

# Hiztegiak eta informatika

E. Agirre, X. Arregi, X. Artola, A. Diaz de Ilarraza eta K. Sarasola  
Donostiako Informatika Fakultateko irakasleak

## Sarrera

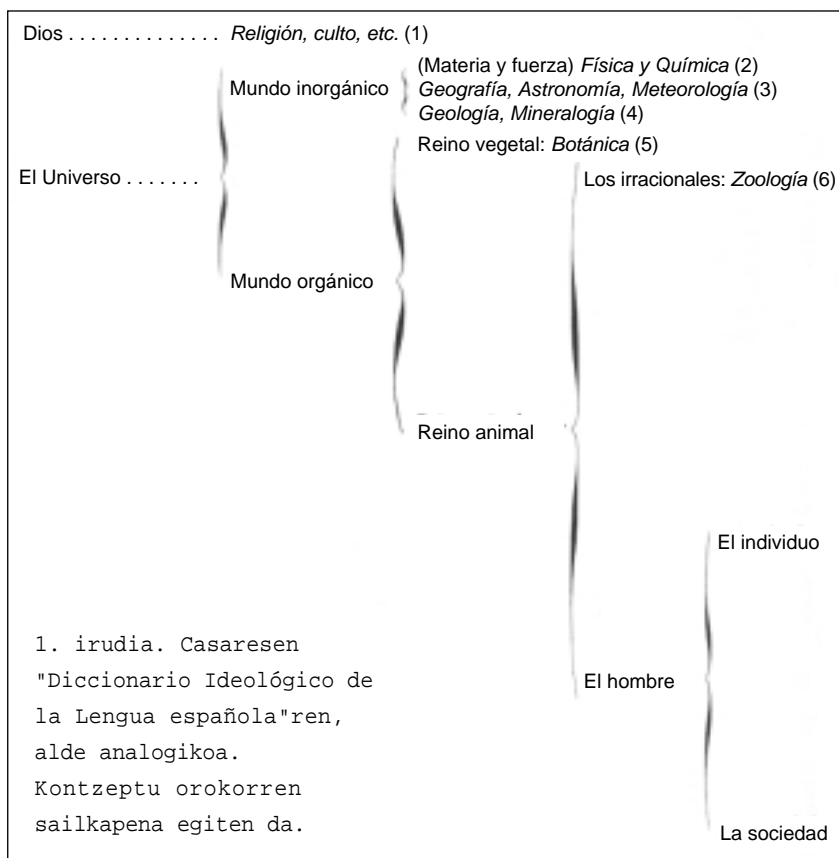
**O**rain dela gutxi arte (eta oraindik ere bai) hiztegia erabiltzeak paperezko liburu mardulen orrietan zehar aurrera eta atzera ibiltzea esan nahi zuen. Ordenadore-munduan sortu diren azken aurrerapen teknologikoei esker, hiztegiak kontsultatzeko bideak erraztu egin dira. Gainera kontsulta-mota berriak ere burutu daitezke, hainbat orrialde eta definiziotan zehar banatuta dauden informazio interesgarriak ediren ahal izateko.

Ordenadorearen aplikazio berri honek zabaltzen dituen aukeren berri eman nahi genuke artikulua honetan. Ezaugarri orokorrak emango ditugu, eta zenbait sistema eta proiektu aurkeztu ere egingo dugu. Bereiztu egingo ditugu merkaturuan erosi daitezkeen sistemak eta oraindik laborategiko prototipo mailan mugitzen direnak zeintzuk diren. Aurkeztuko ditugun sistema gehienak ingeleserako sortuak badira ere, saiatu gara euskararen egoera zertan den azaltzen eta einkizunak motibatzen.

## Hiztegi-motak

**A**plikazio informatikoak azaldu baino lehen, egoki deritzogu ohizkoa ez den zenbait hiztegi-mota aurkezteari. Besterik gabe "hiztegiak" ari bagara, hitz guztien definizioak ordena alfabetikoz dauzkan hiztegi arruntaz ari garela kontsideratzen da. Hiztegi-mota hori oso baliagarri gertatzen da ezezagun dugun berba baten esangura jakiteko. Mota horri hiztegi esplikatzaile esango diogu.

Sinonimo-hiztegiak ere badira. Esplikatzaileak baino gutxiago erabiltzen dira, baina hala ere aski ezagun eta erabiliak



dira; batez ere testu bat idazterakoan, hitz bat esaldi edo parrafo berean errepikatu nahi ez badugu.

Askoz bitxiagoak diren hiztegi analogikoei, hitz bakoitzarentzat esangura aldetik berarekin zerikusirik duten hitzak aurkezten dituzte. Adibide bezala Julio Casares-en Diccionario Ideológico de la Lengua española aipa genezake. Euskarazko hiztegi analogikorik ez dago. Ikus dezagun Casaresen hiztegiaren antolaketa eta erabilera nolakoa den:

Bi atalek osatzen dute hiztegiaren alde analogikoa. Lehenengoan kontzeptu orokorren sailkapena egiten da (ikus 1.

irudia). Sailkapen horren muturretan Zoologia, Hizkuntza, Nekazaritza eta antzeko kontzeptu zabalak agertzen dira. Hiztegiko ondoko orrietan kontzeptu zabal horien barruan kokatzen diren hitzen zerrendak datoz (esate baterako, ikus 2. irudian dagoen zoologiako zerrenda). Hiztegiko bigarren atalean, lehenengo zerrendetako hitzetarako esanguraren arabera hurbil dauden hitzak ematen dira. Bertan sinonimoez gain, kontzeptuaren osagaiak, motak, erabiltzaileak, ofizioak, tokiak, etab. agertzen dira. Adibidez (3. irudian ikus daitekeenez) erdarazko *corporación* berbarako

Animal, bicho (V. Zoología)	Ave
Cuerpo (V. Anatomía, cuadro 7)	Loro, papagayo
Cabeza » »	Halcón
Cuello » »	Pájaro
Pecho » »	Aguzanieves
Espalda » »	Alondra, cogujada
Espinazo » »	Paloma
Pierna, pata » »	Gallina
Pie, garra » »	Pavo
Uña, casco » »	Pato
Cola, rabo	Perdiz
	Pollo
Músculo (V. Anatomía, cuadro 7)	Cría
Nervio » »	Ala
Hueso » »	Pluma
Piel » »	Huevo
Ojo » »	Nido
Boca » »	
Nariz, etc. » »	Reptil, serpiente, víbora
	Tortuga, galápago
	Anfibio, batracio, rana, sapo

2. irudia. Casaresen hiztegian dagoen zoologiako zerrendaren zati bat.

(euskarazko erakunde edo) beste askoren artean honakoak azaltzen dira: organismo, entidad, colectividad eta comunidad sinonimoak; colegio, gremio eta cofradia erakunde-motak; ordenanzas, reglamento eta regla, erakundeek eduki ohi duten arautegiarekin lotutakoak; etab. Bi atal horiez osatzen den alde analogikoa bukatuta, Casaresen hiztegian hitz guztien definizioak azaltzen dira (ordena alfabetikoz emanda, ohizkoa denez).

finizioarekin eta hitzen arteko diferentziak azalduz datorrelarik. Adibidez, norbait gaixorik dagoela adierazten duen hitz bat bilatzen badugu, sailkapenean zehar B111 hitz-multzoraino helduko gara.

- B The body: its functions and welfare  
 B110 Bodily conditions relating to Health, sickness and disability  
 B111 Adjectives: Showing poor bodily condition

Thesaurus-erako hiztegiak	Ohizko hiztegia
Zein hitz erabili K kontzeptua azaltzeko?	Zein da H hitzaren esangura?
Ideiatik hitzera heltzeko bidea	Hitzetik ideiara heltzeko bidea
Testu-sorkuntzarako pentsatuak	Testu-ulerkuntzarako pentsatuak

1. taula.

Ildo berekoak dira thesaurus-erako hiztegiak, baina mota horretako hiztegietan are nabariagoak dira ohizkoekin dituzten diferentziak, 1. taulan ikus daitekeenez.

Euskaraz ez dago horrelako hizteirik. Adibide gisa, ingelesezko Longman Lexicon of Contemporary English aurkeztuko dugu. Hiztegi honetan kontzeptuen sailkapena sistematikoagoa da hiztegi analogikoetan baino. Gero eta murriztagoak diren kontzeptuetan zehar sartuz, azkenean hitz-multzo txiki batera heltzen da, hitz bakoitza bere de-

B111 multzoa 4. irudian dago. Bertako definizioak konparatuz, azaldu nahi dugun ideiarentzat egokia aukeratuko dugu. Adibidez, eritasun-egoera oso arina edo laburra bada, unwell edo upset berbak aukeratuko liriateke, edo dizzy ala funny zorabio-kontua izanez gero.

3. irudia. Casaresen hiztegi 2. atalean, lehengo zerrendetako hitzetarako esanguraren arabera hurbil dauden hitzak ematen dira.

corona del príncipe de Asturias	colegiarse
corona de infante	escribirse
» ducal	inscribirse
» de duque	entrar
» de conde	ingresar
» de vizconde	ser de
» de marqués	pertenecer a
» de barón	formar parte de
coronel	incorporación
corona mural	incorporo
» de laurel	ingreso
» » espinas	recepción
mortero	admisión
	alta
coronar	baja
laurear	reincorporación
<b>ceñir</b>	consenso
	individuo
coronación	miembro
coronamiento	recipiente
coronamento	gremial
	académico
coronario	academista
—	ensor
	séviro
CORONAR	» augustal
(V. Corona)	decano
—	compañero
	colega
	concolega
	miembro podrido
	(Supl.)
<b>CORPORACIÓN</b>	
(25, 30)	
corporación	corporativo
organismo	colegiado
entidad	numerario
<b>admón. pública</b>	de número
colectividad	orgánico
comunidad	
hermandad	corporativamente
<b>asamblea</b>	colegialmente
<b>ayuntamiento</b>	en cuerpo
diputación	en común
<b>consejo</b>	—
compañía	
<b>asociación</b>	
cuerpo	CORYORAL
» facultativo	(V. Cuerpo)
colegio	—
gremio	CORPÓREO
<b>cofradía</b>	(V. M arteria)
<b>orden religiosa</b>	—
ateneo	
academia	<b>CORRAL</b> (37)
instituto	
patronato	corral
patronado	corraliza
patronazgo	cancha
cámara	encerradero
universidad	cortil
mayoría	corte
minoría	cubil
personal	corrincho
promoción	traspuesta
<b>reunión</b>	manguera
junta	aprisco
pleno	apriscadero
directiva	redil
ejecutiva	brosquil
	telera
ordenanzas	ovil
constitución	majada
reglamento	majadal
<b>regla</b>	apero
estatuto	chivetero
escalafón	chivil
	chivital
incorporar	chiquero
reincorporar	boyera
agremiar	boyeriza
afiliar	boíl
formar	avería
admitir	gallinero
recibir	trascorral
	faisanería
incorporarse	paradina
afiliarse	

**B111 adjectives: showing poor bodily condition**

**ill** [F; (B)] not in good health; not well: *She's ill so she can't come. Oh, I'm not ill enough to need a doctor. Oh God. I feel ill!*

**unhealthy** [Wa;B] **1** not generally in good physical and mental condition; often ill; not strong: *They're unhealthy children, because they don't get good food and fresh air.* **2** not likely to give good health: *This place has a hot wet unhealthy climate.* **3** not good for the body or mind: *He has an unhealthy interest in murder and similar crimes.* **4** showing illness or poor health: *Her skin is an unhealthy greenish colour.*

**unwell** [F] ill, esp for a short time: *He was unwell yesterday, but he's fine again today.*

**unfit** [Wa2;B] in bad health: *He's pretty unfit; he needs medical treatment and a lot more exercise.*

**sick** [Wa;B] **1** esp AmE, ScotE not well; ill: *She was sick for three months. He was off sick for a week and missed a lot of work.* **2** about to vomit: *I feel sick; it must be something I ate.* [fi B119 VOMIT]

**sickly** [Wa;B] always or often ill: *What a sickly child she is; she's constantly catching colds.*

**diseased** [B] showing signs of, or damaged by, disease or infection: *The plant is diseased and will soon die.*

**upset** [Wa5 ;B] slightly ill: *The child has an upset stomach.*

**bad** *infml* **1** [F] unwell: *Oh God, I feel bad! He's been bad for a week; it's his back, you know.* **2** [B] weak or unsound: *He's been ill with a bad back for months. She's got a bad heart and can't climb stairs.*

**poorly** [F] esp BrE *infml* not (very) well: *Oh, I am feeling poorly. He's been very poorly lately; he might die.*

**ailing** [Wa5;B] *infml* unwell, esp over a period of time and not seriously: *The doctor was called out to attend to an ailing child. He's an old man now and has been ailing for a long time.*

**run down** [B] *not fml* not in good health; not well: *She's pretty run down; she needs a rest.*

**morbid** [Wa5;B] *med* concerned with disease; showing disease: *The doctor detected a morbid lung condition.*

**dizzy** [Wa;B] feeling as if everything is turning round: *Gosh, I feel dizzy! She felt dizzy and fainted.* **-iness** [U] **-ily** [adv]

**funny** [B] *infml* **1** slightly unwell: *She always feels a bit funny if she looks down from a height.* **2** slightly mad: *He went rather funny (in the head) after the death of his only son.*

4. irudia. "Longman lexicon of contemporary english" hiztegiko zati bat.

Ohizko edo sinonimoen hiztegien erabilera ordenadore bidezko hiztegi-aplikazioetan erosoago bihurtzen da, baina thesaurus-erako hiztegien kasuan ordenadoreak sortuko dituen hobekuntzak askoz ikusgarriagoak dira. Hiztegi-mota horren paperezko bertsioek lortutako zabalkundea ez da handia, baina ordenadoreak erantsiko dituen aztertze-prozedura ahaltsu eta erosoei esker, oso tresna erabilgarri bihurtuko dira.

Hiztegi elektronikoak

**H**iztegi mardulak paperean inprimatuta badaude ere, beste hauek euskarri elektronikoan egoten dira. Hiztegi elektroniko batzuk ez dira ohizko hiztegietatik asko bereizten. Horrelakoei testu hutsezko hiztegi elektroniko deritze. Baina informatika aplikatzen hasi zenez gero, beste errepresentazio- eta funtzionalitate-mota batzuk ere aztertu izan dira; adibidez, datu-base lexikaletan eta baita multimedia eta hipertestu teknologiaz hornitu diren hiztegietan ere. Azkenik, euskarri berri honek beste aukera batzuk ere eskaintzen ditu, hala nola hiztegi anitzekin lan egiteko inguruneak. Banan-banan aztertuko ditugu ekarpen hauek.

Testu hutsezko hiztegi elektronikoak

Esan bezala, ohizko hiztegien antzeko izaten dira hiztegi hauek. Diferentzia honakoa da: karaktere-ak paperean egon ordez, euskarri elektronikoren batean kode bitarrean errepresentatzea. 5. irudian ikus daitekeenez, hiztegi marduletan berba-sarrerera ezberdinak konbentzio tipografikoz

Marigorringo: gorputzeko kolore gorritan izpil beltzak dituen lau hego txikizko intsektua: "marigorringo, marigorringo, bihar eguzki ala euri egingo?" (ikus: 381.or.)

Bereizten dira, baina elektronikoetan aldiz, kode berezi batzuk erabiltzen dira.

Errerepresentazio berri honek abantaila batzuk ditu. Adibidez, bertsio elektronikoari edizio-programa baten bidez begiratu gero, erabilera erraztu egiten da. Abantaila hauek ditu:

- 1) Kontsultak eskuz baino erosoago egitea.
- 2) Hiztegiaren zati ezberdinak pantailan edukitzea, eta horrela

hainbat sarreratako informazioa aldi berean ikusi ahal izatea.

- 3) Ohizko hiztegietan sarrerera bidezko bilaketa bakarrik egin daiteke. Elektronikoetan aldiz, beste bilaketa-mota batzuk ere egin daitezke. Adibidez: hitz bat zein definiziotan (bat ala gehiagotan) agertzen den.
- 4) Informazioa ordenadorean dagoenez, programa espezializatu batzuk aprobetxa daitezke. Ikerkuntza-mailan, adibidez, hiztegi hauetatik abiatzen dira lexikoari buruzko jakintza automatikoki ateratzeko. Hurrengo puntuetan aipatzen diren hiztegi sofistikuagoetarako ere, testu hutsezko hiztegietan dagoen informazioa hartzen da abiapuntu.

Abantaila hauek, erabiltzen den editoreak eskaintzen dituen aukeraren menpe daude, noski. Editore batzuek bilaketak egiteko tresna ahaltsuagoak dituzte; "pattern matching"eko teknikak, adibidez.

Gaur egun ia hiztegi guztiak horrela errepresentatuta daude, nahiz eta gero argitaratzeko orduan paperean inprimatu. Hala ere, bertsio elektroniko hauek ez dira argitaratzen; kopiatzeko dagoen arriskuagatik batez ere.

Esandakoaren adibide ditugu gure artean Euskararako Hiztegia (Adorez) edo Hauta-Lanerako Euskal Hiztegia (Sarasola).

Datu-base lexikalak (DBLak)

Halako hiztegietan, informazioa errepresentatzeko datu-baseak erabiltzen dira. Errerepresentazio honek ezaugarri batzuk inposatzen ditu sarrerak antolatzeke: sarrera osatzen duten eremu guztiak bereiztu eta sistematikoki egituratu behar dira. Horrela datu-baseek eskaintzen dituzten tresna guztiak erabil daitezke hiztegiko informazioa kudeatu eta atzitzeko.

Aipatu dugun antolakuntzaren adibide bezala, balizko DBL bateko hitz-adiera batentzat 6. irudiko eremuak aurki genitzake.

Informazio-egituraketa honek honako bilaketak eta galdeketa egiteko aukera emango luke:

- sarrera-hitzaren araberakoak (edizio-testuekin bezala, baina askoz bizkorrago)

## Hizkuntz industriak

Mende honen barnean teknologien integrazioa aldaketa kualitatibo azeleratuak eragiten ari da: elektronikaren eta informatikaren ezkontzak mikroprozesadoreak ekarri zigun, informatikarenak eta telekomunikazioenak telematika, elektronikarenak eta optikarenak optoelektronika, etab. Azken urte hauetan hizkuntzalaritzaren, informatikaren, telekomunikazioen eta beste zenbait disziplinaren integrazio guztiz emankorra gertatzen ari da. Horren emaitza Hizkuntz Industriaren sektorea da.

Gauden garaian eta gizartean egonda, berebiziko garrantzia duen esparrua dugu hau. Inoiz ez bezalako okasioa dugu jaiotzen ari den sektore batean lehen mailako artean egoteko. Sektore honek behar duen inbertsio nagusia pertsona dira. Gai izan gaitezke euskaldunok arlo honetan produktu esportagarriak egiteko.

Iharduerak hainbat izan daitezke, irudimenak proposatu adina, baina batzuk aipatzearren, hona hemen zenbait adibide: makinak pertsonari mezuak hitzez emateko sistemak, pertsonak makinei hitzez aginduak ematekoak, taula, grafiko eta bestelako errepresentazio abstraktuetatik abiatuz lengoia naturalezko testuak automatikoki idazteko sistemak, ordenadorez lagunduriko irakaskuntza, lengoia naturalez adierazitako ezagutzetatik abiatuz sistema adituen ezagutza-baseak prestatzeko sistema lagungarriak, dokumentuak eskuratzeko sistema automatikoak, testua, grafikoak, soinua eta bideoa integratzen dituzten eta gainera informazioan zehar nabigatzea ahalbidetzen duten hiztegi eta entziklopedia elektronikoak, lengoia naturalez idatzitako testuen zuzenketa ortografikoa, gramatikala edo estilistikoa egiten edo egiten laguntzen duten sistemak, ordenadorez lagunduriko itzulpena eta itzulpen automatikoa, telematika bidez edota CD-ROMetan kontsulta daitezkeen datu-base terminologikoak, hizkuntzalarientzako erreminta informatikoak, etab.

Euskal Herriko enpresa, unibertsitate eta elkarte batzuk ari dira gai hauetan: Deustuko Unibertsitatea, Donostiako Informatika Fakultatea, Gasteizko Filologi Fakultatea, Leioako Zientzi Fakultatea, Baionako IUT, Baionako Hizkia, Uzei, Elhuyar, Erandioko Eudi, Gasteizko Natural Vox, Bilboko Telekomunikazioetako Goi-mailako Injineruen Eskola gutxienez.

Uzei, Elhuyar, Elkar argitaletxea eta Euskal Herriko Ikastolen Elkarteak sustatzen ari diren EUSENOR proiektuak iharduera hauen eragingarri izan nahi luke.

Gutenberg-en asmakuntza euskarak ezinbestekoa zuela aldarrikatzeko Etxeparek egin zuen oihua, gaur errepikatu beharra dugu, Hizkuntz Industriari begira oraingoan.

Andoni Sagarna

- beste edozein eremuren araberrakoak
- eremuak gurutzatuz. Adibidez, medikuntzan erabiltzen diren aditzen bilaketa:

kat\_sint = A.  
erabilera = med.

- hitz bat zein definizioan agertzen den. Dena den, inongo indexaziorik ez badago, testu huteskoetan bezain motela gertatzen da. Indexazioa zertan datzan "multimedia eta hipertestuak: garapen teknologikoa" atalean azalduko dugu.

Datu-baseen teknologiarik esker, erabilera berezituagoetara egokitzea posible da, hala nola:

- programen bidez zuzenean erabili daitezke.
- egiaztapenak egin daitezke:
  - ea eremuen balioak onargarriak diren. Adibidez, 'Adj.' ez da adiera-zenbaki onargarria; ezta 'med.' kategoria sintaktiko zuzena ere.
  - ea definizioetan agertzen diren testu-hitz guztiak definituta dauden.<sup>1</sup>
- etab.

Erabilera horrek aukera ematen du hiztegi-bertsio ezberdinak, baldintza jakin batzuen arabera, DBL beretik ateratzeko. Adibidez, medikuntzan erabiltzen diren terminoen hiztegitxo bat, edo maiztasunez ordenaturiko hitzen zerrenda; maizen erabiltzen diren 500 hitzen zerrenda, edo 2.500 hitzera ...

DBLak, hala ere, hiztegi-gintzarako laguntza bezala diseinatu izan ohi dira batez ere, eta jadanik DBLak zuzenean erabiltzen dira

hiztegiak prestatzeko, testu huteskoekin konparatuz abantaila ugari eskaintzen dutenez.

Gaur egun, urri dira erabiltzaile arruntak lor ditzakeen

DBLak; garatu diren ingurutik ez baitira plazara irteten. Hala ere, bada horrelakorik Euskal Herrian; esate baterako, UZEIk prestatutako EUSKALTERM datu-base terminologikoa.

Urritasun honen arrazoietakoa bat, datu-baseari hedakuntzak egin nahi zaizkionean programatzailea behar izatea da. Beste arrazoi bat datu-basea erabiltzea erabiltzaile arruntarentzat zaila izatea da. DBLari probetxua ateratzeko aurrez ikasi egin beharko litzateke. Esaten dugunaren adibide izan daiteke Donostiako Informatika Fakultatean XUXEN proiektuaren barnean garatu den datu-base lexikala. Datu-base hau

6. irudia.  
Balizko datu-base lexikal baten sarrera bateko hitz-adiera batentzako eremuak

ortografia	(marigorringo, euli, ...)
homografo-zenbakia	(I, II, ...)
adiera-zenbakia	(1, 2, ...)
kategoria sintaktikoa	(I., A., Adj., Iord., ...)
erabilera-ezaugarriak	(fam., kal., med., kim., ...)
fonologia	( /æʃ/ )
maiztasun-neurriak	( % 0.000345 )
definizio-testua	(gorputzeko kolore gorritan ...)
adibideak	(marigorringo, bihar eguzki ...)
sinonimoak	
antonimoak	
etab.	

<sup>1</sup> Hemen aipamen berezi bat egin behar da lematizazioaren arazoari buruz: definizioetako hitzak deklinatuta egon daitezke. Beraz, hiztegian dauden ala ez begiratu ahal izateko, beharrezkoa da lema eskuratzea. Adibidez, "aitonek" hitza hiztegian begiratu ahal izateko, lema "aitona" dela jakin behar dugu. Arazo hau nahikoa larria da, batez ere euskara bezalako hizkuntza aglutinatiboetan, eta gainditzeko beharrezkoa da datu-basean nolabaiteko analisi morfologikoa integratzea.

sortzeko eta hedatzeko RDB datu-baseen gestiorako sistema erabili da. Erabiltzailearekiko interfaceak 7. irudian ikus daitekeen itxura hartzen du. Molde horretatik kanpoko erabilerarik egin nahiko balitz, RDBk berak edo SQL lengoaiak (datu-baseak kudeatzeko lengoia bereziak) kontsultatzeko eskaintzen dituzten aukerez baliatu beharko litzateke.

Bilaketak eragiketa ohizkoenak izaten direnez, datu-baseak kudeatzeko sistemek interface lagungarriagoak dituzte halakoetarako. XUXEN proiekturako erabili den datu-baseak, pantailak definitzeko tresna bat dauka. Horri esker, adibide bezala, "urt"ez hasten diren aditz-formak bilatzeko, 7. irudian azaltzen den bezala bete beharko lituzke erabiltzaileak Osagaia eta Kategoria eremuak. Jada ez dago programa zuzenean idatzi beharrik. Interfaceari esker zuzenean egingo da bilaketa. Bilaketa horren emaitza sarrera anitzez eratua egongo litzateke. Hauetariko bat da 8. irudian azaltzen dena.

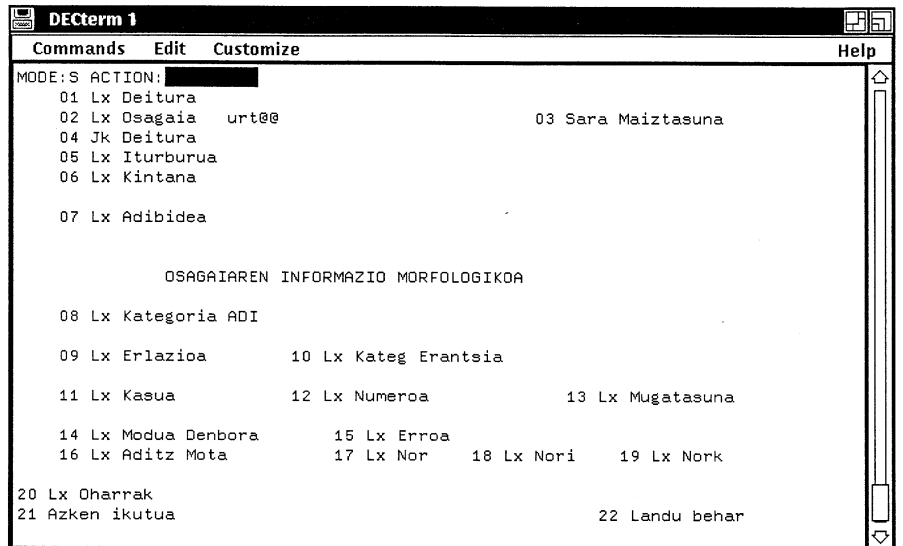
DBLa diseinatzerakoan kontutan har daiteke aipaturiko arazo hau, eta ahal den heinean erabiltzaileari bidea erraztu. Halaxe egin dute "The Oxford English Dictionary" plazaratu dutenek. Hurrengo sailean aipatzen denez, multimedia-tresnak erabiliz DBLa erabiltzaile arruntarengana gerturatzen saiatzen dira.

Multimedia eta hipertestuak: garapen teknologikoa

Informatikaren munduan azken urte hauetan agertu eta garatu diren ideia batzuk bi izenburu hauen inguruan agertzen zaizkigu. Ideia hauek biltzen dituena honakoa da: gaur egun ordenadore pertsonaletan erabil daitekeen aplikazio-mota berri bat garatzea.

Multimedia: testuaz aparte ordenadorean irudi, argazki, soinu, grafiko, animazio, etab. gorde eta maneiatzeko gaitasuna. Azken urteetan CD-ROM edo laser-diskoen hedapenari esker gero eta ezagunagoa da. NEXT delako "workstation"etan edozein momentutan atzi daitekeen ingelesezko hiztegi bat dago. Hiztegi horrek, definizioez aparte, soinu eta irudien laguntzaz azaltzen ditu hiztegi-sarrerak.

Hipertestuak: paper gaineko testuek egitura lineala dute; ezke-



7. irudia. Donostiako Informatika Fakultatean XUXEN proiektuan garatu den datu-base lexikala. Erabiltzailearekiko interfaceak.

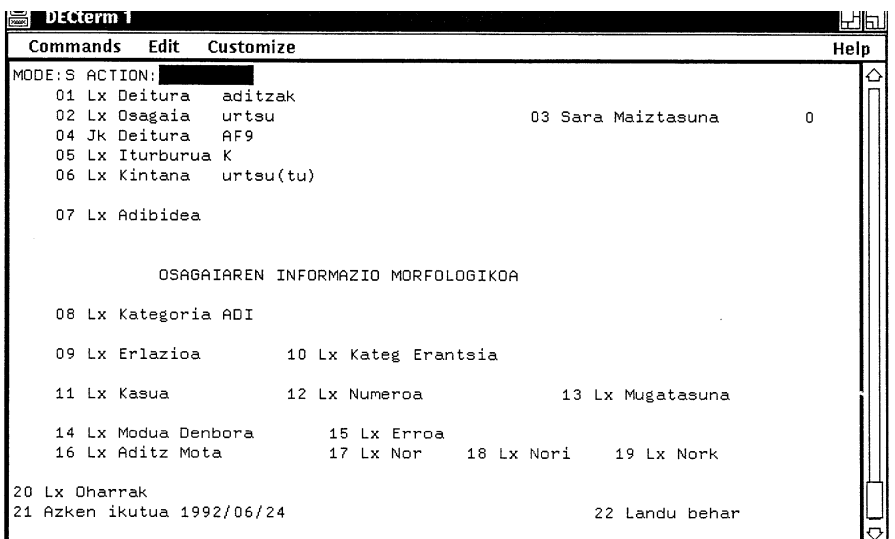
rretik eskuinera, goitik behera, liburuaren lehenengo orritik azkeneraino. Hipertestuetan testu-atalak erlazio anitzez lotuta egon daitezke, egitura linealaz aparte bestelako aukerak eskaintzeko moduan. Adibidez, hiztegiko definizio bat ikusten duenean erabiltzaileak (hipertestuaren bidez) definizio horrekin erlazioen bat duten beste definizioak araka ditzake.

Bestalde, bada testuetan barrena bilaketak errazten dituen, eta datu-base dokumentaletan oso erabilia den beste teknika bat; testu-erako informazio-ataletan (esaterako, definizio eta adibideetan) bilaketak errazteko agertzen diren hitz guztiak indexatzea. Hau da, dokumentu

osoan (hiztegi osoan) dauden hitz guztien agerpenak zerranda batean gordetzea. Erraza dirudien teknika hau, oso lagungarria suertatu da.

Aipatu ditugun teknikak honakoa eskaintzen dute oro har:

- testuetan zehar nabigatzeko erraztasuna.
- galdeketa-lengoia ahalmena galdu gabe erraztea, erabiltzailearengana hurreratuz. Ikus adibidez, 9 eta 10. irudietan gero azalduko den Grolier entziklopediako galdeketa batzuk.
- Edozein hitzek testu osoan dituen agerpen guztiak banan-banan aztertzea.
- merkaturatzeko aukera, CD-ROM euskarri masibo eta merkeari esker.



8. irudia. Aipatu XUXEN proiektuko datu-basean egindako bilaketaren (ikus 7. irudia) emaitzetako bat.

**Word Search**

Word(s)	Occurrences	Articles
martin luther	240	94
martin luther king	146	63
	0	0
	0	0

**Title List**

50 articles  
120 occurrences

16	Luther, Martin
6	Reformation
4	Bull, Olof Jacob Martin Luther
2	antinomianism
2	atonement
2	Augsburg Confession
2	Augustinians
2	Bible
2	Book of Concord
2	Bucer, Martin

9. irudia. Grolier entziklopediako galdeketa.

**Word Search**

Word(s)	Occurrences	Articles
allerg*	167	67
drug, drugs, medic*	156	31
	0	0

10. irudia. Grolier entziklopediako galdeketa.

Jadanik, Estatu Batuetan behinik behin, aipaturiko ezaugarriak dituzten entziklopedia eta hiztegiak prezio onean aurki daitezke. Euskaraz ere ari dira proiektu batzuk garatzen (Elhuyar).

Adibide gisa, aipaturiko "The Oxford English Dictionary" eta "The New Grolier World Enciclopedia" azter ditzakegu. Lehenengo CD-ROM teknologiaren garapenaren adibide argia da. Lehen hamaika tomatan zegoen hiztegi ospetsua, CD-ROM-erako disko bakar batean sartu dute. Sarrerez aparte, definizio eta adibideetan agertzen diren hitz guztiak indexatu dituzte, datu-banku itzel honetan (milioierdi bat sarrera, 2 milioi erabilera-adibide baino gehiago ...) bestela ezinezkoak lirakeen bilaketak azkartu ahal izateko. Alderdi

sofistikatuena, bilaketak egiteko erabiltzaileari ematen zaizkion erraztasunena da, datu-base lexikalek eskaintzen duten ahalmena erabilgarriago aurkeztuz. Adibidez, bilaketa-esparruak mugatzeko iragazkiak erabiltzea, eremuen bidezko bilaketa selektiboa, etab.

"Grolier" entziklopediak multimedia eta hipertestuak erabiltzen ditu. Multimediaren bidez, aztertzen ari den artikularekin zerikusia duten irudiak aurkezten ditu. Hipertestuaren bidez, posible da artikularekin zerikusia duten beste artikulak aztertzea (testuan azpimarratuta dauden hitzak saguarekin aukeratuz) eta gero bidean sistemak eskaintzen duen koaderno elektroniko batean apunteak hartu ondoren jatorrizko artikulura itzultzea.

Lexikografoak (eta erabiltzaile espezializatuak ere bai) hiztegi bat baino gehiagotara jo behar izaten du. Hiztegi-sorta hori ingurune informatiko batean integratuz gero, posible da beren arteko loturak sortzea eta, beraz, erlazioatutako informazioa hiztegi ezberdinetatik atzitzea.

Hiztegi horiek, ohizko esplikatzaileraz aparte, elebidunak, sinonimoenak eta berezituak izan daitezkeenez, erabiltzaileak honelako kontsultak egiteko aukera edukiko luke:

- Hitz baten sinonimoak ezagutu.
- Hitz baten baliokidea beste hizkuntza batean ezagutu; baita hitz baliokide horren definizioa ere; ñabardurak kontutan hartu ahal izateko.
- Hitz baten hiztegi ezberdinetako erabilera-adibideak ezagutu.
- Hitz arraroren bat (bai termino tekniko delako, bai antzinakoa delako, ...) zein hiztegitan dagoen begiratu.

Aukera hauez gain erabiltzaileari lagungarri gerta dakizkiokeen beste tresna batzuk ere gehi daitezke horrelako inguruneetan, hala nola estilo-liburua, etab.

**Etorkizuneko hiztegi informatikoak**

**O**rain arte azaldu ditugun hiztegi-erabilera informatikoetan bide ahalsu eta oso lagungarriak ikusi ditugu, baina azken batean kontsultako erantzunean azaleratzen den informazioa hiztegi berran irakur zitekeen informazioa zen. Aldiz, ikerketa-mailan aztertzen diren sistema-mota berrietan dedukzioak edo inferentziak egiteko gaitasuna hiztegi adierazpidean integratzen denez, esplizitu dagoen informazioaz gain zeharka deduzi daitezkeen informazio implizitua ere erakutsi nahi dute. Ikus dezagun adibide baten bidez. Azter itzazu 11. irudian dauden Euskararako Hiztegiak lau definizioak.

Bihar-etziko hiztegi berrak lengoia naturalaz egin ahal izango direnez, demagun honako galdera hau egiten dela: Zeintzuk dira animalia gorriak?



## ZIENTZIAREN HISTORIOGRAFIA (II) (Errenazimendutik XIX. mendera arte)



Luis Maria Bandrés

7. Antarktida  
Juan Ignacio Abrisketa (182 or.)
8. Homeopatia eta akupuntura  
X. Ibarburu (189 or.)
9. Charles Darwin  
Iñaki Azkune (200 or.)
10. Zientziaren Historiografia (I)  
Luis Maria Bandres (238 or.)
11. Zientziaren Historiografia (II)  
Luis Maria Bandres (206 or.)

Harpidedun egin zaitez gure liburuak merkeago lortuz.

### HARPIDETZA-TXARTELA

Izen-deiturak \_\_\_\_\_  
 Helbidea \_\_\_\_\_ Tel. \_\_\_\_\_  
 Herria \_\_\_\_\_ Post. Kod. \_\_\_\_\_  
 Bankua/Aurrezki Kutxa \_\_\_\_\_  
 Sukurtsala \_\_\_\_\_  
 Kontu-zenbakia \_\_\_\_\_

GAIK argitaldaria / S. Bartolome, 36-behea / Tel. 471304/  
 20007 - DONOSTIA

Marigorringo: gorputzeko kolore gorritan izpil beltzak dituen lau hego txikizko intsektua: "marigorringo, marigorringo, bihar eguzki ala euri egingo?" (ikus: 381.orr.)

Intsektu: trakeetatik arnasa hartu eta zenbait aldaketa izan ohi dituen sei hankako animalia artikulatua: erleak, eltxoak, euliak, kakalardoak, tximeletak, inurriak eta antzeakoak intsektuak dira

Animalia: berez higitzen den eta sentimena duen izaki bizia: zakurra, katua, behia, ahuntza, eta gizakiok ere bai, animaliak gara

Maitasun: atsegin edo gogoko dugunarengana bultzatzen gaituen sentimendua. Amodioa

11. irudia. Marigorringo, intsektu, animalia eta maitasun hitzen definizioak.

Sistemak itzuliko lukeen erantzunean marigorringo ere agertuko litzateke, nahiz eta bere definizioan bertan animalia denik espreski agertu ez. Horrelako sistemak marigorringo-ren definiziotik aterako luke intsektu gorria dela, intsektu-ren definiziotik animalia dela eta bietatik deduzituko luke marigorringo ere animalia gorrien artean sar daitekeela.

Marigorringoak sei hanka al ditu? galderak, era berean, baiezko erantzuna jasoko luke. Marigorringoren definiziotik intsektu klasekoa dela atera eta intsektu-ren definiziotik sei hankakoa dela ikusita.

Nolakoa da marigorringoaren arnasa? galderari trakeetatik erantzuna emango litzaioke, antzeko prozedurari jarraituz gero.

Posible al da maitasuna marigorringo batean? galdera "filosofikoari" ere erantzun ahal izango litzaioke: marigorringo intsektua da, intsektua animalia da, animaliak sentimena du, eta maitasuna sentimendua da. Sentimena eta sentimendua sinonimotzat hartuz sistema irits liteke hori ere deduzitzera. Eta ... ez da egia, ala?

Beste adibide bat aurkeztuko dugu, oraingoan sinonimiako erla-

