



- [Informazioa](#)
  - [Astekaria](#)
  - [Hari@](#)
  - [Larrun](#)
  - [Sarekada](#)
  - [Net Hurbil](#)
  - [Urtekaria](#)
  - [Bilaketa aurreratua](#)
- [Bereziak](#)
- [Multimedia](#)
- [Blogak](#)
- [Aisia](#)
- [Zerbitzuek](#)
- [Gunea](#)
- [Publizitatea](#)
- [Harpidetzak](#)
- 

Bilatu

## Astekaria

[Astekaria](#) [Larrun](#) [Egileak](#) [Gaiak](#) [Elkarrizketatuak](#) [Artxiboa](#)

---

## INTERNET

## 3.062.850 DOKUMENTU TOPATU DITUT

*Bilatzailleak ahalegin handia egin arren, nahi ditugun web-orriak aurkitzea ez da erraza izaten. Bilaketa egiteko eskatu eta lortutako web-orri zerrendak oso luzeak izaten dira, asko ez dira gure beharretara egokitzen, batzuk abandonatutakoak dira...*



**Joxerra  
Aizpurua**

1780. alea  
2000-10-22

0

Gomendatu



Interneten zerbait aurkitu nahi dugunean bilatzaileak erabiltzera jotzen dugu. Nahi adina bilatzaile ditugu eskueran eta askotan zaila izaten da bata eta bestearen arteko ezaugarriak bereiztea. Bilatzaileak datu-base edo informazio-base erraldioetan oinarritzen dira eta hauek etenik gabe automatikoki elikatzen aritzen dira; oro har, elikatzeak bidea robot bilatzailearena izaten da, hauek web-ean argitaratzen dena automatikoki detektatu, indexatu eta indizeak basera kargatzen dituztelarik. Elikatzeko erabiltzen den beste bideetako bat web-orrien kudeatzaileek, interesgarri deritzoten bilatzaileetan, eskuz alta ematearena izaten da.

Interneten eguneroko hainbat eta hainbat web-orri berri sortzen da eta beraz web-orrien kopurua etenik gabe hazten ari da. Aldi berean badago hazkunde esponentziala duen beste prozesu bat, hots, web-orrien desagerketarena edo abandonaketarena; web-orria desagertzen denean, gehienetan erreferentzia ez da bilatzaileen datu-baseetan ezabatzen eta ondorioz, erreferentzia horren gainean klik egitean desagertzen-abisua edo errore-mezua azaltzen zaigu; web-orrien abandonaketa da, ez bairik gabe, Interneten gehien hedatu den gaixotasuna eta horren froga erraz topa dezakegu edozein bilaketa egin ondoren aurkitutako orriaren eguneraketa-data ikusiz. Honen harira esan, oso ohikoa izaten dela enpresek edo partikularrek web-orriak diseinatzea, Interneten kokatzea eta ahaztea, hau da, Interneten banago eta egin beharrekota egin dut pentsatzea. Jokatzeko era honek Internetetik plan estrategikorik ez izatea erakusten du eta hau askorentzat Internet moda baino ez delako gertatzen da.

Beraz, Interneten informazio-posibilitateak zabaltzen doazen neurrian, lastrea ere handituz doa; horregatik gero eta korapilotsuagoa da buruan dugun kontzeptua zehaztasunez eta kalitatez, nolabait esatearren, lortzea.

Kasurik onenean ere, hau da, lastrea alde batera utzita, milaka edo milioika erreferentzia dituen emaitzarekin topo egitea nahikoa arrunta da. Adibidez, artikulua honen izenburuko zenbakia dugu, hau "cat" hitza Altavista bilatzailean idatzi ondoren ateratako emaitza baita.

## NOIZ TOPATUKO DUT BEHAR DUDANA?

Halere, oraindik Interneten nahi duguna aurki dezakegu, gehiago edo gutxiago kostata, baina informazio-bilatzaileak gero eta kezkatuago daude arestian apatu arazoekin. Zer egin, bada, egoera horren aurrean?

Orrien desagerketaren edo abandonaketaren aurrean robot bilatzaile gero eta ahalmentsuagoak garatzen ari dira. Hauek, orriak desagertzean edo abandonatzean, detektatu eta datu-basean erreferentzia ezabatzeak posibilitatea izan behar dute. Garbiketa-prozesuak etenik gabekoa izan behar du eta nahiz eta ez lortu informazio-basea %100 eguneratuta mantentzea, eskaintzen dena kalitate minimo batekin eskainiko da.

Dena den, arazo larriak, buruan dugun kontzeptua ahalik eta fidagarritasun gehienez bilatzea, izaten segituko du. Orain arteko bilatzaile gehienek hitzak edota hauen arteko elkarketak erabiltzen dituzte bilaketarako oinarri gisa eta gainera bilaketa, hitza idatzi dugun lengoian bakarrik burutzen da. Hau da, "etxe" hitza dituzten testuak eskuratu nahi baditugu, ez zaizkigu "maison", "casa" edo "house" dituzten testuak azalduko. Gainera buruan dugun kontzeptuak adierazteko ez da nahikoa hitzen arteko elkarketa hutsa egitea, baizik eta esaldiak, perpausak edo parrafoak dira kontzeptu horiek adierazteko bide naturala. Adibidez, "New Yorkeko burtsen gorabeherak" azaltzen dituen informazioa bilatzeko ez da nahikoa "New York" eta "burtsa" eta "gorabehera" hitzak elkartzea, baizik eta sistemak nolabait ulertu beharko luke indize ekonomiko zehatz baten bilakaerari buruzko informazioa eskatzen zaiola.

Munduan dauden bilatzaileen oinarri edo justifikazio teknologikoa hizkuntzekin erabat loturik doa eta horrela, teknologian zabalduen dagoen hizkuntza ingelesa denez, berau da bilatzaile gehienek motorra. Motor honek euskararen gisako hizkuntza deribatiboetan neurri batean baino ez du balio, kontuan hartu behar baita ingelesezko hitzek oso forma eratorri gutxi dituztela eta aldiz, euskal hitz bakoitzari nahi adina atzizki erants dakiokela.

Ildo honetan Euskal Herrian garatzen ari den informazio-teknologiak, euskara euskarri gisa duen neurrian, atzerriko bilatzaileen teknologiarekin desberdintasun nabarmenak ditu. Batez ere, bi dira euskararen ezaugarriak kontuan hartuz bilatzaileak eta hizkuntzaren teknologia oro har, garatzen ari diren ikerketa-taldeak, Ametzagaia Taldea eta IXA Taldea.

Nahiz eta bi taldeok ikerketarako teknologia desberdinak erabili, lortutako bilatzaileek hitzen lema dute oinarritzat, hau da hitzei itxura kendu eta erroa hartzen da bilaketa-tresna gisa.

#### AMETZAGAIÑA.

Ametzagaiña Taldeak hizkuntzaren ingeniartzan hasieratik orain arte egindako bideak helburu nagusi bat duela esan daiteke, eta hau, sistema informatikoak esaldiak ulertzea da. Ildo honetan, Kapsula deituriko produktu teknologikoa sortu du eta honek hitzen lematizazio automatikoa egiten duenean, automatikoki jasotzen diren hainbat datu gaur indarrean duten bilaketa-sistemarako beharrezkoak ez badira ere, laster helduko zaien analisi semantikoa garrantzi berezia izango dute. Hortaz, euren tresna eta teknologiak prozesu ebolutibo jarraian kokatu dituzte eta bilatzaileak hasieran oso esaldi sinpleak baino ulertuko ez baditu ere, prozesu geldiezin horretan esaldi gero eta konplexuagoak eta beste hizkuntza batzuetan idatzitakoak ere ulertzen joango da.

Ildo honetan, Kapsularen robot web-bilatzaileak, urte honen bukaerarako baliabide garrantzitsu bat izango du, hots, web-orrietan eskuratutako informazioaren katalogazioa automatikoki egitearena. Lehen unean baliabide hau euskarazko informazioarentzat soilik izango da, baina datorren urtean zehar gure kulturaren eragina duten hizkuntzentzat ere implementatzen hasiko da. Honela Kapsula sisteman oinarritutako bilatzaileek, berme teknologikoa eskainiko die bai sistemetako administratzaileei, baita erabiltzaileei ere; batzuek informazioa lortzeko eta antolatzeko tresnak izango dituzten bitartean, besteek kontsulta gero eta zehatzagoak egiterik izango dute.

#### IXA TALDEA.

IXA taldeak jorratu eta jorratzen dituen ikerketa-lerro nagusiak hiru arlotan bana daitezke erabilgarritasunari begira: oinarriak, tresnak eta aplikazioak. Lehenari dagokionez, oinarriko tresna linguistikoa ikeritzen dituzte, hala nola, datu-base lexikala, testu-bildumak eta hauek prozesatzeko tresnak. Tresnei dagokionez, aplikazioak egiteko erabiltzen diren tresna orokorrak honakoak dira: lematizatzailea, analizatzaile sintaktikoa, desanbiguatzaile semantikoa eta abar. Ixa taldeak edota beste enpresek erabiltzeko aplikazioak zuzentzaile ortografikoa, estilo zuzentzailea, Web bilatzaileak eta hiztegi adimendunak dira baitik bat.

Iñaki Alegria Ixa taldeko kideak adierazi digunez, "unibertsitatean gaudenez lehen bi arloak dira inportanteenak guretzako. Aplikazioak hutsuneak betetzeko edo enprekin lankidetzan garatzen ditugu. Horrela Xuxen Hizkia eta Uzeirekin lankidetzan garatu da eta Web bilatzailea Plazaguneak garatu du guk garatutako zenbait teknologia integratuz".

Ixa taldeak garatu duen bilatzaileak lematizazio bidez bilatzen du euskarazko testuetan, baina talde kideek zaila ikusten dute beste hizkuntzetarako baliatzea.

Garatzen duten aplikazioetan ezagumendu linguistikoa oso garrantzitsua da eta hori euskararako garatu da. Prestatutako programak, batzuetan, aukera ematen dute beste hizkuntzetan aplikatzeko baina hizkuntza horietarako ezagumendu linguistikorekin osatu beharko litzateke. Alegriaren ustez, "gure lana baino gurekin aritzen diren enpresen lana dela esango nuke gure teknologia beste hizkuntzetarako baliatzea".

Orri bakar batean ikusi 1 2 »

Gomendatu



INPRIMATU

LAGUN BATI BIDALI

## Iruzkina idatzi

Izartxoaz (\*) markaturiko datuak derrigorrez bete beharrekoak dira.

Izena\*  Eposta\*

Beste norbaitek segidan idazten badu, abisua jaso nahi dut

Idatzi segurtasun kodea:

7x395

Bidali

**argia.com**

Nor gara | Kontakua  
Mapa | Lege oharra | RSS Sindikazioa | W3 WAI-A

© 2008-2011 [Argia.com](http://www.argia.com) - Komunikazio Biziagoa

Industrialdea, 15 - 20160 Lasarte-Oria (Gipuzkoa)  
Tlf: 943 371 545 - 943 373 403

Laguntzaileak

