

DB (DOKUMENTU BILTEGIA): CORPUS AKADEMIKOAK SORTZEKO ETA KUDEATZEKO AZPIEGITURA TEKNOLOGIKOA ¹

Joseba Ezeiza Ramos
(Euskal Herriko Unibertsitatea)

[AURKIBIDEA](#)

0. Sarrera

Lan honetan *Dokumentu Biltegia*[®] izeneko corpus-tresna aurkeztuko dut (DB)², testu-corpusak on-line sortu eta kudeatzeko aplikazio informatiko bat, alegia. Aplikazio hau espezialitate-hizkeren ikerketarako eta irakaskuntzarako sortu da, eta bereziki pentsatuta dago unibertsitatearen alorreko ekoizpenak gorde eta aztertzeke. DB hiru fasetan garatu da 2007-2010 epealdian, Ametzagaiña I+G³ enpresa-ikerketarako gunearekin elkarlanean⁴.

Lehen fasean DBren oinarrian dagoen datu-base dokumentala egin zen. Datu-base horrek aukera ematen du DBren plataformaren barruan nahi adina corpus edota “biltegi” sortzeko. Biltegi horietan modu erraz eta malguan jaso eta sailka daitezke hainbat formatutako materialak (dokumentu idatziak, audioak, bideoak, irudiak...). Fase honetan datu-baseari bilaketa-motor aurreratu bat ere erantsi zitzaion, DBn jasotako dokumentuak modu erraz eta malguan berreskuratu ahal izateko.

Bigarren fasean, testu idatziak prozesatzeko eta arakatzeko moduluak erantsi zitzaizkion DBren oinarritzko egiturari. Modulu horiek sortzeko, Ametzagaiña I+G-ren jabetza diren Kapsula⁵ eta Corpus Arakatzaila⁶ izeneko tresnak erabili ziren. Horiez gain, testuak bihurtzeko modulu bat ere garatu zen. Modulu honek ohiko formatuetan kargatutako testuak (doc, pdf, ppt, docx, html, xml, eta abar) prozesamendu linguistikorako prestatzen ditu, zenbait ezaugarri gordeta: estilo-markak, marka ortotipografikoak, paragrafo- eta lerro-jauziak, eta abar.

Hirugarren fasean, azkenik, tresna osagarri batzuk garatu eta integratu dira DBren egiturari. Batetik, testuak editatzeko lan-mahai bat erantsi zaio DBri; editore horrek aukera ematen du bihurtze automatikoaren fasearen ondoren ere, testuak edozein momentutan eskuz landu edota zuzentzeko. Bestetik, testu-esportazioak XML formatuan egiteko tresna bat ere garatu da, DBn landutako dokumentuetan TEI markatze-sisteman oinarritutako etiketak eta TEI goiburuak txertatzen dituena. Horrez gain, dokumentu baten edota dokumentu-multzo baten bereizgarri den lexikoa erauzteko tresna bat ere garatu da (Lexiko Bereizlea). Eta,

¹ Lan hau hiru proiekturen elkargunean dago: HIZLAN (Bizkaiko Foru Aldundia & EHU: DIPE08/16), RICOTERM-3 (MICINN: HUM2007-65966-C02-02) eta EBALUA (UPV/EHU: EHU 08/53)

² Erregistro-zenbakia: SS-236-09

³ <http://www.ametza.com/euskara/ametza.htm>

⁴ Eskerrak eman nahi dizkiet Aitor Maritxalar, Josu Landa eta Sebi Calonjeri proiektu honi egin dioten ekarpenagatik

⁵ <http://www.kapsula.com/>

⁶ Tresna hau Klasikoen Gordailua [<http://klasikoak.armiarma.com/corpus.htm>] eta Ereduzko Prosa Gaur [<http://www.ehu.es/euskara-orria/euskara/ereduzkoa/>], Ereduzko Prosa Dinamikoa [<http://ehu.es/ehg/epd/>] eta ZIO Corpora [<http://www.ehu.es/ehg/zio/>] garatzeko erabili da, besteak beste.

azkenik, balio terminologikoa izan dezaketen patroi morfosintaktikoak identifikatu eta erazteko balio duen modulu bat ere sortu da DBren datu-basean integratzeko.

Horrela bada, DBk aukera zabala eskaintzen du testu akademikoak modu erraz eta malguan jaso eta sailkatzeko, eta testu horien ezaugarri linguistikoak arakatu eta alderatzeko. Ahalegin honekin erantzun nahi izan zaio, zati batean behintzat, 2006-2007 urtealdian Euskal Filologia Saileko hainbat kidek garatutako PREST⁷ (Ezeiza, 2008) eta GARATERM⁸ (Zabala, 2008) proiektuetan azaleratu zen behar bati: komunitate akademikoaren ekoizpenak jasotzeko, kudeatzeko eta aztertzeke azpiegitura teknologiko egoki baten premiari, alegia.

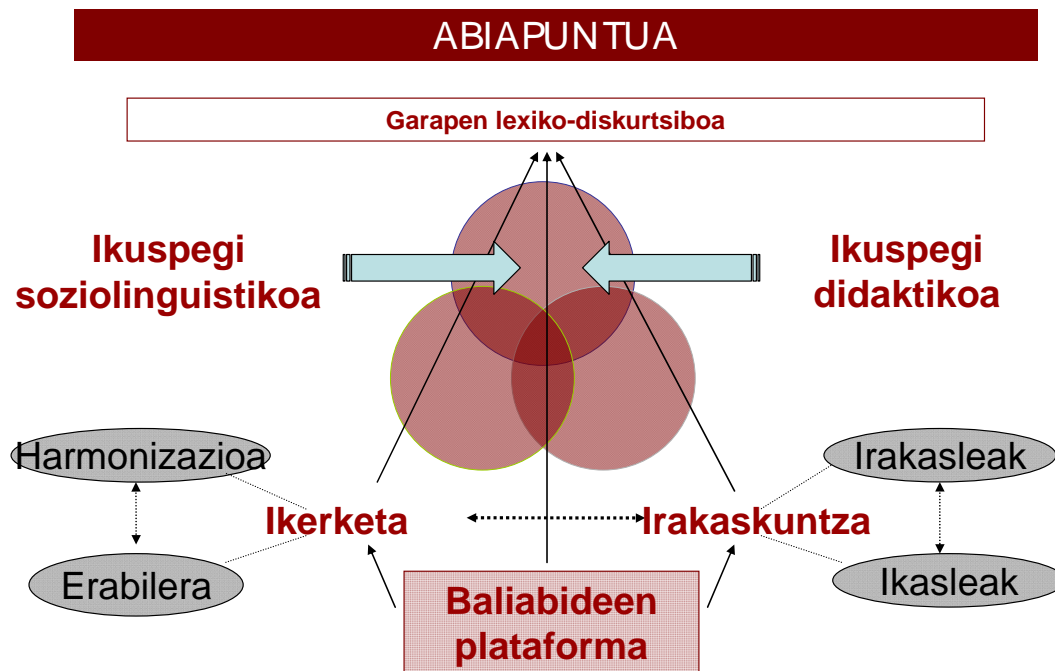
PREST proiektuan saiatu ginen zehazten zer toki eta trataera izan behar duen espezialitate-hizkeren irakaskuntzak unibertsitateko ikasketa-plan berrietan. GARATERM proiektuan bestalde, helburuetako bat izan zen aztertzea zer erronkari egin behar dion aurre hizkera berezituaren didaktikak euskararen testuinguru soziolinguistikoan, kontuan izanda arlo akademikoan euskararen garapen lexiko-diskurtsiboa ez dela amaitu eta ez dagoela erabat normalizatuta, eta ondorioz irakatsi behar dena (espezialitate-hizkera bera, alegia), aldatzen ari den zerbait dela.

PREST proiektuaren emaitzek agerian utzi zuten komunikaziorako gaitasun akademiko eta profesionalak unibertsitateko curriculum berriaren funtsezko osagai direla, eta ikasketa planek bermatu behar dutela unibertsitateko egonaldian ikasleek bereganatzen dutela arloko testu-genero nagusien bereizgarri sozio-kognitiboen, arloko erregistroen adierazpideen eta arloko lexiko eta terminologiaren ezagutza funtzional jaso. Baina, lehen esandakoagatik, helburu horri eraginkortasunez erantzuteko, euskararen kasuan funtsezkoa da espezialitate-hizkeren irakaskuntza euskarak arloz arlo duen eta izan duen garapen lexiko-diskurtsiboaren testuinguruan kokatzea; beharrezkoa da, halaber, arloz arloko diskurtso eta erregistro berezituak karakterizatzeko eta diskurtso eta erregistro horietan gertatzen diren aldaketak sistematikoki jasotzeko modua izatea; eta, azken batean, ezinbesteko dirudi ikerketa, didaktika eta garapena uztartzeko bideak jorratzea. Hori ondorioztatu zen, behintzat, GARATERM proiektuan.

Beste era batera esanda, espezialitate-hizkeren karakterizazioa eta didaktika jorratuko dituen lan-ildo bat zabaldu nahi bada, beharrezko dirudi gune bateratzaile bat sortzea ikuspegi soziolinguistikoa eta ikuspegi didaktikoa uztartuko dituen, euskararen erabileraren eta harmonizazioaren alorrean zein hizkuntzaren irakaskuntzaren alorrean egiten eta egingo diren ahaleginak optimizatzeko; betiere, euskararen garapen lexiko-diskurtsiboari laguntzea jomugan hartuta. Lehen irudian laburbiltzen dira bilgune estrategiko horrek izango lituzkeen zereginak eta lan-ildoak (ikusi 1. irudia).

⁷ Kodea: DGE06/04

⁸ Kodea: EJIE07/07



1. irudia: Baliabideen gune estrategikoaren jomugak

Gune estrategiko horren eginkizuna, funtsean, hauex izango litzateke: azpiegitura teknologiko egokia garatzea eta sare sozial eraginkor bat sortzea, komunitate akademikoaren ekoizpenak jaso, sistematikoki aztertu eta, euskararen corpus akademikoaren karakterizazioa egiten joateko. Karakterizazio-lan horrek motibazio eta jomuga ezberdinak izan ditzake: ikerketa-proiektu formalak egitea, praktika-komunitateen premiei erantzungo dieten neurri-neurriko hizkuntza baliabideak sortzea, euskararen garapen eta normalizazioaren alorrean lanean ari diren beste eragile batzuentzat datuak biltzea... Baina, bigarren irudian ikus daitekeen moduan, prozesu hori abian jarri ahal izateko ezinbestekoa da ekoizpen akademikoak biltzeko, sailkatzeko eta aztertzeko aukera emango duen tresna malgu bat, dokumentu biltegi bat alegia (ikus 2. irudia).

DBren segmentu edota biltegi-tako bakoitzean soil-soilik sartu ahal izango dira horretarako pribilegioak eskuratuta dituzten erabiltzaileak. Horrela bada, erabiltzaile batzuek aukera izango dute testuak sistemaren segmentu jakin batera igo eta testu horien sailkapen-fitxak bete eta aldatzeko. Beste batzuek, testuak igo ahal izango dituzte, baina ezin izango dituzte fitxetan edota testuetan bertan aldaketak egin. Beste erabiltzaile batzuek kontsultak baino ezin izango dituzte egin. Alabaina, administratzaile orokorrak biltegiak sortzeko, aldatzeko eta ezabatzeko aukera izango du. Horrez gain, aukera izango du segmentu edo biltegi guztietan aldaketak egiteko, dokumentuak biltegi batetik bestera mugitzeko edota dokumentu baten kopiak beste biltegi batzuetara esportatzeko.

DBn txertatutako kontsulta-tresnek, azkenik, hainbat informazio eskuratzeko aukera ematen dute. Bilaketa-motorraren bidez, ezaugarri-multzo jakinak betetzen dituzten testuak bilatu eta bereizi litezke. Nahi izanez gero, testu-multzo horiek XML formatuan esporta daitezke, edota biltegi berri batera ere eraman daitezke, berariaz aztertu ahal izateko. Corpus Arakatzaillearen bidez, bestalde, hitzen, lemen eta kategoria morfosintaktikoen arabera bilaketak egin ahal izango dira. Kontsultaren emaitza gisa, konkordantziak (testuinguru laburrak) zein testuinguru zabalak berreskuratu ahal izango dira. Agerpen-maiztasunen datuak ere erraz lortu ahal izango dira, eta datu horiek EXCEL tauletara esportatzeko modua ere eskaintzen da. Lema edota hitz bat barne hartzen duten bi edo hiru hitzen segidak ere erauz litezke DBn (bigramak eta trigramak) eta baita patroi morfosintaktiko jakinen arabera egiturak ere. Lexiko Bereizleak, azkenik, aukera emango digu testu-multzo baten bereizgarri diren elementu lexikoak erauzteko eta ordenatzeko, elementu horien esangarritasun estatistikoaren arabera.

Sistema osoa pentsatuta dago komunitate akademikoa osatzen duten kideek (ikasleek, irakasleek zein ikertzaileek) zailtasun berezirik gabe erabili ahal izan dezaten. Horregatik, hain zuzen, DBk ez du gaitasun tekniko aurreraturik eskatzen. Horrela bada, erabiltzaileek dagokien ingurunean autonomia osoa izango dute testuak kargatzeko, analizatzeko eta kontsultatzeko. Horrek aukera zabalak eskaintzen ditu ekoizpen akademikoak modu errazean jasotzeko, aztertzeko eta ikerketan zein irakaskuntzan erabiltzeko.

2. Administrazio-modulua

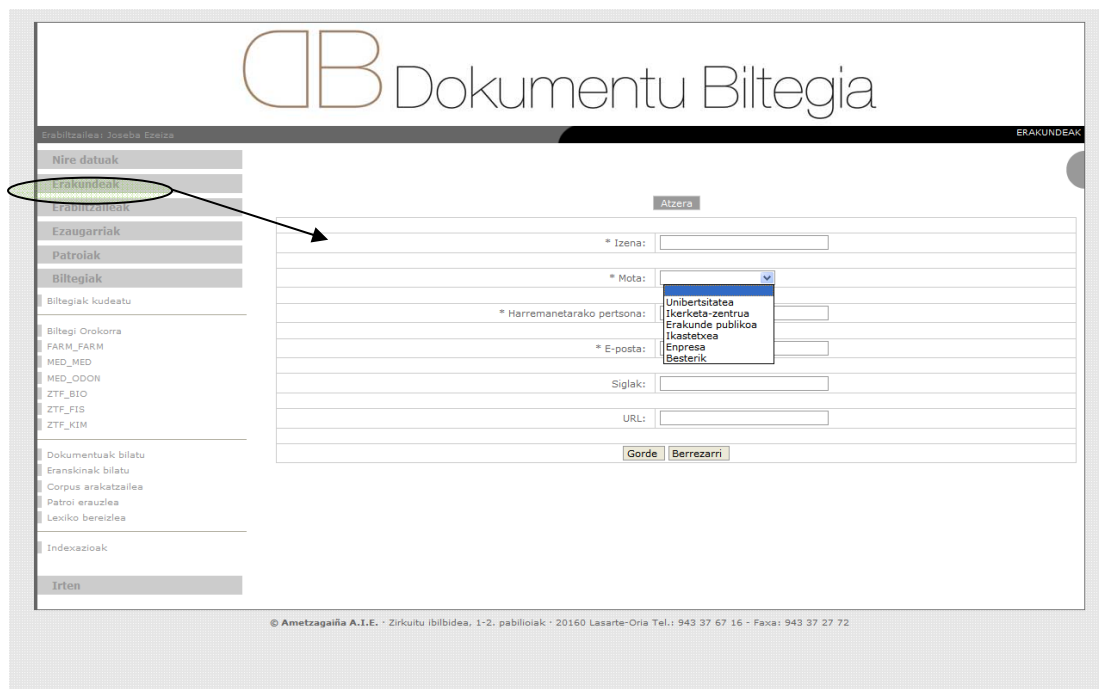
DBren administrazio-moduluaren bidez hiru lan egiten dira: a) biltegiak (corpusak) sortu, aldatu eta ezabatu; b) erabiltzaileei alta eman eta pribilegioak esleitu; eta c) testuak sailkatzeko erabiliko diren kategoriak zehaztu.

Biltegiak sortzeko, aldatzeko eta ezabatzeko aukera administratzaile orokorraren esku dago. Biltegien kudeaketa orokorra *Biltegiak Kudeatu* atalaren bidez egiten da. Biltegiak sortzeko nahikoa da formulario soil bat betetzea (ikusi 3. irudia). Biltegiaren ezaugarri buruzko oharra ere jaso daitezke inprimakian: zein den arduraduna, zeinek duten sarbidea biltegi horretara, zer helbururekin sortu den... Aktibo dauden biltegiak ezkerreko menuan agertuko dira zerrendatuta. Administratzaileak edozein momentutan alda ditzake biltegiaren izena edota fitxan jasotako datuak. Horretarako *Aldatu* funtzioa erabiliko du.



3. irudia: Biltegi berriak sortzeko inprimakia

Erabiltzaile bat sisteman alta emateko hiru urrats eman behar dira. Lehenbizi, zehaztu beharko da zein erakunde, ikastetxe edota taldetako den kide hori. Horretarako, *Erakundeak* izeneko formularioa erabiliko da (ikusi 4. irudia).



4. irudia: Erakundeak sisteman identifikatzeko inprimakia

Ondoren, erabiltzaileari identitate bat eta gako bat esleitu beharko zaizkio *Erabiltzaileak* izeneko formularioa erabilita (Ikusi 5. irudia). Azkenik, formulario berean, beheko aldean, zehaztu behar da zer pribilegio izango dituen erabiltzaile horrek sisteman aktibo dauden biltegi-tako bakoitzean; hau da, zehaztu behar da zein segmentutan sartu ahal izango den eta segmentu horietako bakoitzean zer rol jokatuko duen. Bost rol bereizten ditu DBk. Rolaren arabera, erabiltzaileak aukera izan dezake (edo ez) testuak kargatzeko, testuak sailkatzeko, testuak eta analisi-fitxak aldatzeko, testuak deskargatzeko edota kontsulta linguistikoak egiteko. Erabiltzaileei pribilegio ezberdinak esleiri dakizkieke biltegiaren arabera. Horrela, biltegi batean administratzaile rola duen erabiltzaile batek, beste biltegi batean karga egiteko aukera izan dezake soilik, eta beste batean kontsulta-pribilegioak baino ez ditu izango.

Sisteman sartzen denean, erabiltzaileak berari dagozkion biltegiak baino ez ditu ikusiko ingurunean. Lehen esan dugun bezala, biltegi horiek pantailaren ezkerreko aldean azalduko zaizkio, zerrenda batean antolatuta.

5. irudia: Erabiltzaileak sisteman identifikatzeko inprimakia

Administrazio-moduluak, azkenik, dokumentuak ezaugarritzeko erabiliko den kategoria-sistema egokitzeko atal bat ere eskaintzen du. Atal honetatik (ikusi 6. irudia), eskuz eransten, aldatzen edota ezabatzen dira testuak ezaugarritzeko erabiliko diren "etiketak". Funtzionalitate honi esker, erabiltzaileek beren beharretara egoki dezakete testuak sailkatzeko DBn proposatzen den oinarritzko taxonomia. Irudian ikus dezakegu "antolabide orokorra" izeneko eremuari dagokion aukeren zerrenda (narrazio-egitura, deskribapen-egitura, azalpen-egitura, eta abar). Formulario honen bidez, erabiltzaileak kategorien zerrenda hori laburtu, zabaldu eta aldatu dezake, beharrezko ikusten badu.

DB Dokumentu Biltegia

Erabiltzailea: Joseba Ezeiza EZAUGARRIAK KUDEATU :: ANTOLABIDE OROKORRA

Antolabide orokorra

Ordena	Aukera	Aldatu	Ezabatu
1	Narrazio-egitura	Aldatu	Ezabatu
2	Deskribapen-egitura	Aldatu	Ezabatu
3	Azalpen-egitura	Aldatu	Ezabatu
4	Argudio-egitura	Aldatu	Ezabatu
5	Instrukzio-egitura	Aldatu	Ezabatu
6	Egitura elkarreragilea	Aldatu	Ezabatu

© Ametzagaña A.I.E. · Zirkuitu ibilbidea, 1-2, pabiloiak · 20160 Lasarte-Oria Tel.: 943 37 67 16 · Faxa: 943 37 27 72

6. irudia: Dokumentuak sailkatzeko erabiliko diren kategoriak finkatzeko inprimakia

3. Karga- eta berreskuratze-modulua

Dokumentuen karga eta testuen berreskuratzea modu erraz eta intuitiboan egingo da DBn. Karga egiteko biltegi bat aukeratu behar dugu ezkerreko zerrendan dauden artean. Dokumentuak bilatu eta berreskuratze *Dokumentuak Bilatu* funtzioa erabiliko dugu.

Dokumentuen kargarako formularioa zazpigarren irudian ikus liteke. Agerikoenez, formulario horrek hainbat atal ditu (dokumentua, edukia, testuingurua, funtzioa eta egitura). Atalez atal dokumentuaren ezaugarrien multzo zabal bat gorde daiteke datu-basean (ikus 7. irudia). Datu horiek oso baliagarriak izango dira dokumentuak sailkatzeko eta berreskuratze. Edozein modutan, ezinbestez bete beharreko eremu bakarra dokumentuaren izenari dagokiona da.

7. irudia: Dokumentuak DBn kargatzeko modulua

Dokumentua DBra igo denean, bertan jasota gelditzen da, dagokion biltegian (ikusi 8. irudian). Dokumentuen fitxetan edozein unetan egin ahal izango dira aldaketak *Aldatu* funtzioa erabiliz. Dokumentuak beste biltegi batera ere eraman ahal izango dira *Beste biltegi batera kopiatu* funtzioaren bidez. Ezabatu ere nahi denean egin liteke. Dokumentuek izan dituzten gorabehera guztien historia *Historikoa* izeneko atalean gordetzen dira (noiz eta nork igo duen, zer aldaketa egin diren analisi-fitxetan, nork egin dituen aldaketa horiek...)

ZTF_KIM

Dokumentuak

Dokumentuak | Gaiak | Generoak | TEI esportazioa

Gehitu

Izena							
MaterialenZientziak_2.gai_a_riketak		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_2.gai_a_difusioa		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_3.gai_a_riketak		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_4.gai_a_riketak		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_4.gai_a_Prop.Mekanikoak		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_5.gai_a_riketak		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_5.gai_a_Prop.Termikoak		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_6.gai_a_Metalak_b		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_6.gai_a_Metalak_c		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_6.gai_a_Metalak_d		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_7.gai_a_zeramikoak		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_7.gai_a_zeramikoak_b		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_9.gai_a_Konpositeak		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	
MaterialenZientziak_9.gai_a_konpositeak_b		Beste biltegia kopiatu	Eranskinak	Historikoa	Aldatu	Ezabatu	

* Dokumentua ez da publikoa.

1

1

© Ametzagaña A.I.E. - Zirkuitu ibilbidea, 1-2. pabilioiak - 20160 Lasarte-Oria Tel.: 943 37 67 16 - Faxa: 943 37 27 72

8. irudia: Biltegian jasotako dokumentuen zerrenda

Dokumentu nagusiei nahi adina “eranskin” atxiki dakizkieke (irudiak, taulak, testu-eranskinak, bideoak, audioak...). Horretarako *Eranskinak* funtzioa erabiliko dugu (ikusi 9. irudia). Aukera hau oso baliagarria izan daiteke, adibidez, hipertestu-egitura erabiltzen duten dokumentuen kasuan edota hipermedia baliatzen dutenen kasuan ere. Aukera hau oso baliagarria izan daiteke, halaber, dokumentu itzulien kasuan, hizkuntza ezberdinetako bertsioak elkarrekin gordetzeko; baita dokumentu beraren hainbat bertsio batera gordetzeko ere. Audio- edo bideo-dokumentuen kasuan, transkripzioak ere erantsita joan daitezke funtzio honi esker. DBk ezin ditu, noski, ahozko dokumentuak prozesatu, baina, bai, adibidez transkripzioak. Azken kasu honetan, eranskinak atxikitze funtzio honek aukera ematen du bi materialak (ahozkoa eta idatzia) batera gorde eta kudeatu ahal izateko.

9. irudia: Dokumentu nagusiari eranskina atxikitzeko pantaila

Dokumentuak sailkatzeko DBk erabiltzen duen taxonomia finkatzeko hainbat iturri erabili dira: alde batetik, artikularen serreran aipatu dugun PREST proiektuaren emaitzak¹², eta, bestetik, gai honen inguruko hainbat lan (Alcaraz, Mateo eta Yus, 2007; Biber, 2004 eta 2006; Cassany, 2006; Ciapuscio, 1994, 2003 eta 2005; Ezeiza, 2008; Ezpeleta eta Gamero, 2004; García, 2000; García eta Gamero, 2003; Maldonado, 2003; Parodi, 2003, 2004, 2005, 2007-a, 2007-b eta 2009; Plazaola eta Alonso, 2007; eta Swales, 1990).

Lehen esan dugunaren arabera, taxonomia horrek egitura modularra du, eta lau alderdi hartzen ditu kontuan:

- Dokumentuaren edukia: gaia, azpi-gaia, gaiaren trataera, hitz gakoak, laburpena
- Dokumentuaren sorburuko testuingurua: eremua, arloa, komunikazio-mota, solaskidetza...
- Dokumentuaren ezaugarri funtzionalak: dokumentu mota, generoa, azpi-generoa...
- Dokumentuaren egiturazko ezaugarriak: antolabide orokorra, osagaiak...

Datu-multzo zabal hau jasotzea aukerakoa bada ere, oso baliagarria izan daiteke bilaketa aurreratuak egiteko. Horretarako, bilatzailearen interfazea erabili beharko da. Bilatzailea *Dokumentuak Bilatu* atalean aurkituko du erabiltzaileak (ikusi 10. irudia). Eranskinen artean

¹² <http://www.euskaraztrebatzen.org>

bilaketak egiteko aukera ere eskaintzen du DBk. Horretarako, formulario berezi bat eskaintzen da *Eranskinak Bilatu* atalean.

10. irudia: DBren bilatzailearen interfazea

Bilaketaren emaitzak XML formatuan esportatu ahal izango dira, TEI ereduaren arabera etiketatuta. Bilaketaren emaitzekin biltegi berri bat ere sortu ahal izango da (*Emaitzekin biltegia sortu* funtzioa erabilita). Azken funtzio honek erraztu egiten du ezaugarri ezberdinak dituzten testu-multzoen alderaketa linguistikoa. Egin behar dena da bilatzailearen laguntzaz bi testu-multzoak berreskuratu eta multzo bakoitzarekin biltegi berri bat sortu. Behin testu-multzoak bakoitza biltegi ezberdinetan bereizirik daudenean, erraz alderatu ahal izango dira DBk dituen analisi-tresnak baliatuta (Corpus Arakatzaila, Lexiko Bereizlea, Patroi Erauzlea, eta abar).



11. irudia: Bilaketaren emaitza ematen duen pantaila

4. Kontsulta-modulua

Dokumentu Biltegiak kontsulta linguistikoak egiteko hiru tresna ditu integratuta: Corpus Arakatzaila, Lexiko Bereizlea eta Patroi Erauzlea. Tresna horiek baliatu ahal izateko, aurrez testuak prozesatu egiten dira DBra kargatzen direnean. Prozesu horrek hainbat ataza hartzen ditu barne: testua HTML bihurtu, normalizatu eta datu-basean gorde; testua txt formatura aldatu; testua tokenizatu eta lematizatu; elementuen maiztasun-taulak eguneratu; monogramak, bigramak eta trigramak erauzi; desanbiguazio lexiko eta morfologikoa egin; kategoria eta patroi morfosintaktikoak markatu; eta abar. Azkenik, informazio hori guztia indexatu egiten da, kontsultak egiteko unean atzipena errazteko.

Erabiltzaileak baimendutako biltegi guztietan egin ditzake kontsultak. Baina, DBk aukera eskaintzen du, halaber, kontsultak murrizteko ezaugarri jakinak dituzten dokumentuetara (erabiltzaileak finkatuko ditu ezaugarri horiek). Horretarako, DBren datu-basea baliatzen da kontsulta interfazeetan eskaintzen diren inprimakien bitartez. Kontsulten emaitzak, bestalde, XLS formatuko fitxategietara esporta daitezke. Modu horretara, errazagoa gertatzen da datuen bilketa eta analisisa, batez ere, datu ugari eskuratu eta landu nahi direnean.

Corpus Arakatzaila hainbat kontsulta mota egiteko aukera ematen du. Kontsulta sinpleak hitzen, lemen edota kategoria gramatikalaren arabera egin litezke. Emaitzak konkordantzia gisa eskuratu ahal izango dira, edo, hala behar izanez gero, testuinguru zabalagoak ere berreskuratu ahal izango dira. Ondoren dugun irudian (ikus 12. irudia) ikus ditzakegu *larik* atzizkiaren erabilera aztertzeke egin den kontsultaren emaitzetako batzuk.

Erabiltzailea: Joseba Ezeiza

CORPUS ARAKATZAILEA

Nire datuak

Erakundeak

Erabiltzaileak

Ezaugarriak

Biltegiak

Irten

Bilaketa berria

Dokumentu kopurua: 22 (99 agerpen denera)

BILTEGIA: ZUZENBIDE PRIBATUA

»Erregistro Zibilarren Legea eta erregelamendua (2 agerpen)

1. hemen ahaleginaren lerro nagusia: euskara ofiziala non izan eta lurralde horietako herritarrei halako tresna sendoa eskaintzea, alor horretako hizkuntzaren normalizazioan bultzagile izateko gogoia agerian uzten **delarik**.
2. d) Alderatu eta bateratu egin dira, ahal **delarik**, lege-testu hauen terminoak eta Erregistro Zibilarren formularioetan Agiri Judizialen Itzupenerako Batzordeak ezarritakoak. Bion ahaleginak egokitu dira, eta berebiziko gogopeta egin da, modu bertsuan jokatzeko. Hartara,

»Foru Lege Zibilak (5 agerpen)

1. a) Zibil mailan instantzia eta gradu guztietara, Euskadiko Zuzenbide Zibil Foralaren gaietan lasazio eta berrikuskatzeak errekutuzkoak baina **diralarik**.
2. Zuzenbide Zibilarren baterako bereritasunak agertuz joango dira eurotan sakondu eta aztertu ahal, oraingoz idatziz dauden sistemek eta euren ezarpen-esparru zehatzek eragin dezaketen erakarpen-indarra ekidin ezinezkoa **delarik**.
3. ez zituen beste batzuk, 11. atalean agertzen diren oinordetzako itunak edo Aiarako podere osoko gozamina adibidez, behar bestean garatu beharra ezaren, azken hori garapen-egazatik ez **delarik** benetan ezarriari izan.
4. eta Bizkaian agiri-bidez dastozkigunetan baserri eskualdatzeko erarik erabiliena oinordetza-itunak izsteak, beharrezko egiten dute itunok legez arautzea, Bildumaren 11. atalak egiten zuen aipamen hutza nahiko ez **delarik**.
5. aitzina, oro har, elikagaietarako eskubidea amaitzeagatik eta horren onuraduna ezkontzen den momentutik; edo ezkontzak bazala bizitzen hasten denetik aitzina, b letraren kasuan aldiz, arrazola edozein **delarik** ere seme-alabenganako ardura eteten denean edo horiek adinez nagusi izatera iritsi edo emantzipatzen direnean; salbuespen izanen dira egaitasun kasuak.

»Hipoteak Legeria2 (1 agerpen)

1. behar da: ondasun higiezinak 526. artikuluko ziurtagiriaren arabera sortutako errenta 100eko 3tan kapitalizatuta, horren salmenta-balioa zama guztien bikoitza izatea, zamen berruz fidantza berria ere sartzen **delarik**.

»Hipoteka Legeria (1 agerpen)

1. behar da: ondasun higiezinak 526. artikuluko ziurtagiriaren arabera sortutako errenta 100eko 3tan kapitalizatuta, horren salmenta-balioa zama guztien bikoitza izatea, zamen berruz fidantza berria ere sartzen **delarik**.

BILTEGIA: HIZTEGI BATUA

»Hiztegi Batua (1 agerpen)

1. deitzeako hitza to 2 'tori' tobera tobera-mustra tobera (hizkuntza) togoar (herritarra) toli 1 Mustelus mustelus toli 2 Izond. Batez ere Ipar, 'pisua, motela'; gorputza toli **delarik**, loa uxatzea zail da, toildu, toli/toildu, toiltzen; da/du ad. tok pisin (hizkuntza) toka 1 adlag, 'gizoneko bati hika' toka 2 iz. 'jokoa' tokamen Heg.: irabazietan

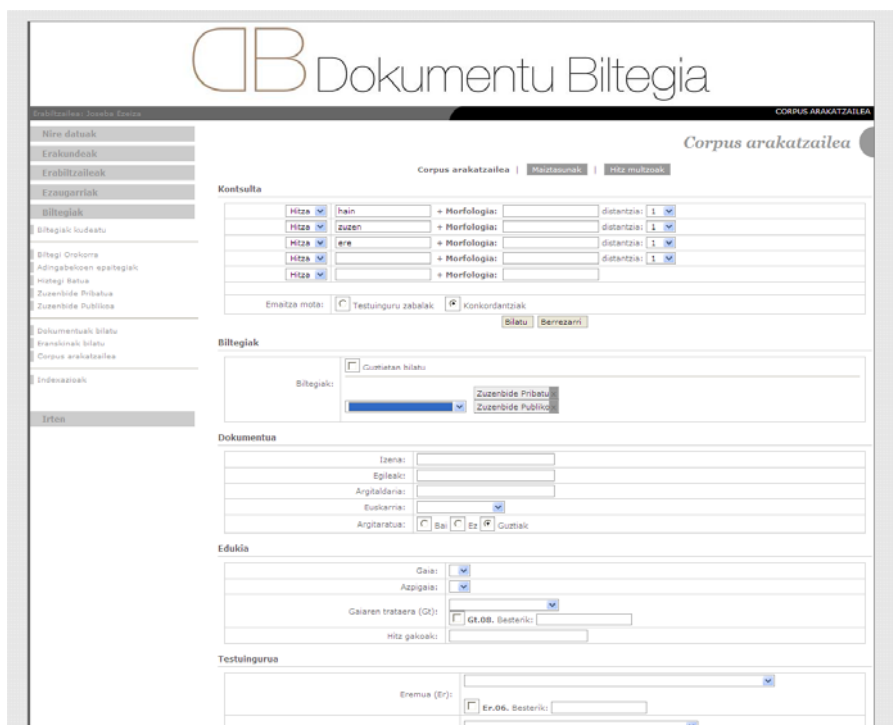
BILTEGIA: ZUZENBIDE PRIBATUA

»Jabetza Intelektualari Buruzko Legea (12 agerpen)

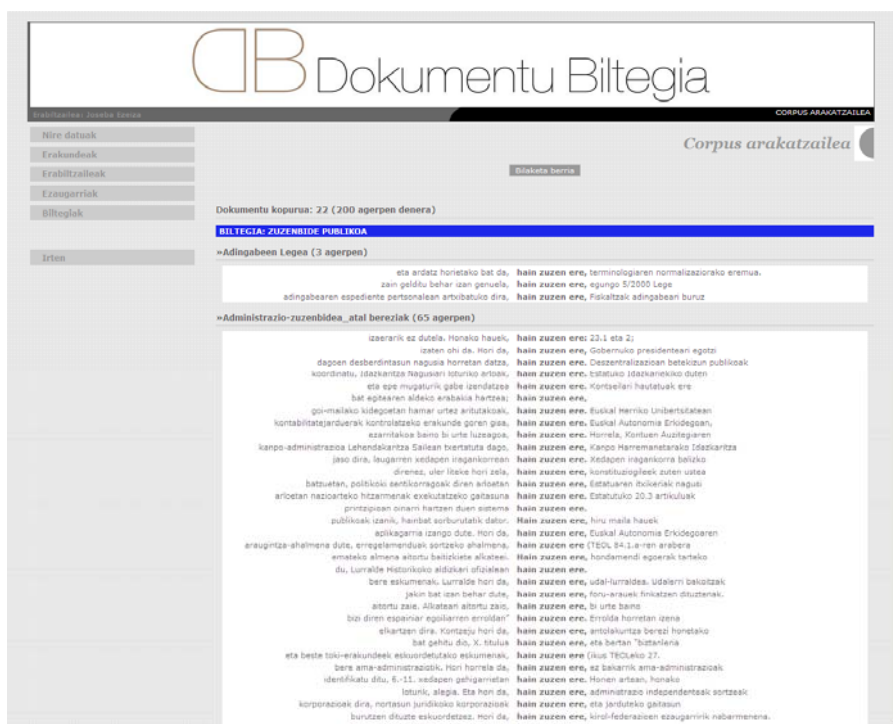
1. a) Beren jardueran ari **diralarik** obrak, prestazio artistikoak, fonogramak edo bideoogramak erreproduzitzeko nahitaezkoa den baimena dutenek eskuratutako ekipoa, tresnak eta euskarri materialak. Baimen horren egiaztapena erakutsi beharko diete zordunei eta,
2. Aurreko paragrafoan xedaturikoa aplikatzekoa izango da edozein aldaketa gertatzen **delarik**, dela ordezkarri bakarraren edo osaturiko elkartearen pertsonan, dela haien egoitzetan, dela enbitate kudeatzaileen kopuruan nahiz motan, ordezkatuak zein elkartuak izan enbitateak, Orobat, izango da aplikatzen
3. Independentiak ez diren zabi, liburuak, entsegu edo faszikulutan jendarteratutako obraren kasuan, eta haien babes-apea hainke zifre den moduan jendarteratzen direnetik zenbatzen hasten **delarik**, elementu bakoltzarentzat bereiz zenbatuko da epe hori.
4. Baldin eta, egilearen heriotzaren edo heriotza-deklarazioaren ondoren, haren eskubidedunek obra ez jendarteratzeko eskubidea baliatzen badute, Konstituzioaren 44. artikuluan xedaturikoa urratzen **diralarik**, epaileak degozkion neurriak agindu ahal izango ditu, Estatuaren, autonomia-erikidegoen, toki-korporazioen, izatera kulturalako instituzio publikoen edo interes legitimoko duen beste edozein pertsonaren eskariz.
5. e) Baldin eta, argitaraldi bat baino gehiago aurreikusit **diralarik** eta egindako azkena agortu, argitaratzaileak ez badu hurrengo argitaraldia egiten egileak

12. irudia: Kontsulta arruntaren emaitzak

Corpus Arakatzailleak kontsulta aurreratuak egiteko aukera ere ematen digu. Horrela bada, 13. irudian ikus litekeen moduan, Corpus Arakatzailleak bost hitz, lema edota kategoria gramatikalen sekuentziak ere berreskura ditzake. Adibide honetan *hain zuzen ere* antolatzailearen erabilera-testuinguruen kontsulta egiteko modua erakusten da. Bilaketa murriztu egin da, eta DBri eskatu zaio kontsulta hori zuzenbide publikoaren eta zuzenbide pribatuaren inguruko biltegiatan egiteko. Kontsultaren emaitzetako batzuk 14. irudian agertzen dira.



13. irudia: Kontsulta aurreratuaren adibidea



14. irudia: Kontsulta aurreratuaren emaitzak

Corpus Arakatzailera badu beste modulu bat hitzen eta lemen agerpen-maiztasunaren datuak ematen dituena. Datu horiek taula-formatuan eskaintzen dira, eta zutabe zutabe kontsultan hautatu ditugun biltegiei dagozkien emaitzak ematen dira bereizirik (ikusi 15. irudia). Taula hori XLS formatura esportatzeko aukera ere ematen da. Emaitzen testuinguruak

Ikusi funtzioaren bidez berreskura daitezke aparteko leiho batean, pantailatik irten gabe (ikusi 16. irudia). 15. eta 16. irudietako adibidean, *kuntza* atzikia duten lemen agerpen-datuen kontsultaren emaitzak ikusten dira. Kontsulta hau zuzenbide alorreko testuen hiru biltegitan egin da.

The screenshot shows the 'DB Dokumentu Biltegia' interface. On the left is a navigation menu with categories like 'Nire datuak', 'Erakundeak', 'Erabiltzaileak', 'Ezaugarriak', 'Biltegiak', 'Biltegiak kudeatu', 'Biltegi Orokorra', 'Adingabekoen epaitegiak', 'Hiztegi Batua', 'Zuzenbide Pribatua', 'Zuzenbide Publikoa', 'Dokumentuak bilatu', 'Eranskinak bilatu', 'Corpus arakatzailerak', 'Indexazioak', and 'Irten'. The main content area displays a table titled 'Maiztasunak' (Frequency) with columns: 'Bilaketa', 'Lemak', 'Agerpen kopurua', 'Maiztasuna (10.000ko)', 'Zuzenbide Publikoa', 'Zuzenbide Pribatua', and 'Adingabekoen epaitegiak'. The table lists various forms of the word 'kuntza' and their frequencies across different corpora. A 'Guztira:' (Total) row is at the bottom of the table.

Bilaketa	Lemak	Agerpen kopurua	Maiztasuna (10.000ko)	Zuzenbide Publikoa	Zuzenbide Pribatua	Adingabekoen epaitegiak
Ikusi	aldakuntza	45	0.36	23	22	0
Ikusi	antolakuntza	155	1.23	153	2	0
Ikusi	asmakuntza	2	0.02	0	2	0
Ikusi	aurkikuntza	14	0.11	2	12	0
Ikusi	aurrerakuntza	4	0.03	2	2	0
Ikusi	autoikaskuntza	1	0.01	1	0	0
Ikusi	berreizkuntza	5	0.04	1	4	0
Ikusi	berokuntza	1	0.01	1	0	0
Ikusi	berrikuntza	109	0.86	29	80	0
Ikusi	doikuntza	1	0.01	1	0	0
Ikusi	ebakuntza	4	0.03	0	4	0
Ikusi	eraikuntza	84	0.67	42	42	0
Ikusi	eraldakuntza	1	0.01	1	0	0
Ikusi	ernalkuntza	2	0.02	0	2	0
Ikusi	errakuntza	26	0.21	21	5	0
Ikusi	gaikuntza	55	0.44	30	25	0
Ikusi	gehikuntza	28	0.22	6	22	0
Ikusi	hazkuntza	4	0.03	1	3	0
Ikusi	hezuntza	510	4.04	475	34	1
Ikusi	hizkuntza	385	3.05	197	188	0
Ikusi	hobekuntza	209	1.65	36	173	0
Ikusi	hornikuntza	4	0.03	1	3	0
Ikusi	ikaskuntza	38	0.30	35	3	0
Ikusi	ikerakuntza	9	0.07	8	1	0
Ikusi	irakaskuntza	293	2.32	263	30	0
Ikusi	medikuntza	6	0.05	1	5	0
Ikusi	normalkuntza	3	0.02	1	2	0
Ikusi	prestakuntza	122	0.97	112	10	0
Ikusi	saiakuntza	2	0.02	0	2	0
Ikusi	sorkuntza	8	0.06	5	3	0
Ikusi	trebakuntza	6	0.05	4	1	1
Ikusi	zabalkuntza	7	0.06	2	5	0
Ikusi	zapalkuntza	1	0.01	1	0	0
Guztira:		2144	17.01	1455	687	2

© Ametzagaña A.I.E. · Zirkuitu ibilbidea, 1-2, pabilioiak · 20160 Lasarte-Oria Tel.: 943 37 67 16 · Faxa: 943 37 27 72

15. irudia: Agerpen-maiztasunen kontsultaren emaitza-taula

DB Dokumentu Biltegia

Corpus Arakatzaileria

Maiztasunak

Bilaketa	lema	Agerpen kopurua	Maiztasuna (10.000ko)	Zuzenbide Publikoa	Zuzenbide Pribatua	Adingabelekin epategiak
Ikusi	aldakuntza	45	0.36	23	22	0
Ikusi	antolakuntza	155	1.23	153	2	0
Ikusi	asmakuntza	2	0.02	0	2	0
Ikusi	aurkikuntza	14	0.11	2	12	0
Ikusi				2	2	0
Ikusi				1	0	0
Ikusi				1	4	0
Ikusi				1	0	0
Ikusi				29	90	0
Ikusi				1	0	0
Ikusi				0	4	0
Ikusi				42	42	0
Ikusi				1	0	0
Ikusi				0	2	0
Ikusi				21	5	0
Ikusi				30	25	0
Ikusi				6	22	0
Ikusi				1	3	0
Ikusi				475	34	1
Ikusi				197	108	0
Ikusi				36	173	0
Ikusi				1	3	0
Ikusi				35	0	0
Ikusi				8	1	0
Ikusi				283	30	0
Ikusi				1	5	0
Ikusi				1	0	0
Ikusi				112	10	0
Ikusi				0	2	0
Ikusi				5	3	0
Ikusi				4	1	0
Ikusi				2	5	0
Ikusi				1	0	0
Ikusi				1455	987	2

Bilaketa: asmakuntza
Dokumentu kopurua: 3 (3 agerpen denera)

BILTEGIA: ZUZENBIDE PRIVATUA

Kode Zibila (1 agerpen)

1. Haren ondarearen inguruan, Kode Zibila arautzearen, behar besteak ziren jardura honen arauterako arindua. Zehazki den ere, berorak aurrera egin ahal, agindua, euren eraginarekin jarduera jarduera. **Asmakuntza** baimen ondorio, komunikabideak eratu egin dira, eta horiek, beraz, ezin du bermari gero eta askarrago jaso, eta gozate-baitzen dabilen harenaren: gero eta gehiago eratu.

Lege Zibila: Hondsun higiezinak (1 agerpen)

1. Itzulpuntu junditoki, figura horiek ondarearen luperatze eskubidea denboran zehar zatitza dalar. 1960ko urtarraren 16ko Errege Dekretuaren arabera, **Asmakuntza-patestea** inakribatu zen; poliza honen bidez, titularen aurreneko bittegi bat egi finko edo aldakortan luperatze eskubidea erakiko zitzen. Hala ere, ondarearen higiezin funtzio-antzerako sistema hon ezaugarri.

Zigor Kodea (1 agerpen)

1. Sei hilabeteko lu urte arteko espekzio-egitura ezarriko zena, bai eta sei hilabeteko hopena lau arteko izate ere, zehazki-egituraren ondorengo, zehazki-egituraren eskubidearen ga den **asmakuntza** hala nahia zabalzen duenari, baldin eta zabaluntza honi nazio-defentaren kalteen bidez.

16. irudia: Testuinguruen kontsulta *Ikusi* funtzioaren bidez

Corpus Arakatzaileraren hirugarren moduluak, azkenik, hitz edo lema jakin baten bigramak eta trigramak berreskuratzen ditu (ikus 17. irudia). Bilaketa hauek ere murriztu egin daitezke, hainbat gauza zehaztuta: kontsultatu nahi dugun hitza edo lema bigramaren edota trigramaren zein tokitan agertuko den, gutxieneko maiztasunak, eta abar. Aurreko kasuan bezala, kasu honetan ere biko eta hiruko hitz-multzo horien testuinguruak pantailatik bertatik berreskura daitezke zuzenean. Irudian dugun adibidean *dakioke* aditz laguntzailearen bigramak eta trigramak kontsultatu dira.

DB Dokumentu Biltegia

Corpus Arakatzaileria

Hitz multzoak

Bikoteak

Bilaketa	Bikotea	Agerpen kopurua	Maiztasuna (10.000ko)	Zuzenbide Publikoa	Zuzenbide Pribatua	Adingabelekin epategiak
Ikusi	eska dakioke	12	0.10	4	8	0
Ikusi	jar dakioke	12	0.10	4	8	0
Ikusi	aplica dakioke	8	0.06	4	4	0
Ikusi	eman dakioke	7	0.06	4	3	0
Ikusi	egin dakioke	6	0.05	4	2	0
Ikusi	egin dakioke	5	0.04	0	5	0
Ikusi	aurkez dakioke	3	0.02	3	0	0
Ikusi	lortu dakioke	2	0.02	0	2	0
Ikusi	erakiko dakioke	2	0.02	1	1	0
Ikusi	eskuorde dakioke	2	0.02	0	2	0
Ikusi	ezar dakioke	2	0.02	1	1	0
Ikusi	egin dakioke	2	0.02	0	2	0
Ikusi	adurdua dakioke	2	0.02	0	2	0
Ikusi	dustea	65	0.55	25	40	0

Hirukoteak

Bilaketa	Hirukotea	Agerpen kopurua	Maiztasuna	Zuzenbide Publikoa	Zuzenbide Pribatua	Adingabelekin epategiak
Ikusi	erreturtoze jar dakioke	7	0.06	1	6	0
Ikusi	egin eska dakioke	5	0.04	1	4	0
Ikusi	lortu eska dakioke	2	0.02	0	2	0
Ikusi	justiza-ministroen jar dakioke	2	0.02	0	2	0
Ikusi	bekarri, egin dakioke	2	0.02	0	2	0
Ikusi	idatzi eska dakioke	2	0.02	0	2	0
Ikusi	edo eska dakioke	2	0.02	0	2	0
Ikusi	ekin egin dakioke	2	0.02	0	2	0
Ikusi	erreturtoze aurkez dakioke	2	0.02	0	2	0
Ikusi	kalte-ondare eman dakioke	2	0.02	0	2	0
Ikusi	hori eska dakioke	2	0.02	0	2	0
Ikusi	ekin eskuorde dakioke	2	0.02	0	2	0
Ikusi	dustea	32	0.30	4	28	0

© Ametragoia A.S.E. - Zirkulu Biltegia, I.G. jarkitzaile; 2010ko Lezarte-Oriz Tel.: 943 37 87 16 - Faxa: 943 37 27 72

17. irudia: Bigramen eta trigramen kontsultaren emaitza

Aurkeztu berri ditugun hiru tresna horiek Ametzagaña I+G ikerketa-guneak aurrez garatuak dira; hau da, ez dira izatez berriak. Hori bai, DBren egituran integratuta, funtzio eta erabilera berriak bilatu zaizkie tresna hauei, besteak beste, kontsultak murrizteko aukera ematen duten formularioak erantsiz eta datuak esportatzeko modua erraztuz. Baina, hiru horiez gain DBk beste bi kontsulta-tresna ere eskaintzen ditu: Patroi Erauzlea eta Lexiko Bereizlea. Bi hauek originalak dira eta proiektu honen testuinguruan garatu dira.

Patroi Erauzleak hainbat elementuz osatutako patroi morfosintaktikoak erauzteko aukera ematen du. Patroi horiek eskuz finkatzen ditu erabiltzaileak *Patroiak* atalean duen formularioa erabilia (ikusi 18. irudia). Besterik ezean, DBk balio terminologikoa izan dezaketen patroi batzuen zerrenda eskaintzen ditu; hain zuzen, Elosegik (2002) bere doktore-tesian identifikatutakoak. Baina, esan dugun moduan, DBren erabiltzaileek aukera dute zerrenda hau egoki ikusten duten moduan aldatzeko edota zabaltzeko, beren premien arabera.



18. irudia: Patroi morfosintaktikoak finkatzeko formularioaren interfazea

Hurrengo irudian [izena + adjektiboa] patroiarene bilaketaren emaitzetako batzuk ikus litezke (ikusi 19. irudia). Bertan agertzen diren emaitzak medikuntzaren alorreko testuei dagozkie: *nodo sinusal*, *egoera patologiko*, *zelula automatiko*, *zelula borobil*, *bolumen sistoliko*, *takikardia sinusal*, eta abar. Bilaketa hauek, tarteka, nolabaiteko “zarata” ere eman dezakete, zenbait elementuren kategoria gramatikalaren anbiguotasunaren ondorioz: *bihurri urrun*, *mota bihur*, *kliniko orokorra*, eta abar.

19. irudia: Patroi Erauzlearen bidez egindako kontsulta baten emaitzen pantaila

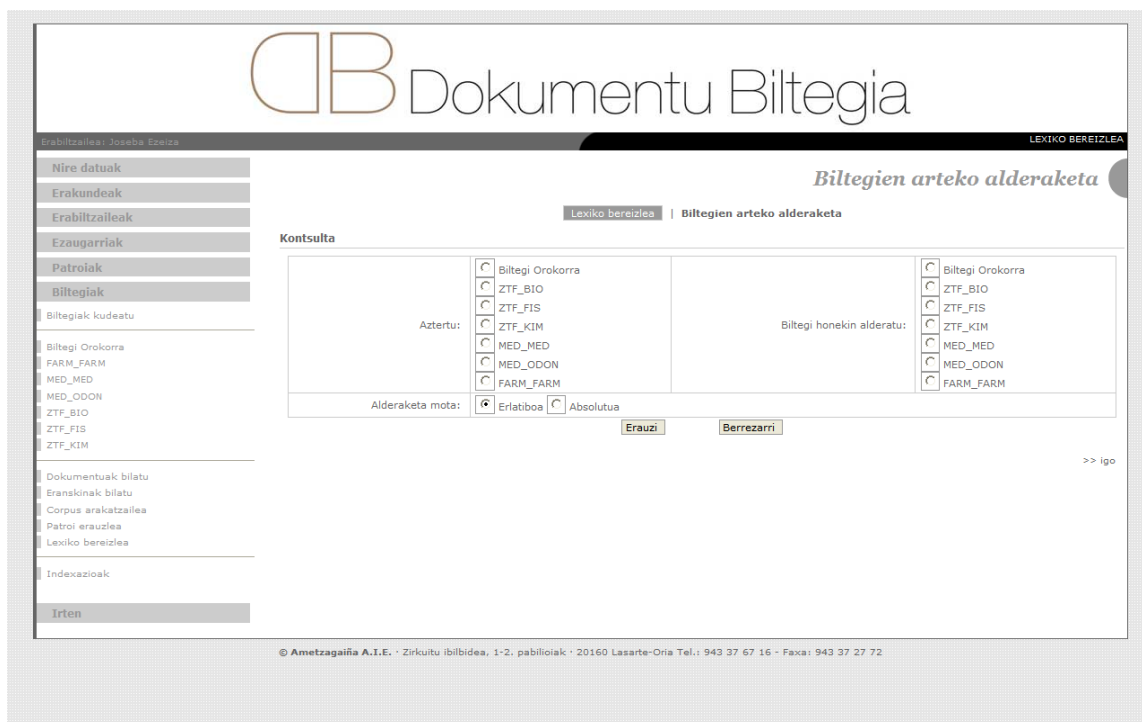
Lexiko Bereizlea, azkenik, corpus baten elementu lexiko bereizgarriak erauzten dituen tresna da. Tresna honek ezaugarritu nahi den corpora beste corpus batekin alderatzen du. Modulu estatistiko baten bidez, elementu lexiko bereizgarrienak identifikatzen ditu, eta beren esangarritasun estatistikoaren arabera zerrendatzen ditu. Hiru alderaketa mota egin daitezke Lexiko Bereizlea erabiliz.

- Lehenik eta behin, Ereduzko Prosa Gaur¹³ corpusaren maiztasun datuekin egin liteke alderaketa (ikus 20. irudia). Alderaketa honen bitartez, testu orokorretan agertu ohi ez diren elementu lexikoen eta testu orokorretan erabili arren, ezaugarritu nahi den corpusean agerpen-maiztasun nabarmen ezberdinak dituzten elementuen zerrenda erauziko du Lexiko Bereizleak.
- Bestetik, Lexiko Bereizleak aukera ematen digu ezaugarritu nahi dugun corpora DBn dagoen beste edozein corpusen kontra alderatzeko. Aukera honi esker ("alderaketa absolutua") hainbat alderaketa jakingari egin daitezke: adibidez, zuzenbide pribatuko corpus bat zuzenbide publikoko corpus baten kontra; kimika organikoko corpus bat kimika ez-organikoko beste baten kontra; eta abar. Modu honetara corpus horietan aldea egiten duten elementu lexikoak eskuratu ahal dira bi norabideetan (ikus 21. irudia).
- Azkenik, Lexiko Bereizleak aukera ematen digu bi corpusen arteko alderaketa "erlatiboa" egiteko. Kasu honetan, elementu lexikoen erauzketa testu orokorren maiztasun-zerrendaren bahetik ere pasatzen da. Erauzketa honen emaitza alderaketa absolutuarena baino mugatuagoa da, bazterrean uzten baititu testu orokorretan antzeko agerpen-maiztasunak dituzten elementu lexiko guztiak.

¹³ <http://www.ehu.es/euskara-orria/euskara/ereduzkoa>



20. irudia: Lexiko Bereizlean oinarritzko kontsultak egiteko pantaila



21. irudia: Lexiko Bereizlean kontsulta aurreratuak egiteko pantaila

22. irudian ikus liteke bilaketa baten emaitza. Ezkerreko leihatilan erazutako elementuen zerrenda agertzen da (monogramak); eskuinekoan, aldiz, elementu horiei dagozkien bigramak eta trigramak. Kasu honetan ere pantailatik irten gabe berreskura daitezke elementu horien guztien testuinguruak. Eta pantailatik bertatik agindu daiteke, halaber, datuen esportazioa XLS fitxategietara.

DB Dokumentu Biltegia

Erabiltzailea: Joseba Ezeiza LEXIKO BEREIZLEA

Nire datuak
Erakundeak
Erabiltzaileak
Ezaugarriak
Patroiak
Biltegiak
Biltegiak kudeatu

Biltegi Orokorra
FARM_FARM
MED_MED
MED_ODON
ZTF_BIO
ZTF_FIS
ZTF_KIM

Dokumentuak bilatu
Eranskinak bilatu
Corpus arakatzaila
Patroi erazlea
Lexiko bereizlea

Indexazioak

Irten

Biltegien arteko alderaketa

Lexiko bereizlea | Biltegien arteko alderaketa
Bilaketa berria

Alderaketa: "ZTF_BIO" versus "ZTF_FIS"

Emaitzak: <TXT moduan> <Excel moduan>

- proteina: 1.0000
- gene: 0.4181
- lotu: 0.4036
- molekula: 0.4007
- sintesi: 0.3827
- erreakzio: 0.3544
- sekuentzia: 0.3211
- entzima: 0.3130
- mintz: 0.2611
- lotura: 0.2319
- gune: 0.2151
- egitura: 0.2024
- de: 0.1945
- erregulazio: 0.1880
- sintetizatu: 0.1843
- eratu: 0.1814
- prozesu: 0.1806
- desberdin: 0.1775
- konplexu: 0.1753
- garraio: 0.1733
- edo: 0.1728
- azido: 0.1708
- glukosa: 0.1671
- aminoazido: 0.1643
- transkripzio: 0.1626
- funtzio: 0.1621
- bakotz: 0.1589
- kromosoma: 0.1542
- genetiko: 0.1469
- genoma: 0.1468
- espezifik: 0.1454

gene ekoizkin

- geneen ekoizkinak (2) <arakatu>

gene eraketa

- geneen eraketa (1) <arakatu>

gene erregulazio

- geneen erregulazioa (1) <arakatu>
- geneen erregulazioarekin (1) <arakatu>

gene espresio

- geneen espresioa (3) <arakatu>
- geneen espresioaren (6) <arakatu>
- geneen espresiorako (1) <arakatu>

gene funtzio

- geneen funtzioa (1) <arakatu>
- geneen funtzioak (1) <arakatu>

nona hiru

© Ametzagaña A.I.E. · Zirkuitu ibilbidea, 1-2, pabilioiak · 20160 Lasarte-Oria Tel.: 943 37 67 16 - Faxa: 943 37 27 72

22. irudia: Lexiko Bereizlean kontsulta aurreratua egiteko pantaila

22. irudiaren adibidean biokimikako corpus bat aztertu da optikako beste corpus baten kontra. Biokimikako corpusaren lexiko bereizlearen zerrendan honelako elementuak agertu dira esangarrienen artean: *proteina, gene, molekula, sintesi, erreakzio, sekuentzia, entzima, mintz, lotura, gune sintetizatu...* Optikako corpusaren lexiko bereizlearen zerrendan, aldiz, oso bestelako elementuak aurki daitezke esangarrienen artean: *cm, irudi, distantzia, kalkulatu, intentsitate, objektu, uhin, argi, irekidura, foku, angelu, perpendikular, prisma...* Ageriko denez, oso tresna lagungarria izan daiteke Lexiko Bereizlea terminologia erazi eta aztertzeke, besteak beste.

5. Dokumentu Biltegia aurrera begira

Lerro hauetan erakutsi dugu hiru urteko ibilbidearen emaitza. Hasieran ideia lauso bat basterik ez zena, itxuratzen eta gauzatzen joan da arian-arian, eta une honetan erabiltzeko moduan da jada. Tresna honek aukera berriak eskaintzen ditu esparru akademikoan egiten den euskarazko ekoizpenaren azterketari ekin ahal izateko. Izan ere, DB dokumentuak biltzeko, prozesatzeko eta aztertzeke baliabide malgua eta erabilgarria da, testu-masa handiak neke handiegirik gabe kudeatzeko aukera ematen duena.

Orain, tresnaren funtzionaltasuna ebaluatu behar da, eta, gutxienez, bi alderdi aztertu behar dira: zer nolako ikerketak egiteko izan daitekeen egokia, eta nola aprobetxa litekeen espezialitate-hizkeren irakaskuntzan. Ebaluazio horren ondorioek esango digute zer etekin atera ahal diogun DBri, euskararen garapen lexiko-diskurtsiboaren dinamika esparru akademikoetan bizkortu eta indartzeari begira.

Bitartean ere, DB laugarren fasean sartuko da aurki. Fase horretan bi modulu osagarri erantsi nahi zakizkio: datu-base lexiko bat (GlosaTegia) eta bestelako informazio linguistikoa jasotzeko beste datu-base bat (Estilo Lantegia). Bi modulu horiek DBn jasotako testuetan detektatzen diren jokabide eta ezaugarri linguistikoak dokumentatu eta administratzeko balioko lukete, hau da, azterketa linguistikoaren emaitzak biltzen joateko eta komunitatearen esku jartzeko. Bosgarren fasean, azkenik, beste bi biltegi sortu nahi dira: bata administrazioaren eta zuzenbidearen alorretako dokumentuak kudeatzeko, eta bestea komunikabideetako dokumentuak jaso eta horiekin lan egiteko.

Azken jomuga hau da: azpiegitura teknologiko egokia izatea, espezialitateko testuetan egiten den euskararen erabileraren inguruan, jakintza-banku aktibo, dinamiko eta sendo bat abian jartzeko. Gauzak okertzen ez badira, 2010. urtearen amaierarako egongo da azpiegitura osoa prest. Beraz, laster izango gara proiektu horri bultzada emateko moduan.

Bibliografia

- ALCARAZ, M.; YUS (arg.) (2007): *Las lenguas profesionales y académicas*, Bartzelona: Ariel.
- BIBER, D. (2004): *Dimensions of Register Variation. A Cross-linguistic Comparison*, Cambridge: Cambridge University Press.
- BIBER, D. (2006): *University language: A corpus-based study of spoken and written registers*, Amsterdam: John Benjamins.
- CASSANY, D. (2006): *Taller de textos. Leer, escribir y comentar en el aula*, Bartzelona: Paidós.
- CIAPUSCIO, G. (1994): *Tipos textuales*, Buenos Aires: Eudeba.
- CIAPUSCIO, G. (2005): "La noción de género en la Lingüística Sistémico Funcional y en la Lingüística Textual", *Rev.Ssignos*.57, 31-48.
- ELOSEGI, A. (2002): *Zuzenbideko euskal hizkera berezia. Lege-corpus itzuli bateko terminologiaren azterketa linguistikoa eta terminologia*, Bilbo: UPV/EHU
- EZEIZA, J. (2008): "Especialitate hizkeren didaktika europako unibertsitate eremuan. Gaitasunetara zuzendutako irakasbidea", in UGARTEBURU, I. eta SALABURU, P. (arg.): *Especialitate Hizkerak eta Terminologia III*, Bilbo: UPV/EHU.
- EZPELETA, P.; GAMERO, S. (2004): "Los géneros técnicos y la investigación basada en corpus: proyecto GENTT", in GASER, R., GUIRADO, C. eta REY, J. (arg.): *Insights into Scientific and Technical Translation*, Bartzelona: PPU, 147-156
- GARCÍA, I.; MONZÓ, E. (2003): "Corpus de géneros GENTT. Una enciclopedia para traductores", *Traducción y Comunicación* 4. 31-53.
- GARCÍA, I., (2000): *Análisis textual aplicado a la traducción*, Valentzia: Tirant Lo Blanch.
- MALDONADO, M. (2003): *Texto y comunicación*, Madril: Fundamentos.
- PARODI, G (2004): "Textos de especialidad y comunidades discursivas técnico-profesionales: una aproximación basada en corpus computarizado", *Estudios Filológicos* 39, 7-36.
- PARODI, G. (arg.) (2005): *Discurso especializado e institucionse formadoras*, Valparaiso: Universidad Pontificia de Valparaiso.

- PARODI, G. (arg.), (2007): *Working with Spanish corpora*, London: Continuum.
- PARODI, G. (2005): "Discurso Especializado y Linguística de Corpus: Hacia el desarrollo de una competencia psicolingüística", *Boletín de Linguística*.23, 61-88.
- PARODI, G. (2007): "El discurso especializado escrito en el ámbito universitario y profesional: Constitución de un corpus de estudio", *Rev. Signos*, 63.
- PARODI, G. (2009): "El Corpus Académico y Profesional del Español PUCV-2006: semejanzas y diferencias entre los géneros académicos y profesionales", *Estudios Filológicos* .44., 123-147.
- PARODI, G.; GRAMAJO, A. (2003): "Los tipos textuales del corpus técnicoprofesional PUCV 2003: una aproximación multiniveles", *Rev. Signos*..54, 207-223 .
- PLAZAOLA, I.; ALONSO, M. P. (2007): *Testuak, diskurtsoak eta generoak. Euskal testuen azterketa korrontea*, Donostia: Erein.
- SWALES, J. (1990): *Genre Analysis*, Cambridge: Cambridge University Press.
- ZABALA, I. (zuz.) (2008): "Diskurtso akademiko-profesionalaren didaktika eta garapena uztartzeko tresna informatikoen diseinu eta integrazioa helburua duen proiektua", in UGARTEBURU, I. eta SALABURU, P. (arg.): *Espezialitate Hizkerak eta Terminologia III*, Bilbo: UPV/EHU.

[AURKIBIDEA](#)