

# INTERNETEKO

## etengabeko eboluzioa



IGOR LETURIA AZKARATE  
Informatikaria eta ikertzailea

**I**nternet da gizadiak inoiz eskura izan duen ezagutza-bankurik handiena. Nahi dugun informazioa han aurkitzeko, beharrezkoak ditugu bilatzaileak, hala nola Google, Yahoo edo Bing. Hasieran hitzen bilaketa hutsa egiten zuten, baina gero eta aukera gehiago dituzte. Eta bilatzaile-mota berriak ere hasi dira agertzen.

Weba sortu zenean bilatzaileak oso sinpleak ziren. Emandako hitza edo hitzak zituzten dokumentuak bilatzen zituzten; gero irizpide orokor batzuen arabera ordenatu, eta kito. Baina arazo ugari zituzten: adibidez, askotan, emaitzak ez zeuden guk nahi genuen hizkuntzan; edo bilatutako hitzaren eratorriak ez zituzten bilatzen; ezta bilagaiaren sinonimoak ere; eta abar. Denborarekin, arazo horiei konponbidea eman diete bilatzaileek. Aukera batzuk zuzenean integratu dituzte bilaketetan, eta beste batzuk aukera aurreratuetan eskaintzen dituzte. Horrez gain, bilatzaile espezializatuak sortu dira, arazo horietako batzuen konponbidea ematen dutenak.

### HIZKUNTZAREN ARABERAKO HOBEKUNTZAK

Lehen hobekuntzetako bat hizkuntzaren detekzioa izan zen. Hizkuntza-teknologiak erabiliz, webeko orri bat zer hizkuntzatan dagoen detektatzeko gai dira bilatzaileak, eta, horrela, hizkuntza jakin bateko orriak soilik eskaintzen dituzte. Gainera, orrien hizkuntza ezaguna dutenean, hizkuntza horren arabera tratamendu berezia eskaintzen diote orriari. Adibidez, *stemming* edo *sasi-lematizazioa* aplikatzen diete hitz guztiei, eta horrela bilaketa hitzaren lema-aren arabera izatea lortzen dute, aipatutako eratorrien arazoa gaindituz.

Hala ere, bilatzaile nagusiek hizkuntza garrantzitsuenekin edo sarean presentzia handiena dutenekin soilik egiten dute hori; eta hori ez da euskararen kasua. Euskarazko emaitzak soilik eta lema bidezkoak nahi ditugunean, Elebila bilatzailea erabil dezakegu (<http://www.elebila.eu>),

Elhuyar Fundazioko I+G taldearen teknologian oinarrituta Eleka enpresak komertzializatutako bilatzailea.

### ELEANIZTASUNA

Beste kasu batzuetan kontrakoa interesatuko zaigu, hau da, gauza jakin bati buruz dauden web-orriak interesgarrienak lortzea, zernahi hizkuntzatan daudela. Hau ahalbidetzea helburu duen ikerketa-lerroari hizkuntzarteko informazio-bilaketa deritzo. Bilatu nahi den hitza edo hitzak beste hizkuntzetara itzultzen dira, bai eta horietan bilaketa egin ere; ondoren, hizkuntza bakoitzeko emaitzarik esanguratsuenak bueltatzen dira. Zirkulua ixteko, emaitza guztiak itzulpen automatiko bidez hasierako hizkuntzara itzul daitezke.

Horrelakoen adibide esperimental batzuk <http://terpconnect.umd.edu/~dlrg/clir/systems.html> helbidean aurki daitezke. Bilatzaile komertzialetan Google da horrelako zerbait egiten duen bakarra, Google Translated Search zerbitzuaren bidez ([http://translate.google.com/translate\\_s](http://translate.google.com/translate_s)). Hala, esaterako, "bars in Moscow" ("tabernak Moskun") bilatzeko eska diezaiokegu errusieraz dauden orrietan. Hark galdera errusierara itzuli, bilaketa egin, eta ondoren ingelesera itzuliko ditu emaitzak.

Euskarari dagokionean, Elhuyar Fundazioaren I+G taldeak laster kaleratu behar du Zientzian hizkuntzarteko zientzia-bilatzailea. Bilatu nahi duguna euskaraz emanik, euskarazko, gaztelaniazko eta ingelesezko zientzia-webgune esanguratsuenetan bilatuko du eskatutakoa.

### ESANAHIA OINARRI

Gerta daiteke guk bilatu nahi dugun kontzeptua adierazteko hainbat hitz egotea. Baina bilatzaileak hitz zehatz hori duten orriak soilik bueltatuko dizkigu. Emaitzak hobetzeko, galderaren hedapena deritzon teknika erabil daiteke, hitzaren sinonimoak edo aldaerak ere bilatzean

# BILATZAI LEAK

datzana. Googlek, adibidez, sinonimoak ere bilatzen ditu hitzaren aurretik ~ ikurra jarriz gero. Elebila euskarazko bilatzaileak automatikoki bilatu ez, baina aukeran emango dizkizu hitzaren aldaera edo sinonimoak.

Bestalde, bilatu nahi dugun hitzak esanahi bat baino gehiago baditu, guri horietako batekin lotutako emaitzak soilik interesatuko zaizkigu normalean. Horiek bakarrik itzultzea oso lagungarria da, edo gutxienez emaitzak esanahi ezberdinen arabera taldekatuta agertzea. Horrelako zerbait egiten saiatzen dira Microsoft-en Bing Reference bilatzailea (<http://www.bing.com/reference>) —Wikipedia-ko artikuluen gainean soilik, hala ere— edo Hakia bilatzailea (<http://www.hakia.com>).

Edonola ere, aukera hauek inplementatzeko, bilatzaileak asmatu beharra du hitzaren esanahietako zein interesatzen zaion erabiltzaileari. Hori egiteko hainbat modu daude. Bat da erabiltzaileari zuzenean galdetzea zein den interesatzen zaion esanahia edo hitza ongi itzuli den. Beste bat da hizkuntza-teknologiaren bidez esanahia asmatzen saiatzea beste hitzek ematen duten testuingurua baliatuz, baina horretarako hitz bat baino gehiagoz osatua egon behar da bilaketa. Eta beste bat da erabiltzailearen bilaketa-historialaz edo kokapen geografikoaz baliatuz esanahia asmatzen saiatzea. Azken hau egiten du Googlek, berariaz horretarako baime-na ematen badiogu.

## GALDERAK ERANTZUNEZ

Hainbat kasutan, galdera baten erantzun zehatzaren bila jotzen dugu Internetera. Bilatzaile arrunt bati galdera eginez gero, galderako hitzak dituzten dokumentuen zerrenda itzuliko digu; baina badaude galderei erantzuteko gai diren sistemak ere. Batzuek testuak eta informazio-bilaketako eta hizkuntza-teknologietako

teknikak darabiltzate, hala nola MIT-eko START sistemak (<http://start.csail.mit.edu/>) edo IXA Taldeak garatutako Ihardetsi, euskarazko galderari erantzuten diena. Beste batzuek ezagutza egituratua eta arrazonomendu automatikoa erabiltzen dute, adibidez Wolfram Alpha (<http://www.wolframalpha.com>) edo TrueKnowledge (<http://www.trueknowledge.com>). Eta web semantikoa erabiltzen dutenak ere ari dira garatzen, DBPedia (<http://dbpedia.org>) kasu.

Zalantzarik ez dago bilatzaileek asko eboluzionatu dutela euren sorreratik, eta gaur egun ere hobetzen jarraitzen dutela. Horiei esker, eta oraindik nahiko egoera esperimentalean dauden



© LEO BLANCHETTE/123RF