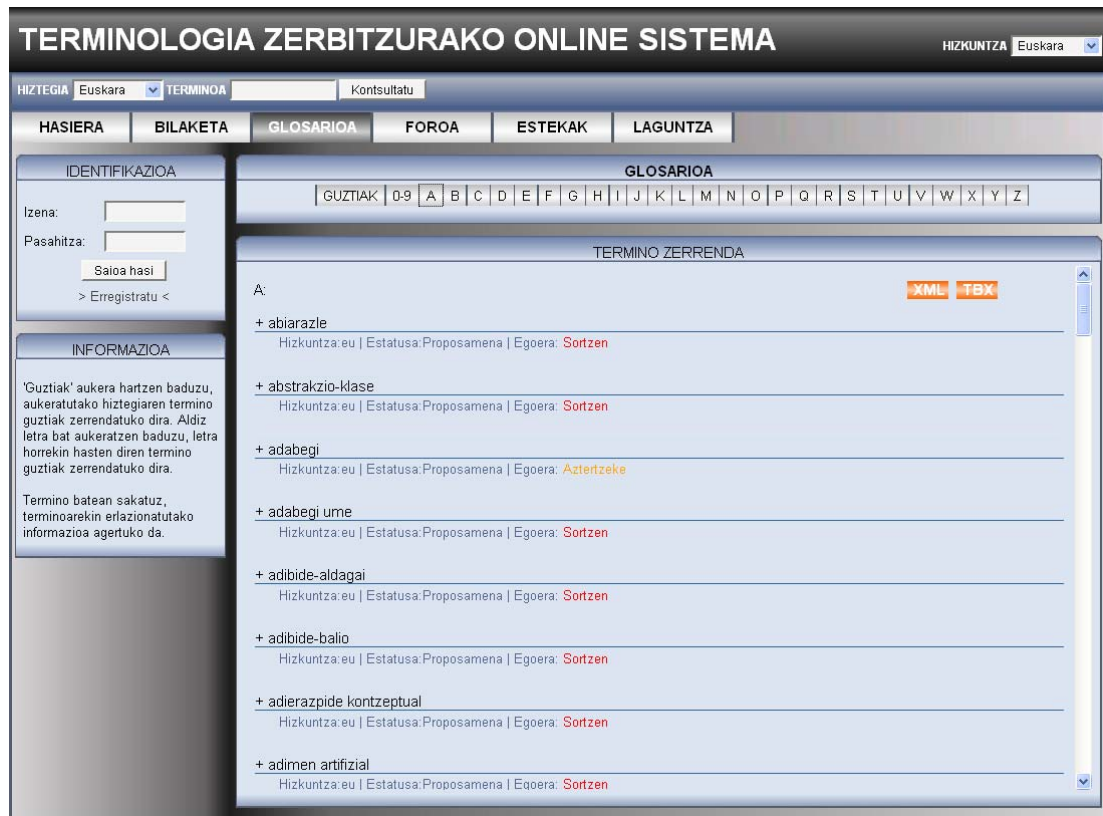
#### 4. Kontsulta aurreratuen interfazea

- Glosarioko bilaketak

Beste kontsulta aukera bat da ikustea glosario osoa edo letra jakin batekin hasten diren terminoak kontsultatzea. 0. irudian "A" letraz hasten diren terminoen zerrendaren hasiera dago.

Bilaketak edozein erabiltzaile motak egin ditzake.





## 5. Glosarioaren interfazea

- Kontsulta-emaitzen esportazioa

Egindako kontsulta guztien emaitza esportatzeko aukera du. Horretarako zein formatutan nahi dituen zehaztu beharko du: RTF, XML, TBX, TXT...

- Proposamenak

TZOSen terminoak jasotzeaz gain, termino horiei buruzko hainbat datu gordetzen ahal dira: terminoa (euskaraz, ingelesez, gaztelaniaz, frantsesez), jakintza-arloa(k), iturria, definizioa, adibideak, iruzkina, testuingurua eta erlazionatutako terminoak. 6. irudian dago "Eredu erlazional" terminoaren proposamena egiterakoan betetako erdua.

Proposatzaileak, proposamenen bat egin nahi izanez gero, sisteman kautotu beharko du, eta ondoren nahi duen terminoa proposatu.

6. Terminoak proposatzeko interfazea

Termino bat sisteman sartzeko proposatzen denean, badira derrigorrean bete beharreko eremu batzuk. Eremu horiek zeintzuk diren aldatzen ahal da.

Terminoak proposatzen den unetik, bere ezaugarri batzuk gordetzen dira. Besteak beste, beti gordetzen da nork proposatu duen, terminoaren egoera administratiboa (edizioan edo sortzen, bukatua...) zein bere onarpenari buruzkoa (egoera bera: estandarra, onartua...; eta autoritatea: Hiztegi Batua, Euskalterm...) sisteman zein den. Era berean, termino batek sistemaratu denetik izan duen bilakaera ere erregistratuta geratuko da, hots, hasierako proposamenetik egungo egoerara izan dituen aldakuntzak (terminoaren formari dagokionez, gainbegiratzailer batek eginak, adibidez) gordeko dira, nork egin dituen eta zergatik jasoz.

Bestalde, termino bat proposatzen denean, gerta daiteke dagoeneko sartuta egotea sisteman, eta orduan proposatzaileak hori adierazten duen mezua jasoko du. Orduan proposatzaileak, sistemaren laguntzaz, erabaki beharko du ea sartzera doan terminoak, lehendik sisteman dagoen termino baten berdina izanik, beste kontzeptu bat denotatzen duen ala ez. Kasu horretan, termino berria homografo gisa geratuko da sisteman; bestela, terminoak ez du sartu beharko.

Beste talka kasu bat da proposatzailea sartzera doan terminoak lehendik sisteman dagoen kontzeptu bat adierazteko beste termino bat denean. Kasu horretan sinonimo gisa sartu behar litzateke termino berria. Hala ere, sistemarentzat aurreko kasuan baino zailagoa da automatikoki detektatzea horrelako baten aurrean gaudela, non eta ez den beste hizkuntza batean sartzera goazen terminoaren baliokidea existitzen dela.

- Terminoaren inportazioa

Terminoak banan-banan sartzeko aukeraz gain, proposatzaileak termino-zerrendak inportatu ditzake. Inportazioa egiterakoan, termino horiek sisteman txertatuko dira, edizioan edo sortzen, gehienetan (maiz ez baitira erabat osaturik egongo). Inportazio-fitxategiek hainbat

informazio-eremu (terminoa euskaraz; terminoa euskaraz eta jakintza-arloa; terminoa euskaraz eta terminoa ingelesez...) izan dezakete, eta horren arabera, inportazio mota bat edo beste aukeratu beharko da.

- Laguntza

TZOS sistemak laguntza-aukera erabiltzea ahalbidetzen du. Laguntza-menu ezberdinak ditu erabiltzaileen arabera. Erabiltzaile mota bakoitzak bere erabiltzaile motari buruzko laguntza besterik ez du ikusiko.

- Eztabaidak eta foroak

TZOSek zalantzak eta eztabaidak bideratu eta ebazteko bideak eskaintzen ditu: besteak beste, gainbegiratzailerak, eztabaida-foroak eta bozketa-sistemak.

- Estekak

Aurrekoez gain, sistemak baditu hainbat esteka erlazioren bat duten baliabideekin (baliabide lexikografikoak batik bat), eta erabiltzaile guztiek kontsultatzen ahalko dituzte.

#### 4. Informazio terminologikoaren errepresentazioa TZOSen

##### 4.1. Estandarrak: TBX

Terminologia informatikoki errepresentatzeko ahalegina kontzeptua oinarri duen printzipioak gidatu du, hasiera-hasieratik. Joan den mendeko 70eko hamarkadatik hasita hainbat estandar egon dira, markatze-lengoaian oinarritzen diren gaur egungo proposamenetara iritsi arte. Aro berri honetan, TEI ekimenak (*Text Encoding Initiative*<sup>24</sup>), bere gidalerroen hirugarren edizioan (P3, 1994) kapitulu oso bat eskaini zion terminologiari, datu terminologikoak errepresentatzeko SGMLn oinarritutako kodeketa proposatuz. TEIren lanak ISOren bidea hartu zuen gero (*ISO Technical Committee 37*<sup>25</sup>), eta hortik sortu ziren MARTIF (ISO 12200:1999) eta datu-kategorien erregistroaren lehen bertsioa (ISO 12620).

Beste mugari garrantzitsu bat historia honetan Europa mailako SALT proiektua (*Standards-based Access service to Lexicons and Terminologies*<sup>26</sup>) izan da. Proiektu horretan aipatutako *ISO Technical Committee 37*-rekin batera aritu ziren, besteak beste, LISA OSCAR taldea (*Open Standards for Container/content Allowing Re-use*<sup>27</sup>), OLIF partzuergoa<sup>28</sup> eta TEI bera, eta lankidetzaren emaitza izan zen TBX (*Term Base Exchange*) formatuaren lehen zirriborroa, gerora TMFren garapena (*Terminological Markup Framework*, ISO 16642:2003) ekarriko zuena.

TBX estandarra (LISA OSCAR, 2008), beraz, hainbat erakunderen arteko lankidetzaren baten emaitza da. LISA OSCAR taldeak 2002. urtean hartu zuen TBX terminologia elkartrukatzeko formatu gisa. Duela bi urte, 2008an, ISO 30042 bilakatu den LISAREN estandar honek informazio lexiko zehaztua adieraztea ahalbidetzen du. ISO 30042 hiru estandarretan oinarritzen da: ISO 12620, ISO 12200 eta ISO 16642. Lehenak, alde aurretik modu estandarrean izendatu eta ondo definitutako datu-kategorien bilduma bat jasotzen du. Bigarrenak, ISO 12200 edo MARTIF ere deiturikoak, TBXren funtsezko egituraren oinarria eskaintzen du. Azkenik, ISO 16642 edo TMF delakoak, terminologiako markatze-lengoaian metaeredu-egitura orokorra definitzen du.

<sup>24</sup> <http://www.tei-c.org/>

<sup>25</sup> [http://www.iso.org/iso/standards\\_development/technical\\_committees/other\\_bodies/iso\\_technical\\_committee.htm?commid=48104](http://www.iso.org/iso/standards_development/technical_committees/other_bodies/iso_technical_committee.htm?commid=48104)

<sup>26</sup> <http://www.ttt.org/salt/>

<sup>27</sup> <http://www.lisa.org/OSCAR-LISA-s-Standa.79.0.html>

<sup>28</sup> <http://www.olif.net/consortium.htm>

TBX datu terminologikoak kudeatzeko formatu estandar eta irekia da, XMLn oinarritua. Besteak beste, datu terminologikoak analizatzeko, modu deskribatzailean adierazteko, zabaltzeko eta konputagailu-inguruneen artean elkartrukatzeko aukera ematen du. Esan daiteke TBXren helburu nagusia datu terminologikoen elkartrukatzea dela. XML formatu bat izanik, ez da egokia informazio terminologikoa nola bistaratuko den adierazteko, bere egitura-edukiak deskribatzeko baizik. Datu-base terminologikoak erabiltzen dituzten software-pakete gehienek eskaintzen dute TBX fitxategiak esportatzeko eta inportatzeko aukera. Markatze-eredu unibertsal bat eskainiz, enpresek eta erakundeek erabiltzen duten barne-terminologia kudeatzeko ez ezik, beste edonorekin partekatzeo aukera ezin hobea ere ematen du.

Datu-base terminologikoetan (*termbases*) jaso edo erabiltzen diren datu-kategoriek lan egiteko, TBXk bi modulu ditu: funtsezko egitura definitzen duena, batetik, eta datu-kategoriak eta horien arteko murriztapenak identifikatzeko formalismoa (XCS, *eXtensible Constraint Specification*), bestetik. Zentzu honetan, esan daiteke TBX modularra dela. Bata zein bestea XMLz adierazten dira. Beraz, TBX kontzeptua bere horretan aipatzen denean, bi modulu horien elkarrekintzaren emaitzari buruz ari garela jakin behar da.

Egungo datu terminologikoen elkarreragingarritasuna maximizatzeko aldera, TBXk datu-base terminologikoetan erabiltzen diren datu-kategorien multzo lehenetsi bat eskaintzen du: besterik ezeko XCSa. Katetoria lehenetsi horien helburu nagusia errepresentazio-mekanismo itsu bat ematea da. Errepresentazio itsu honek datuak emailearekin kontsultatu gabe interpretatzea ahalbidetuko du.

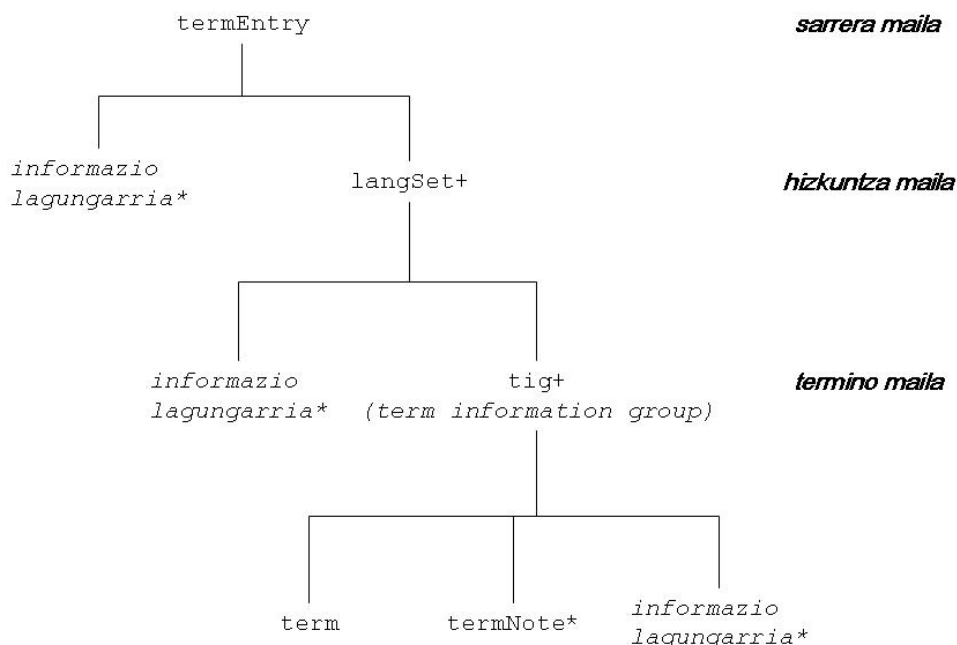
Nahiz eta datu-kategoria lehenetsiak eskaini, ez dira maiz datu-kategoria berdin-berdinak erabiltzen dituzten terminologia bildumak topatzen. Zentzu honetan, TBX formatua malgua da. XCS mekanismoaz baliatuz, erabiltzaile talde bakoitzak, bere eskakizunak kontuan hartuta, dagozkion datu-kategoria propioak defini ditzake. Datu-kategoriak eta haien arteko murriztapenak XCS fitxategi batean jasoko dira. Horrela, talde bakoitzak bere beharretara egokitutako TMLa (*Terminological Markup Language*) definitzen du, informazio terminologikoa markatzeko XML lengoia edo aplikazio jakin bat, alegia.

#### **4.2. TZOSen egindako aukera, adibide baten laguntzaz azaldua**

TZOSen terminologia errepresentatzeko aukeratu dugun datu-eredua deskribatuko dugu orain. Horretarako, TBX estandarraz baliatu gara.

Goian esan bezala, TBX estandarrek informazio terminologikoa elkartrukatzeko XML formatu bat eskaintzen du. Formatu horretaz baliatuz, TZOSen erabiliko dugun TMLa definitu dugu: TZOS\_TML. TZOS\_TML definitzeko garaian, XCS lehenetsia ere egokitu behar izan dugu, euskal terminologiaren eta eraiki nahi dugun sistema terminologikoaren zenbait berezitasun bildu ahal izateko.

TZOS\_TML horretaz baliatuko gara orain aipatutako datu-eredua azaltzeko. Hori baino lehen, ordea, argi dezagun kontu bat. Hemendik aurrera adibide batez azalduko dugunaren funtsa datu-eredua da, TZOSen informazio terminologikoa errepresentatzeko aukeratu duguna, TBXn oinarrituz. Beraz, horrek ez du esan nahi TZOSen terminologia, fisikoki, horrela gordeko denik. Aurretik egin dira saioak XML datu-base natiboak erabiliz (Berkeley DBXML), eta bai datu-base erlazionalak erabiliz ere (MySQL). Proiektua, egun, birdiseinu-fase batetik ateratzen ari da, eta aurki erabaki beharko da datu-eredua inplementatzeko zer teknologiaz baliatu, datuak fisikoki nola gordeko diren, alegia. Behin hori argituta, goazen aurrera.



## 7. Sarrera terminologiko baten egitura eta mailak

TBX ereduari jarraituz, sarrera terminologiko batek 7. irudian azaltzen den egitura izango du. Informazio terminologikoko bilduma bat horrelako sarrerek osatzen dute, kontzeptu bakoitzeko sarrera bat (`termEntry`). Hiru maila bereizten dira bertan, eta bakoitzean mailari dagokion informazioa adierazten da. Sarrera mailan, kontzeptuari dagokion informazioa errepresentatzen da; hizkuntza mailan, kontzeptuaz dugun informazioa errepresentatzen da, hainbat hizkuntzatan (irudiko + ikurrak adierazten du `langSet` bat edo gehiago ager daitezkeela hor, bat hizkuntzako); azkenik, termino maila dator, non `tig` elementu batean errepresentatuko baita terminoa bera (`term`) eta dagokion informazioa. Kontzeptua denotatzen duten hainbat sinonimo adierazi nahi badira, `tig` bat baino gehiago erabiliko dira, bat termino bakoitzerako. Irudian ikusten denez, edozein mailatan informazio lagungarria zehazteko aukera dago (\* ikurrak 0 edo gehiagotan ager daitezkeela adierazten du): definizioa, adibideak, informazio administratiboa (nork sortu duen sarrera, zer jakintza-arlotan erabiltzen den, zer bilakaera izan duen datu-basera iritsi denetik...) eta abar. Bestalde, termino bakoitzari lotuta hainbat `termNote` ager daitezke, terminoaren egoera edo estatusa zehazteko, adibidez, edo, araututa baldin badago, autoritate arauemailea zein den aipatuz; halaber, `termNote` elementuak erabiliko dira terminoaren kategoria gramatikala eta beste ezaugarri batzuk adierazteko ere.

Atal honi bukaera emateko, berriz, ikus dezagun adibide oso samar bat: TBXren arabera errepresentatutako sarrera terminologiko bat, atalez atal azalduta<sup>29</sup>. Sarrera informatikako *datu-base* kontzeptuari dagokio. Demagun kontzeptu hori sisteman unibokoki bereizten duen identifikadorea `c2` dela. Hona hemen `termEntry` elementuaren lehen atala, non

<sup>29</sup> Esan gabe doa hau adibide bat baino ez dela, eta litekeena da sarreraren benetako adierazpidea TZOSen ez izatea berdin-berdina.

kontzeptuari buruzko informazio lagungarria adierazten baita:

```
<termEntry id="c2">
  <admin type="conceptOrigin">DIF</admin>
  <admin type="originatingPerson">jiparbex</admin>
  <admin type="productSubset">Apunteak</admin>
  <admin type="productSubset">Tesiak</admin>
  <descripGrp>
    <descrip type="explanation" xml:lang="eu">Ez da datu-banku
      zentzu zabalean
    </descrip>
    <note>Kontzeptua bereizteko oharra</note>
  </descripGrp>
  <descrip type="subjectField">Informatika</descrip>
  <descrip type="subjectField">LSI</descrip>
  <descrip type="subjectField">Datu-baseak</descrip>
  <descrip type="example" xml:lang="eu">Datu-baseak datuak
    gordetzeko erabiltzen diren fitxategi-sistema
    konplexuak dira
  </descrip>
  <descrip type="superordinateConceptGeneric" target="c3">
    biltegi
  </descrip>
  ...
</termEntry>
<termEntry id="c3">
  ...
</termEntry>
```

Hor goian agertzen den zuhaitzeko informazio lagungarria da, c2 kontzeptuari dagokiona, hain zuzen: kontzeptua non sortu den, sistemara nork ekarri duen<sup>30</sup>, zer produktu klasetan erabiltzen den, kontzeptua bera zehazten eta antzekoetatik taxuz bereizten laguntzeko azalpen moduko bat, erabilera-eremu edo jakintza-arloak, adibideren bat, eta, bukaera aldera ikus daitekeenez, hauxe da leku egokia kontzeptuen arteko erlazio zenbait zehazteko ere, hala nahi izanez gero (*datu-base*, c2, eta *biltegi*, c3, kontzeptuen artean hiponimia/hiperonimia erlazio bat dagoela adierazten da adibidean, esate baterako)<sup>31</sup>.

Informazio lagungarri horren ondoren, hizkuntzen atalak datoz, langSet elementuak. Hona hemen euskarazko atalaren lehen parte, hizkuntza-atalaren informazio lagungarriari dagokiona:

...

<sup>30</sup> Horrelakoetan kodeak erabiltzen dira. Kode horiek TZOSen beste atal batzuetan deskribatuta egongo dira, beren ezaugarriekin: iturburuaren atala, sistemaren erabiltzaileena eta abar.

<sup>31</sup> Ikusten denez, *descrip* elementuei oharren bat gehitu nahi zaienean (*note* elementuaren bidez, adibidean), *descripGrp* elementuetan bildu behar dira *descrip* elementua bera eta dagozkion beste elementu lagungarriak. Antzera jokutzen da TBXn *admin* elementuekin eta beste zenbaitekin ere.

```

<langSet xml:lang="eu">
  <descripGrp>
    <descrip type="definition">Datuak biltegitratzeko eta
      erabiltzeko azpiegitura eskaintzen duen
      baliabide informatikoa</descrip>
    <note>Bat-bateko definizioa</note>
  </descripGrp>
  <descrip type="example">Datu-baseak hiru mailako arkitekturan
    oinarritzen dira</descrip>
  ...

```

Hizkuntza-atalaren hizkuntza zein den zehazten da `xml:lang` XMLren atributu orokorraren bidez. Ondoren, terminoa euskaraz definitzen duen esaldi bat dator, bere ohar eta guzti, eta, bukatzeko, adibide bat. Hizkuntza-atal honen senide dirateke ingeleseko, gaztelaniako eta frantseseko hizkuntza-atalak ere, non kontzeptua erdara horietan adierazteko erabiltzen diren termino baliokideak egongo baitira:

```

<langSet xml:lang="en">
  ...
  <tig>
    <term>database</term>
    ...
  </tig>
  ...
</langSet>
<langSet xml:lang="es">
  ...
</langSet>
<langSet xml:lang="fr">
  ...
</langSet>
...
</termEntry>

```

Jarrai dezagun, baina, euskarazko hizkuntza-atalarekin, ez baitugu oraindik erabat azaldu. Aipatu dugun informazio lagungarriaren ondoren, `tig` elementuak datoz, hau da, terminoari buruzko informazioa biltzen duen atala. Elementu horietan deskribatzen dira kontzeptua denotatzen duten termino guztiak, eta, beraz, dokumentuan senide diren `tig`ak hizkuntza batean kontzeptu bera adierazteko erabiltzen diren termino sinonimoak liriateke.

```

...
<tig>
  <term>datu-base</term>
  <termNote type="administrativeStatus">
    standardizedTerm-admn-sts
  </termNote>

```

```

<termNote type="normativeAuthorization">
  standardizedTermHB
</termNote>
<termNote type="termType">fullForm</termNote>
<termNote type="partOfSpeech">izena</termNote>
<admin type="entrySource">Euskalterm</admin>
<transacGrp>
  <transac type="transactionType">origination</transac>
  <note>...</note>
  <date>1999-11-12</date>
  <transacNote type="responsibility">jiparbex</transacNote>
</transacGrp>
</tig>

```

...

Hemen goian, *datu-base* terminoa deskribatzen da: bere egoera administratiboa (termino estandarra da), erakunde arauemailea eta erakunde horrek aitortzen dion estatusa (*standardizedTermHB*: Hiztegi Batuan arautua), forma osoa dela (eta ez laburtzapean bat), izena dela, Euskaltermen duela bere iturburua, eta, azkenik, terminoa TZOSera sartu zeneko zenbait datu: data, nork sistemaratu duen eta abar.

Transakzioen bidez erregistratzen da termino bakoitzaren bilakaera TZOSen barruan, terminoa TZOSeratu denetik izan dituen gorabehera guztiak, alegia. Transakzioek dokumentuan duten ordenak adieraziko luke transakzio horien kronologia.

Terminoaren bilakaeraren kontu hori argitzearren, ikus dezagun, bukatzeko, *datu-base* terminoaren laburtzapean gisa erabiltzen den *DB* siglaren kasua. Kontzeptu bera izanik, *datu-base* terminoa ageri den sarrera terminologiko berean egongo litzateke, euskarazko hizkuntzatan, gure kasuan, bere *tig* propioan. Demagun, transakzioen gaia hobeto argitzearren, *DB* baino lehen baten batek *db* proposatu duela, eta ikus dezagun bi laburtzapean horien historia nola geratzen den erregistratuta TZOSen:

```

<tig>
  <term>db</term>
  <termNote type="administrativeStatus">
    supersededTerm-admn-sts
  </termNote>
  <termNote type="termType">acronym</termNote>
  <transacGrp>
    <transac type="transactionType">origination</transac>
    <date>2000-11-12</date>
    <transacNote type="responsibility">
      jiparbex
    </transacNote>
  </transacGrp>
  <transacGrp>
    <transac type="transactionType">withdrawal</transac>
    <note>Maiuskulatan behar duela iruditzen zait</note>
    <date>2000-11-13</date>
  </transacGrp>
</tig>

```



```

        <transacNote type="responsibility">
            jiparlaa
        </transacNote>
    </transacGrp>
    <descrip type="sampleSentence">
        db-a, db-rik... erabili behar da, marratxo
        eta guzti
    </descrip>
</tig>
<tig>
    <term>DB</term>
    <termNote type="language-planningQualifier">
        proposedTerm
    </termNote>
    <termNote type="termType">acronym</termNote>
    <transacGrp>
        <transac type="transactionType">origination</transac>
        <date>2000-11-13</date>
        <transacNote type="responsibility">
            jiparlaa
        </transacNote>
    </transacGrp>
    <descrip type="sampleSentence">
        DBa, DBrik... erabili behar da, maiuskulatan
        eta marratxorik gabe
    </descrip>
    <descrip type="example">
        Informazio guztia DBratu zuten bi ordutan
    </descrip>
</tig>
...

```

Bi tig ditugu hor, *datu-base* terminoaren *db* eta *DB* laburtzapenei dagozkienak, hain zuzen. Lehenengoan, *db*-ri dagokionean, bi transakzio ageri dira, jatorrizkoa, jiparbex erabiltzaileak egindakoa, eta jiparlaak egindakoa hurrena (*withdrawal* motakoa). Lehenengo erabiltzaileak proposatutakoa zuzentzera etorri da, antza denez, bigarrena, eta ohar baten bidez adierazten du bere ustez maiuskulatan behar lukeela laburtzapenak: *DB*. Bigarren tigean dator, hain zuzen ere, jiparlaa erabiltzaileak proposatutako *DB* laburtzapena, erabilera-oharrez eta adibidez hornitua.

## 5. Ondorioak eta aurrera begirakoak

Terminologiaren sorkuntza eta hedapena prozesu neketsua denez, prozesu hori arinagoa, biziagoa eta emankorragoa egingo bada, elkarlanean aritzea proposatzen da artikulu honetan. Hemen aurkeztu den TZOS tresna elkarlaneko termino-sorkuntzari eta hedapenari lagunduko dion tresna da. Baliabide tekniko da, baina lan-molde baten euskarri izateko bokazioa du. Beraz, TZOSek inplikazioa eskatzen du, auzolana eta parte-hartzea, benetan emankorra

izango bada. *Terminogintza 2.0* izenda zitekeen joera hori, terminogintza klasikoa gaur egungo komunikazio- eta informazio-teknologiez mamitzeko proposamena den neurrian.

Molde horri eusteko, aplikazioaren diseinu funtzionalak erabiltzaile-mota aktiboek garrantzi handia ematen die, eta, kontsultarako ez ezik, terminoak proposatzeko eta eguneratzeko aukerak lantzen ditu. Gainera, terminologiari buruzko eztabaida sustatzeko, foro bat integratzen du.

Informazio terminologikoaren errepresentazioari dagokionez, TZOSen TBX estandarrean oinarritu gara, eta gure beharretara egokitu dugu TBXk berak proposatzen duen bideari jarraituz. Nazioartean onartutako estandar bat erabiltzeak informazioaren kudeaketarako nahiz elkartruckerako eskaintzen dituen abantailak ugariak dira.

Esan behar da, bestalde, TZOSek ez duela integratzen corpus-lanketarik, ez termino-erazketa (erdi-)automatikorik. Ez du, hortaz, terminologiako jarduera-katea bere osotasunean jasotzen. Etorkizuneko lan interesgarria izan daiteke beste zenbait tresnaren eta baliabideren erabilpena integratzea.

Prototipo mailakoa da gaur egungo bertsioa. Ezin izan da orain arte testuinguru erreal batean martxan jarri. EHUko Euskara Zerbitzuak sustatuta, ekimen bat jarri da abian TZOS EHU ezartzeko eta erabiltzeko. Hori da, hain zuzen ere, helburu gertukoena. Erabiltzea eskatzen baitu TZOSek, hots, partekatzea, eztabaidatzea, kontsultatzea, norberetik ematea, besterenetik jasotzea. Eta erabilerak esango du zenbateraino den egokia.

## Bibliografia

- POCIELLO E., GURRUTXAGA A., AGIRRE E., ALDEZABAL I. and RIGAU G. (2008): "WNTERM: Combining the Basque WordNet and a Terminological Dictionary", *Proceedings of the 6th International Conference on Language Resources and Evaluations (LREC)*. Marrakech.
- DÉSILETS A., HUBERDEAU L-P., LAPORTE M., QUIRION J.: *Building a Collaborative Multilingual Terminology System*, Internet: [http://wiki-translation.com/tiki-download\\_wiki\\_attachment.php?attId=57](http://wiki-translation.com/tiki-download_wiki_attachment.php?attId=57)
- LISA Open Standards for Content/Container and Reuse (OSCAR) (2008): *Systems to manage terminology, knowledge and content - TermBase eXchange (TBX)*.

[AURKIBIDEA](#)