

ERAUZTERM

Erauzterm Elhuyar Fundazioak eta Euskal Herriko Unibertsitateko IXA taldeak garatutako euskararako terminologia-erauzlea da. Tresna honek teknika linguistikoak eta estatistikoak erabiltzen ditu, eta XML teknologiaz inplementatu da. Tresna nazioarteko bi kongresutan aurkeztu da (LREC 2004 y GLAT 2004), eta aurkezpen horiekin lotutako artikuluak eskuragarri daude ondoko helbide hauetan:

http://ixa.si.ehu.es/Ixa/Argitalpenak/Artikuluak/1079630425/publikoak/Term_Erauzketa.pdf
eta

http://ixa.si.ehu.es/Ixa/Argitalpenak/Artikuluak/1078851980/publikoak/Erauzterm_LREC

Erauzterm proiektua Hizking21 proiektu estrategikoaren barruan eraman zen aurrera. Proiektu estrategiko horren helburua izan zen beste hizkuntza batzuetarako existitzen diren funtsezko tresnak sortzea euskararako ere. Euskara hizkuntza eranskaria da eta informazio morfologiakoa ezinbestez hartu behar da kontuan testu batean termino-hautagaiak bilatzeko. Gainera, euskarak normalizatutako hizkuntzen aldean, dispersio terminologiko handiagoa du corpusetan.

Erauzterm tresnak izen-sintagmak erauzten ditu. Euskarazko izen-sintagmarik ohikoenak eta adierazgarrienak hautatzeko, IXA taldearen barruan alde aurretik egindako lan bat erabili zen (Urizar *et al.*, 2000). Hasierako lan hori patroï gehiagorekin osatu zen aurrerago. Hobekuntza lortze aldera, 50.000 hitzeko corpus bat eskuz prozesatu zen. Corpus hori Elhuyar Fundazioaren *Zientzia.net* web orrian argitaratutako 48 artikuluz osatuta dago. Terminologok corpuseko terminoak erauzi eta egitura morfosintaktikoa aztertu zuten. Erreferentziako corpus honek terminoen patroïak finkatzeko ez ezik, Erauzterm tresnaren termino-erauzketaren emaitzak balioztatzeke ere balio du.

Erauzketa automatikorako lehen urratsa da testuak XML formatuan inplementatzea. Horretarako, corpora alde aurretik prozesatzen da Eustagger lematizatzaile eta analizatzaile morfologikoaren bitartez. Horrela lortzen da terminoen lemen eta tasun morfosintaktikoen informazioa izatea.

Terminoen patroïen parekoak diren testu-kateak detektatzeko, gramatika bat sortu da. Corpora Eustagger tresnaren bitartez analizatzen da eta lortutako patroï morfosintaktikoak gramatikakoekin erkatzen dira termino-hautagaiak erauzteko. Hasteko, terminorik luzeenak erauzten dira eta, ondoren, gainerakoak lantzen dira. Prozesu honen bitartez lortutako hautagaiak eredu estatistikoaren bidez sailkatu eta probabilitateen arabera aurkezten dira.

Erabilitako teknika estatistikoaren arabera, termino-hautagai bakunak eta konposatuak lortzen dira. Termino bakunen kasuan, maiztasun erlatiboa hizkuntza orokorreko maiztasunarekin erkatzen da. Hainbat hitzez osaturiko terminoen kasuan aldiz, lotura-tekniken bitartez egiten da azterketa.

Erauzterm tresnari esker, erabiltzaileek termino-hautagaia testuinguruan ikus dezake eta terminoa balioztatzeke eta erauzteko aukera daukate. Egun zaila da estaldura/doitasun harremana erabat optimizatzea eta tresnak eskuzko balioztatzerako modulua duenez, tresnak estaldurari ematen dio lehentasuna doitasunaren aldean.

Hona hemen terminologia-erauzle elebakarraren bilaketa baten adibidea:

Leia1	Lema2					
	Lema3	Bilboko 2000 indize	M S	AprepNN	Ekon.	18 (15)
	Lema4	Espaniako ekonomia ministro		AprepNN	Ekon.	18 (7)
		etengabeko lan prestakuntza	S	AprepNN	Ekon.	18 (11)
		petrolio konpainia nagusi		NNApos	Ekon.	18 (13)
		idazkari nagusi Jose		NNN	Ekon.	18 (1)
		indioi pezetako aurrekontu	S	NAprepN	Ekon.	18 (18)
		lan istripuen kopuru	M	NAprepN	Ekon.	18 (12)
		indioi pezetaren inbertsio	S	NAprepN	Ekon.	17 (17)
		euroaren eremuko inflazio	M	AprepAprepN	Ekon.	17 (10)
		petrolioaren prezioaren igoera	M	AprepAprepN	Ekon.	17 (9)
		Gipuzkoako Foru aldundi		AprepNN	Ekon.	17 (17)
		petrolio upelaren prezio	S	NAprepN	Ekon.	17 (15)
		eusko jaurlaritzako nekazaritza		NAprepN	Ekon.	17 (12)
		Bizkaiko batzar nagusi		AprepNApos	Ekon.	16 (13)
		Elorrieta ELAko idazkari	S	NAprepN	Ekon.	16 (11)
		industria partaidetzarako estatu	S	NAprepN	Ekon.	16 (5)
		eskual herriko	S	NAprepN	Ekon.	16 (11)

...Euroaren ahulezia gainera, **euroaren eremuko inflazioak** gora egiteko arriskuak nabariak direla onartu du EBZko...

... **Euroaren eremuko inflazioak** gora egin zuen nabarmen ekainean KPI Euroaren eremuko ...

...k eta interes tasak puntu laurden igo zituen (% 0,25), **euroaren eremuko inflazioa** murrizteko. Bileraren ostean plazaratutako agiriari, ha...

...oaren gainbeherak erregaien prezioa igo du, eta horrek **euroaren eremuko inflazioa** igoaraziko du. LPEE Lurralde Petrolio-Esportatzaileen ...

...ioaren aurrikuspenak " ukitzeke jarraitzen dute ", eta **euroaren eremuko inflazioa** munduko txikienetakoa dela gogoratu nahi izan zuen. Be...

...koa dela gogoratu nahi izan zuen. Bere ustez, gainera, **euroaren eremuko inflazioa** goia jo du eta hurrengo astean hori ikusiko dela aurre...

...Zk igaro abuztuan esan zuen petrolioaren prezio altuek **euroaren eremuko inflazioa** areagotu zezaketela. OCDEren arabera, hamabi luhabetet...

... **Euroaren eremuko inflazioaren** beste igoera bat irailan...

...konomia komisarioak lehenbiziko aldiz onartu zuen atzo **Euroaren eremuko inflazioa** %2tik gora egongo dela 2000ko amaieran. Hala ere, joan...

Ritortan **euroaren eremuko inflazioa** kontrolpean dagoela, ez dindi ERZF interes inbert...

Goialdeko ezker aldean kontsulta mugatu dezakegu. Kasu honetan hiru hitzeko terminoak eskatu ditugu. Eskuinaldean terminoak azaltzen dira dagozkien patroi eta maiztasunekin batera.

Behaldeen terminoa bere testuinguruan ikus dezakegu. Badago terminoa testuinguru zabalagoan ikusteko aukera ere.