



# Hizkuntz Teknologiak

Nola lagun zaitzake teknologiak hizkuntzarekin  
borrokan zabiltzanean?

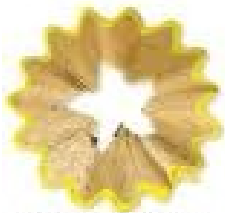
<http://bit.ly/9L7gtR>

Iñaki Alegria, Kepa Sarasola

Ixa taldea

[i.alegria@ehu.es](mailto:i.alegria@ehu.es)

<http://ixa.si.ehu.es>

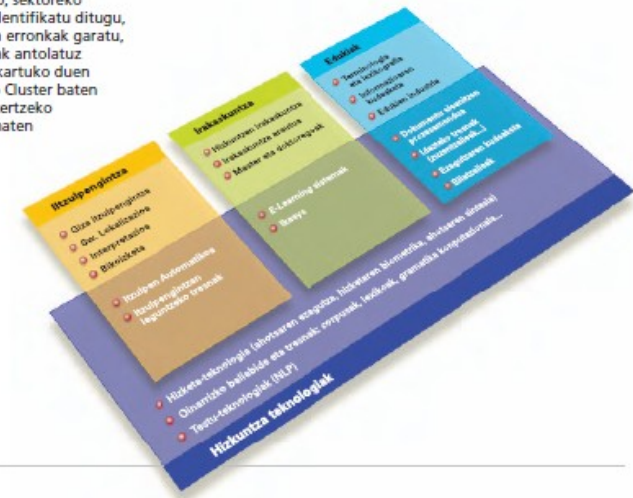


*EHU Campus de Leizor  
Calle Beatoorok  
EHU Campus Fisicopre  
Euzkoan Beatoorok*

# Aurkezpen plana

- Sarrera
- Oinarrizko tresna linguistikoak eta zenbait aplikazio
- Itzulpengintza
  - Adibide bat: *OpenTrad*
- Edukiak (XABIER GARCIA DE KORTAZAR)
  - IR - IE
  - Adibide bat: *Kyoto*
- Irakaskuntza/Ikaskuntza (ENTZI ZUBIRI, JOSUNE GEREKA)
  - Adibideak

izateko eragileak, baina  
ienara jarri dugu abian.  
etarako, sektoreko  
iteak identifikatu ditugu,  
erearen erronkak garatu,  
iragileak antolatuz  
erearen elkartuko duen  
irra edo Cluster baten  
irra aztertzeke  
tsak ematen  
gara.



# Sarrera: Ixa taldea



- <http://ixa.si.ehu.es>
- EHUko ikerketa-taldea
- 10 esplotazio-lizentzia
- Spin-off enpresa (Elhuyarrekin batera): Eleka ([www.eleka.net](http://www.eleka.net))
- Hainbat produktu merkatuan (lankidetzan)
  - *Eleka, Elhuyar, Euskaltzaindia...*
  - *Microsoft, Scansoft, Eaton...*
- Hainbat prototipo aurrekomertzial
- Hainbat europar proiektu (Meaning, Kyoto...)
- Adibideak (lankidetzan):
  - *Xuxen* zuzentzaile ortografikoa: [www.xuxen.com](http://www.xuxen.com)
  - *Opentrad*-Matxin itzultzaile automatikoa: [www.opentrad.com](http://www.opentrad.com)
  - *Bilaketa* semantikoa: [www.kyoto-project.eu](http://www.kyoto-project.eu)
- HAP masterra: <http://ixa.si.ehu.es/master>



# Sarrera: Hizkuntz teknologiak

- Oinarriak, tresnak eta aplikazioak
  - Oinarrizko baliabideak:
    - hiztegiak, corpusak, web-a, ontologiak, ezagutza-baseak...
    - corpus/testuak gordinak baino etiketatuak (XML)
  - Tresnak:
    - analizatzaile morfologikoa, etiketatzaile morfosintaktikoak, analizatzaile sintaktikoa, adieraren desanbiguazio, arrazonamendua... (gehi ohikoak: estatistikak, ML...)
  - Aplikazioak:
    - *Plug-in* gehigarriak
    - Bilaketak (IR-IE)
    - Itzulpengintza automatikoa
    - Hizketa
    - ...

# Hizkuntz teknologiak





# Tresnak

- Hizkuntzalaritzaren funtzio nagusiak ebaztea: morfologia, sintaxia, semantika, eleaniztasuna...
- Hizketaren analisia/sorkuntza
- Anbiguotasuna eta sena-eza dira arazo nagusiak
  - "zuen"                      *"zapatos de piel de señora"*
  - *"haserretu egin zen senarra beste batekin ikusi zuenean"*
- Bi hurbilpen nagusi:
  - ezagutzan oinarrituta (arauak)
  - adibideetan oinarrituta (ikasketa automatikoa-ML)
    - etiketatutako corpusetan oinarritua askotan
  - morfologian izan ezik, azken urteetan adibideetan oinarritutakoak nagusi
- Merkatua: kostu/etekina, eskalagarritasuna, integrazioa...



"You can't believe a word he says since he invented the subjunctive."

# Language Technology

## De Rerum Natura





# Tresna linguistikoak

- morfologia/lematizazioa/*stemming*
- sintaxia
- semantika (gehi pragmatika, arrazonamendua...)
- eleaniztasuna (hiztegiak, itzulpen automatikoa)
  - hiztegiak: [www.yourdictionary.com/languages.html](http://www.yourdictionary.com/languages.html)
- hizketaren ezagutza/sorkuntza (ahots-testu-ahots)
  - modu multimodala (hizketaren ezagutza, irudi/bideo sailkapena)
  - Adib.: <http://aholab.ehu.es/tts/>





# Tresna linguistikoak: Morfologia

- Tresnak:
  - analizatzaile/sortzaile morfologikoa (lematizazioa)
    - deskribapena: morfemak, morfotaktika, fonologia
    - sinplifikazioa: *stemming*
  - desanbiguazioa (*tagging*) (testuingurua kontuan hartuta)
- Oinarria: ohiko gramatikak, hiztegiak, corpusak
- Teknologia (doitasuna %90)
  - ezagutzan oinarrituta
  - desanbiguazioa: estatistika edota ML
- Aplikazioak:
  - zuzentzaile ortografikoak (plug-in / sarean)
  - estaldura handitzea bilaketan (IRn)
  - beste tresnetarako aurreprozesu gisa (sintaxia, itzulpeng., sorkuntza)
- Ikergaiak:
  - aldaera: dialektala, diakronikoa, erregistroak
  - eratorpena, elkarketa

# Morfologian oinarritutako IR aplikazioak

- Lematizazioa/*stemming*
  - funtzioak:
    - dokumentuen lematizazioa datu-basera eraman aurretik
    - galderan oinarritutako sorkuntza
  - galderaren lematizazioa
- Zuzenketa
  - hitz arraroen (erantzun gabe/gutxi) aurrean proposamenak

eu | es

Txertatu nabigatzailearen tresna-barran **Berria!** | Laguntza | Honi buruz

eLebila

zuhaitsetikan

BILATU

Bilaketa aurreratua  
Hobespenak



inguruetan  
EUSKALTZAINDIA  
inguruetan

Euskarazko web orrietan Edozein hizkuntzatan

Emaitzak: 12300 orri

- ! Erabilitako analisia: [zuhaitz izena](#).
- Sartutako testuan aldaerak daude: [zuhaitzetik](#)
- Aldaerak: [zuhaitzetik](#), [zuhaitzetio](#)

Google  
Scholar BETA

diachronic linguistics

Search

Scholar All articles - [Recent articles](#)

Did you mean: [diachronic linguistics](#)

[Arabic Phonology](#)

[F. Corriente - Jewish Quarterly Review 1972 - JSTOR](#)



zientziaren  
ELHUYAR  
komunikazioa

# zientzia.net

Eguneratze-data 2004/6/23



## Bilaketaren emaitza

284 kasu aurkitu dira **saguarekin** hitzarekin



< Aurrekoa Hurrengoa >

- Artikuluak
- Argazkiak
- Sarean
- Agenda
- Hileko zerua
- Dosierak
- Elhuyar aldizkaria
- Zientzia-hiztegiak
- Euskara teknikoa
- Dibulgazio-liburuak
- CAF-Elhuyar sariak

[asteko laburpena]  
[gaiak]

## Artikuluak euskaraz (284)



### 6.- Saguaren obesitatea murriztu duen genea klonatu dute 1999/04/01

Massachusetsko Millennium Pharmaceutical etxeko zientzialari-talde batek obesitatea kontrolatzeko erabili daitekeen gene bat klonatu du.

**Osasuna** > Osasuna



### 7.- Saguaren hiru dimentsioko atlas osatu dute 2001/04/11

Erresonantzia magnetikoaren bidezko irudiekin osatutako sagu-enbrioiaren atlas aurkeztu dute Kaliforniako ikertzaileek hil honetan.

**Biziaren zientziak** > Biologia

**Biziaren zientziak** > Genetika



### 8.- Etxe-sagua, gizakiaren aprobe txategia 1995/04/01

Munduan 1.500etik gora karraskari-espezie daude, ia ugaztunen erdia, alegia. Orden honen barruan, muridoen familia nabarmentzen da handiena delako eta familia honetakoak diren arratoiak eta etxe-saguak ikaragarri ugartu daitezkeelako.

**Biziaren zientziak** > Zoologia



### 9.- Diabetea sendatzeko bidean? Sagueta behintzat lortu da

2002/05/14



# Tresna linguistikoak: Sintaxia

- Tresnak
  - azaleko sintaxia: sintagmak bereiztea
  - sakoneko sintaxia: menpekotasunak, osagaiak, erlazioak...
- Oinarria: sintaktikoki etiketatutako corpusak (*TreeBank*), ohiko gramatikak
- Teknologia (doitasuna: % 75)
  - ezagutzan oinarritutakoa:
    - maila batetik aurrera zaila da hobetzea
    - esaldi konplexuetan arazoak
  - estatistikoa: eskuz etiketatutako corpus handiak behar dira
- Aplikazioak:
  - azaleko sintaxia:
    - zuzentzaile gramatikala, estilo-zuzentzailea
    - terminologia-erazketa, pertsona/toki/erakunde izenak
    - informazioaren iragazketa, informazioa lotzea (IE)
  - beste tresna batzuen aurreprozesua
- Ikergai nagusiak
  - analizatzaile estatistikoak, gramatiken eta estatistiken konbinazioa

# Azaleko sintaxian oinarritutako IR aplikazioak

The image displays three browser windows illustrating different IR (Information Retrieval) applications:

- Clusty Search:** Shows search results for "Clinton Obama" with a sidebar of clusters. The top cluster is "Hillary Clinton and Barack Obama" (28 items), followed by "Debate" (28), "Pennsylvania" (22), "McCain" (24), "Bill Clinton" (12), "Obama, Edwards" (12), "Blog" (9), "Bush" (6), "Clinton-Obama ticket" (5), and "Clinton Answers" (4).
- Quintura:** Shows search results for "clinton obama" with a word cloud. The most prominent words are "clinton", "barack", "obama", "rodham", "senator", "debate", "campaign", "ticket", "clash", "clinton", "obama", "rodham", "senator", "debate", "campaign", "ticket", "clash".
- KartOO:** Shows a network visualization of search results for "clinton obama". The network consists of nodes representing search results and their relationships. Key nodes include "report", "blog.washingtonpost.com", "www.nytimes.com", "www.usnews.com", "www.cnn.com", "www.factcheck.org", "www.truthdig.com", "www.politifact.com", "www.cnn.com", "www.usnews.com", "www.factcheck.org", "www.nytimes.com", "www.usnews.com", "www.cnn.com", "www.factcheck.org".

# Azaleko sintaxia:terminologia

## *Erauzterm* eta *Elexbi* (Elhuyar)

esportatu

Forma	Eredu	Maiz.	Neur.	Test.	Ter.?
IP helbide ↑ ↓	NN	41	222.26	T/KWIC	✓✗
IP protokolo ↑ ↓	NN	9	91.17	T/KWIC	✓✗
sare lokal ↑ ↓	NApos	14	82.32	T/KWIC	✓✗
sare lokal arrunt	NAposApos	1	1.86	T/KWIC	✓✗
sare lokalaren atzipen	NAprepN	1	5.31	T/KWIC	✓✗
sare lokaleko konexio	NAprepN	1	4.35	T/KWIC	✓✗
sare lokaleko zerbitzu	NAprepN	1	1.42	T/KWIC	✓✗
sare lokalen txartel	NAprepN	1	3.23	T/KWIC	✓✗
sare-txartel ↑ ↓	NN	13	67.77	T/KWIC	✓✗
TCP/ip ↓	NN	8	63.37	T/KWIC	✓✗
TCP/IP	NN	1	63.37	T/KWIC	✓✗
ethernet helbide ↓	NN	1	45.08	T/KWIC	✓✗
Ethernet helbide ↑ ↓	NN	6	45.08	T/KWIC	✓✗
Ethernet protokolo ↑ ↓	NN	6	43.29	T/KWIC	✓✗
UDP protokolo ↑ ↓	NN	5	36.77	T/KWIC	✓✗
Ethernet sare ↑ ↓	NN	4	30.58	T/KWIC	✓✗
bideratze-taula ↑ ↓	NN	5	29.22	T/KWIC	✓✗
TCP/IP protokolo ↑	NNN	5	27.94	T/KWIC	✓✗

Corpus paraleloen nabigatzailea

Sartu corpus bat:  Hautatu indize bat:  Neurrirako atalasea:  Maiztasunerako atalasea:

Hautagai kopururako atalasea:

Hitz antzeko terminoak:  Termino bakunak:  Esportatu

fiesta <a href="#">IR</a>	fiesta <a href="#">IR</a>	6	Kognatuak	<a href="#">T</a>	✓✗
informática <a href="#">IR</a>	informatika <a href="#">IR</a>	6	Kognatuak	<a href="#">T</a>	✓✗
condición <a href="#">IR</a>	kondizio <a href="#">IR</a>	6	Kognatuak	<a href="#">T</a>	✓✗
experiencia <a href="#">IR</a>	esperientzia <a href="#">IR</a>	5	Kognatuak	<a href="#">T</a>	✓✗
operador_de_telecomunicación <a href="#">IR</a>	telekomunikazio-operadore <a href="#">IR</a>	5	Kognatuak	<a href="#">T</a>	✓✗
especialidad <a href="#">IR</a>	espezialitate <a href="#">IR</a>	5	Kognatuak	<a href="#">T</a>	✓✗
gestión <a href="#">IR</a>	gestio <a href="#">IR</a>	5	Kognatuak	<a href="#">T</a>	✓✗
transferencia_de_datos <a href="#">IR</a>	datu-transferentzia <a href="#">IR</a>	5	Kognatuak	<a href="#">T</a>	✓✗
sincronización <a href="#">IR</a>	sinkronizazio <a href="#">IR</a>	5	Kognatuak	<a href="#">T</a>	✓✗
escritura <a href="#">IR</a>	eskritura <a href="#">IR</a>	5	Kognatuak	<a href="#">T</a>	✓✗
acción <a href="#">IR</a>	akzio <a href="#">IR</a>	5	Kognatuak	<a href="#">T</a>	✓✗
secreto <a href="#">IR</a>	sekretu <a href="#">IR</a>	5	Kognatuak	<a href="#">T</a>	✓✗

Que Euskaltel , S.A. se ha constituido como **operador de telecomunicaciones** en el País Vasco , siendo su vocación , además de la propiamente empresarial , la de mantener un alto grado de compromiso con el País , participando en el desarrollo tecnológico del mismo y en el acceso de la sociedad en general a servicios avanzados de telecomunicación . [T](#)

Euskaltel, SA Euskal Herrian **telekomunikazio operadore** bihurtu dela, eta enpresa-helburuak dituen arren, Euskal Herriarekin konpromiso-mailarik gorenari eusteko helburua ere baduela, Herriaren teknologi garapenean parte hartuz eta gizarte osoari telekomunikazio-zerbitzu aurreratuetarako sarbidea erraztuz. [T](#)

Adquirir una visión particularizada de los **operadores de telecomunicaciones** ; estrategias , estructura organizativa , sistema de gestión . etc. . [T](#)

**Telekomunikazio-operadore**: buruzko ikuspegi zehatza izatea; estrategiak, antolaketa-egitura, gestio-sistema, etab ... [T](#)

Maila fisikoa dispositiboen ezaugarriak definitzen dira, konektoreak eta kableak nola izan behar diren eta seinale elektrikoaren ezaugarriak zein abiadurak esaterako. Maila honetan estandar asko definitu dira, adibidez RJ-45 telefono-konexiorako konektorea edo 10base - T kable-mota sare lokaletarako . Halere, liburu honen esparrutik at geratzen da hau guztia, hardware hutsa da eta.

Find:  Match case Reached end of page, continued from top

http://213.96.132.225:8085/cgi-bin/XMLErauzT...sare\_lokal;sesioa=ARMEuUMnQU;testmota=erauz



# Tresna linguistikoak: Semantika

- Semantika (+ pragmatika, arrazonamendua...)
- Tresnak:
  - sinonimoak, hiponimoak, hiperonimoak...
  - adiera-desanbiguazioa (WSD), hautapen lexikala
- Oinarria
  - Datu-base lexiko-semantikoak (adib. *WordNet*)
  - Ontologia orokorrak (adib. *Sumo*)
  - Eremuko ontologiak (medikuntza, genetika...)
  - Eskuz etiketatutako corpusak
- Teknologia: ikasketa automatikoa nagusiki (wikipedia eta antzekoak)
- Aplikazioak
  - informazioa erlazionatzea
  - inferentzia egitea
  - estaldura handitzea bilaketan (epaile/magistratu)
  - informazioaren iragazketa (erauzketa)
- Ikergaiak: WSD, entziklopedietatik ikastea, domeinu-aldaketa...

# WordNet & Sumo

Web MCR Interface MEANING - Microsoft Internet Explorer

Fitxategia Editatu Ikusi Gogokoak Tresnak Laguntza

Atzera Bilatu Gogokoak

Helbidea <http://nipadio.lsi.upc.edu/cgi-bin/wei4/public/wei.consult.perl>

Google eneko agirre

java

Word: Nouns English\_1.6

Synonyms: near\_synonym English\_1.6

Gloss  Score  Rels  Full

Engl  Spai  Cata  Basc  Italia

Multilingual Central Repository

poisoning (Poisoning)

A **Poisoning** is caused by an external substance. Since **Poisoning** is not possible without some biologic function which affects the **Organism** being injured, it is a subclass of **BiologicalProcess**.

**Ontology**

SUMO / PROCESSES

**Superclass(es)**

```

graph BT
    poisoning --> injuring
    poisoning --> pathologic_process[pathologic process]
    poisoning --> biological_process[biological process]
    poisoning --> internal_change[internal change]
    poisoning --> process
    poisoning --> physical
    poisoning --> entity
    pathologic_process --> damaging
    biological_process --> internal_change
    internal_change --> process
    process --> physical
    physical --> entity
    
```

**Related WordNet synsets**

**poisoning**  
the physiological state produced by a poison or a toxic substance

**poison**  
administer poison to; "She poisoned her husband"

**poison, envenom**  
add poison to; "Her husband poisoned her drink in order to kill her"

See more related synsets on a separate page.

**Axioms (2)**

If **poison** is an instance of **poisoning**, then there exists **thing** so that **thing** is a patient of **poison** and **thing** is an instance of organism or **thing** is an instance of anatomical structure.

05948884n 11 coffee\_1 java\_2

alimention

food

Beverage+ 05948884n 17 café\_1

Comestible+ 05948884n 17 café\_1

Liquid+ 05948884n 12 kafe\_2

Natural+ 05948884n 12 caffè\_1

Substance+ 05948884n 12 caffè\_1

a beverage consisting of an infusion of ground coffee bean

Bebida estimulante que se obtiene de los granos del cafec

Beguda estimulant que s'obté dels grans de l'arbre del cafí

5 gloss 5 role agent 2 has mero madeof 38 rgloss 1 has hyperonym 44 role patient 11 has hyponym

2 has mero madeof 1 has hyperonym 16 has hyponym

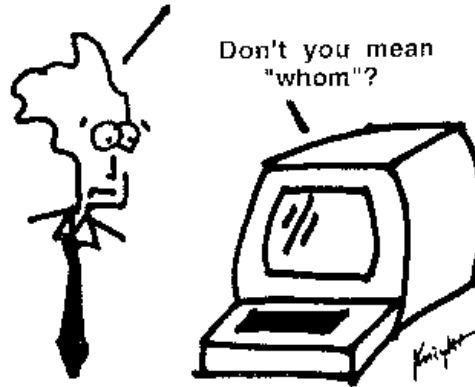
2 has mero madeof 1 has hyperonym 16 has hyponym

3 role agent 2 has mero madeof 1 has hyperonym 33 role patient 11 has hyponym





Computer! Translate into Russian:  
"We need a courier who we can  
trust with sensitive documents."





# Itzulpen automatikoa

- Sarrera
  - Zailtasunak
  - Itzulpen-memoriak eta itzulpen automatikoa
  - Erabilerak, historia, ebaluazioa
- Hurbilpenak
  - RBMT / SMT
- Adibideak
  - *OpenTrad*
- Bibliografia:

Hutchins, J. *MT-Archive*. [www.mt-archive.info](http://www.mt-archive.info)



# Itzulpen automatikoa: Sarrera

- Arazo nagusia: Anbiguotasuna
  - Lexikala, sintaktikoa, semantikoa, pragmatikoa, ...
    - Adibideak:
      - *it is raining cats and dogs / revista Time*
      - *autobusak: bihar eta berandu / azken batasunak*
    - Testuingurua, munduko ezagutza
  - Faktoreak
    - Hizkuntzen arteko antzekotasuna
    - Itzulitako testuak eta baliabide digital eleanitzak
    - Domeinua eta testu mota (domeinu orokorra? / espezializatua)
    - Testuaren zuzentasuna
  - Lagunduta?
    - Aurre-edizioa: Hizkuntza kontrolatua, Sinplifikazioa, Sorkuntza eleanitza
    - Postedizioa: zuzenketa



# Sistemak

- [www.opentrad.com](http://www.opentrad.com)
- [translate.google.com](http://translate.google.com)
- [www.systran.co.uk](http://www.systran.co.uk)
- [appserv.lsi.upc.es/multiTranslator/](http://appserv.lsi.upc.es/multiTranslator/)
- ...

# Itzulpen automatikoa

demo - OpenTrad - Mozilla Firefox

Etikategia Editatu Ikusi Historia Laster-markak Irresnak Hitzegiak Laguntza

http://www.opentrad.org/demo/libs/nabigatzalea.php?language=eu

UPV-EHU Google http://www.euskaltz... Inscripción en la Que...


Aukerak ES EU Euskalterm 3000 Elhuyar Hitzegi Batua Itzul Harluxet Mokoroa ZT Corpusa XLIXENweb Opentrad

demo - OpenTrad quijote - Buscar con Google CVC. Don Quijote de la Mancha, Capit... Google News España

Portadan **Espainia** Ir


Generada automáticamente hace 15 minutos

**Ratzinger: beldurraren estrategia**  
Egunean (Mexiko) **1 ordu Egin**  
Haren 1 bisita da Eliza katolikoaren bastioira. Benedicto XVI Brasilen egongo da Kariberen landetxe Conferencia General De El Episcopado Latinoamericano inauguratzeko eta (CELAM), gertakatu den martxoaren 13an ekintza. Latinamerika-rengan ia bizi da ...  
[Sao Paulori blindatzen da eta edertzen da hartzeko Benedicto XVI La Razón-i](#) (Espainia)  
[Munduaren herri katoliko gehiagok Benedicto XVI La Gaceta Tucumán-i zain daqokio](#)  
[Merkataritza \(Ekuatorea\) Terra España - Univisión - EITB eta 572 gai erlazionatu - »](#)



Ideal Digital

**Dick Cheney Irak-era iristen da ustekabe bisitan**  
**2 ordu egin** Terra España -  
Lehendakariorde estatubatuarra, Dick Cheney, Asteazken honetan iritsi zen ustekabe bisita batean Bagdadera erreklamazioa egiteko irakiar liderretara bikoitz dezaten adiskidetze nazionalaren alde haren esfortzuak, bomba kamioi batek utzi zuen mementoetan ...  
[Irak Clarín-i EE.UU. lehendakariordearen ustekabeko bisita](#) Com-a  
[Dick Cheney Bagdadera iristen da ustekabe bisita batean](#) ABC  
[Télam - Swiss Info - Voz De América - Abangoardia eta 58 gai erlazionatu - »](#)



Voz De América

**Recomendada berriak kontsulta egin ditzan**  
Recomendada berriak kontsulta egin ahal edun-tzeko onets beza bilaketan

http://www.opentrad.org/demo/libs/nabigatzalea.php?markatu=&norantzta=es-eu&inurl=http://www.lagaceta.com.ar/vnortae.asp?k

Inicio Microsoft Ms doc... demo - ... six04... Posta E...

**Orrialde hau izena aipatzea**

**Bulegoa**  
Diano Palentino - [eta 212 gai erlazionatu - »](#)

**Google jarraera hartzearen kontra borrokatzen da bilatzaileetan**  
Herria (Espainia) [Eta 16 gai erlazionatu - »](#)

**Garbajosa-k soilquene uzten du ahal badu.**


Google translate http://news.google.es/ Translate

Translate from: Spanish Translate into: Basque ALPHA View: Translation Original

Espainia


**Spotlight**  
Ko 9 minutes ago

**eraso deiei du Atzerritar Marokoko azalpenak ...**  
The País.com (Espainia) - **1 hour ago**  
Secretaría de Estado de Asuntos Exteriores, Juan Pablo de Laiglesia, del egin zuen Marokoko zuden eta "eskarri azalpen bat" idatzi polizia arratsaldea Larunbata arabera hiru Espainiako 14 ekintzalle parte hartu duten ...  
[Espainia Maroko from nahi ... azalpen on eraso eta](#) Diario de Sevilla  
[Ekintzalleak eraso txostena Kanarietako Marokoko Poliziak](#) International Herald Tribune  
[Cadena Ser - La Vanguardia - Canarias 7 - Canariasaldia.com eta 1095 artikuluen - »](#) [Bidali berriak](#)




Dela

**Real '9 beharrak jaten adibidez'**  
marca.com - **2 hours ago**  
Ligako lehen eguna zalantza gela ez uzten. Berria Real Madrid erailkitzen ar da. Eta proiektua martxan jartzea, arkitektoak argi pieza bat falta aurrelaria izateko atera eronka bat: ...  
[Asmorik zailtzen Mallorca](#) Faro de Vigo  
[Cristiano dagokionez](#) COPE  
[ADN.es - La Vanguardia - Yahoo! Eurosport Espainia - Kirola 890 artikulua - »](#) [Bidali berriak](#)



Intereconomía

**Alderdiak espero Montilla katalana deitu gaur ...**  
The País.com (Espainia) - **duela 21 minutu**  
Udako oporak amaitu ondoren, alderdiek espero Generalitateak, José Montilla presidentek, eta legegintzaldian emandako clinched eta gaur



e-berriak

**Telefonica esan negozio eredu ez da iraunkorra**  
EFE- 31 minutes ago - [40 artikulua - »](#)

**Youtube edukien arduraduna**  
ABC.es- 49 minutes ago - [52 artikulua guztiak - »](#)

**"Mad Men" eta "Modern Family" gata Emmy du garaipena at**  
EFE- 4 hours ago - [612 artikulua - »](#)

**The ohikoa baino Villa Barça**  
Espainia The New- 4 hours ago - [650 artikulua - »](#)

**Jimenez onartu duten hautagal izan dute eragina batean ...**  
Europa Press- 28 minutes ago - [51 artikulua guztiak - »](#)

**Beraz, TV da adi MARKA**  
marca.com- 26/08/2010

**Obama New Orleans berreskurapen osoa promises a**  
Europa Press- 5 hours ago - [800 artikulua - »](#)

**Bit ostareak zeru eta temperatura aldaketarik gabe**  
ABC.es- Duela 3 ordu - [254 artikulua guztiak - »](#)

**Berriak**  
[Real Madrid](#) [José Luis](#)


# Itzulpen-memoriak (TM) vs. Itzulpen automatikoa (MT)

- Itzulpen-memoriek eta hiztegi elektronikoen giza itzultzaileen lana errazten dute
  - Programa hedatuak
    - *OmegaT, Trados, WordFast, Deja Vu, ...*
- Oso baliagarriak dira MT tresnak eraikitzekeo garaian
  - Itzulpen-memoriak (esaldiz esaldi) TMX formatua
    - Corpus paraleloak (testu baliokideak)
    - Corpus konparagarriak (antzekoak)
  - SMT: funtsa
  - RBMT: ebaluazioa + baliabideen erauzketa
- Konbinazioa



# Itzulpen automatikoa: Erabilerak

- Zabalkundeko itzulpena (*dissemination*)
  - Kalitateko itzulpena: Aldaketak %10 edo gutxiago
  - Itzulpen-enpresek eskatzen dutena
  - Inbertsio handiak, hizkuntza-pare gutxirako
  - Sinplifikazioa
    - Domeinua, hizkuntza kontrolatua, ...
      - Katalogoak itzultzea hizkuntza askotara
    - Eguraldia (Meteo, MultiMeteo)
- Asimilazioa (*assimilation*)
  - Komunikazioa da helburua
  - Bestela itzuliko ez liratekeen testuak itzultzen dira
  - Internetek asko handitu du interesa: azkartasuna
  - Erabilpen berriak



# Erabilpen berriak: hainbat eszenatoki

- Gai zehatz baten inguruko artikulu bat bilatzeko, gero ondo itzultzeko asmoz
- Turismoko iritziak hainbat hizkuntzatan
- Enpresa baten atzerrian delegazioarekiko:
  - dokumentazio formala itzultzea: ohiko itzulpen-enpresak
  - langileen arteko komunikazioa edo talde moduan sendotzeko ekintzak: ingelesa edo asimilizaioa
- Ikastolatik SMS mezuak bidaltzeko guraso erdaldunei
- *Chat* moduko aplikazioak



# Itzulpen automatikoa: Historia

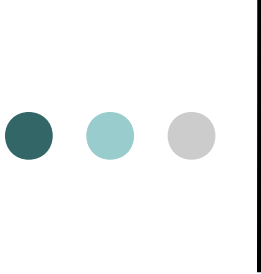


- Checking translation quality by **back-transliteration**
- *The spirit is willing, but the flesh is weak*
- English → Russian → English
- *The vodka is good but the meat is rotten*



# Itzulpen automatikoa: Ebaluazioa

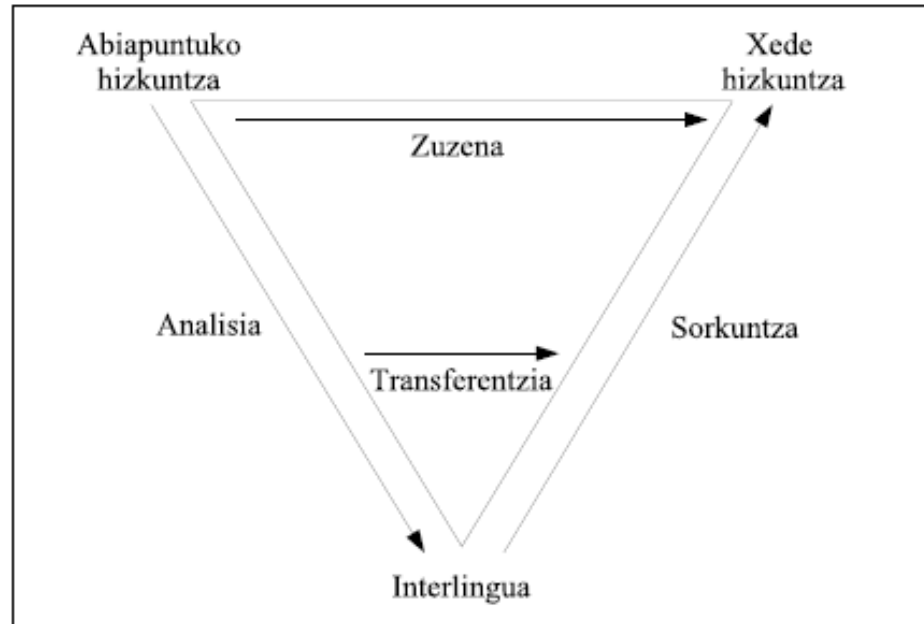
- Automatikoa
  - Erreferentziazko itzulpenetan oinarrituta, antzekotasuna/desbideratzea neurtu neurri estatistikoez
  - Neurriak: *Bleu*, *NIST*, ...
  - Behin eta berriz erabil daiteke gain-kosturik gabe
  - Eraitza eskasak, erreferentzia kopurua 3tik gora
  - Sistema estatistikoak lehenesten ditu
  - Adibidea: NIST [www.itl.nist.gov/iad/mig/tests/mt/2009/](http://www.itl.nist.gov/iad/mig/tests/mt/2009/)
- Postedizioan oinarrituta
  - Sistema bakoitzeko eta bertsio bakoitzeko kostu gehigarria
  - Egin beharreko aldaketen proportzioa
  - Zehatzagoa eta neutroagoa
- Puri-purian dagoen gaia



# Itzulpen automatikoa: Hurbilpenak

- Teknikak
  - **RBMT**: Ezagutzan/erregeletan oinarrituta
  - Egindako itzulpenetan oinarrituta
    - **EBMT**: Adibideetan oinarrituta
    - **SMT**: Estatistikan oinarrituta
- Beste ikuspuntu bat
  - Metodo linguistikoak: RBMT, EBMT
  - Metodo estatistikoak: SMT
- Konbinazioa (MEMT) eta hibridazioa

# RBMT



- Hurbilpen klasikoa, adimen artifizialetik:
  - hizkuntzalarien eta informatikarien artean
  - oso programa konplexuak
  - baliabide lexikal, sintaktiko eta semantiko konplexuak
- Kostu handikoak, urte luzetako lana
- Sistema komertzial gehienak duela gutxi arte
  - *Systran (babelfish), Opentrad (apertium, matxin)*...



# SMT

- Arrakasta handia azken urteetan
- Hizketaren ezagutzan erabilitako teknologiaren jarraipena
- Software librean oinarrituta
- Itzulpen-memorien eta corpus handien beharra
- Hizkuntzatik independentea
  - Hobeto hurrenkera finkoko hizkuntzekin eta flexio murrizkeotan
- Google-ko apustua
- 3 osagai nagusi:
  - itzulpen-eredua: itzulpen-memoretan oinarrituta
    - Programa: *Giza++*
  - xede-hizkuntzaren eredua: xede-hizkuntzaren corpusetan oinarrituta. Hainbat programa
  - Probabilitatetan oinarritutako deskodetzailea
    - Programa: *Moses*



# SMT: Adibideak

## French input

*Nous savons très bien que les Traités actuels ne suffisent pas et qu'il sera nécessaire à l'avenir de développer une structure plus efficace et différente pour l'Union, une structure plus constitutionnelle qui indique clairement quelles sont les compétences des États membres et quelles sont les compétences de l'Union.*

## Statistical machine translation

*We know very well that the current treaties are not enough and that in the future it will be necessary to develop a different and more effective structure for the union, a constitutional structure which clearly indicates what are the responsibilities of the member states and what are the competences of the union.*

## Human translation

*We know all too well that the present Treaties are inadequate and that the Union will need a better and different structure in future, a more constitutional structure which clearly distinguishes the powers of the Member States and those of the Union.*



# SMT: Adibideak (2)

## Arabic input

اعتبر تيرى رود لارسن الموفد السابق للامم المتحدة الى الشرق الاوسط ان الوضع فى هذه المنطقة لم يكن يوما على درجة الخطورة التى هو عليها اليوم مشبها المنطقة بـ " برميل بارود بفتيل مشتعل " .

## Statistical machine translation

*Former envoy Terje Roed Larsen, the United Nations to the Middle East that the situation in this region was not as serious as it is today, comparing the region "powder keg burning".*

## Human translation

*Terje Roed-Larsen, the former United Nations Middle East envoy, considered the situation in the region as having never been as dangerous as it is today and compared the region to a "powder keg with a lit fuse".*



# OpenTrad

- [www.opentrad.com](http://www.opentrad.com)
- software librean oinarrituta - kode irekia
- RBMT (MEMT eta hibridazioa lantzen)
- unibertsitateak + enpresak
  - ikerketarako laguntza
- 2 teknologia
  - Apertium: antzeko hizkuntzatarako
    - [apertium.sf.net](http://apertium.sf.net)
  - Matxin: besteetarako
    - [matxin.sf.net](http://matxin.sf.net)
- arrakasta handia, hizkuntza latinoen artean batez ere



Eitxategia Editatu Ikusi Historia Delicious Laster-markak Tresnak Hiztegiak

← → ↻ × 🏠 🖼️ 📄 TAG 🍊 http: ☆ 🌐 ins 🔍

Translate English → Euskara >>

🍊 Open... × 🌐 Google Tr... 🖨️ Traductor... 📺 diariovasc... +



# Opentrad

traducción automática

Hasiera Opentrad Zerbitzuak Berriak

## ITZULI

### Testua

El Gobierno Vasco abre mañana el curso político con un Consejo en Donostia.

Hitz ezezagunak markatu

Gaztelania-Euskara

Eusko jaurlaritzak bihar irekitzen du ikastaro politikoa Consejo batekin Donostian.

demo - OpenTrad - Mozilla Firefox

Eitxategia Editatu Ikusi Historia Laster-markak Tresnak Hiztegiak Laguntza

← → ↻ × 🏠 http://www.opentrad.org/demo/libs/nabigatzailea.php?language=eu

UPV-EHU Google http://www.euskaltz... Inscripción en la Que...

Aukerak ES → EU Euskalterm 3000 Elhuyar Hiztegi Batua Itzul Harluxet Mokoroa ZT Corpusa XUXENweb Opentrad

demo - OpenTrad quijote - Buscar con Google CVC. Don Quijote de la Mancha. Capit... Google News ES

Portadan Espainia Ir Generada automáticamente

**Portadan**

- Nazioartekoa
- Espainia
- Ekonomia
- Tecn zientzia .
- Kirolak
- Ikuskizunak
- Osasuna
- Herritar GEHIAGO

Berrien alertatuak

[Atom RSS](#) | [Feeds-ei buruz](#)

[Higikorrenzat berriak](#)

**Portadan** Espainia Ir

**Ratzinger: beldurraren estrategia**

Egunean (Mexiko) **1 ordu Egin** - Haren 1 bisita da Eliza katolikoaren bastioira. Benedicto XVI Brasilen egongo da Kariberean landetxe Conferencia General De El Episcopado Latinoamericano inauguratzeko eta (CELAM), gertatuko den martxoaren 13an ekintza. Latinamerika-engan ia bizi da ...

[Sao Paulori blindatzen da eta edertzen da hartzeko Benedicto XVI La Razón-i](#) (Espainia)

[Munduaren herri katoliko gehiagok Benedicto XVI La Gaceta Tucumán-i zain dagokio](#)

[Merkataritza \(Ekuatorea\) Terra España - Unibisión - EITE eta 572 gai erlazionatu - »](#)

**Dick Cheney Irak-era iristen da ustekabe bisitan**

**2 ordu egin** Terra España - Lehendakariorde estatubatuarra, Dick Cheney, Asteazken honetan iritsi zen ustekabe bisita batean Bagdadera erreklamazioa egiteko irakiar liderretara bikoitz dezaten adiskidetze nazionalaren alde haren esfortzuak, bomba kamioi batek utzi zuen mementoetan ...

[Irak Clarín-i EE.UU. lehendakariordearen ustekabeko bisita](#) Com-a

[Dick Cheney Bagdadera iristen da ustekabe bisita batean ABC Télam - Swiss Info - Voz De América - Abanqardía eta 58 gai erlazionatu »](#)

**Recomendada berriak kontsulta egin ditzan**

Recomendada berriak kontsulta egin ahal edun-tzeko onets beza bilaketen

**Orrialde hau izen**

**Bulegoa**  
Diario Palentino - [eta 212 gai erla...](#)

**Google jarrera hartzearen konbilatzaileetan**  
Herria (Espainia) [Eta 16 gai erla...](#)

**Garbajosa-k soilgune uzten du Europarra jolastuko da**  
Marka - [eta 17 gai erlazionatu »](#)

**José Coronado espainiar Griss Telecinco-n**  
ABC - [eta 23 gai erlazionatu »](#)

**Hobetzeko plana lehen arreta landatarra**  
Europa Sur - [eta 14 gai erlazionatu](#)

**Berria da:**

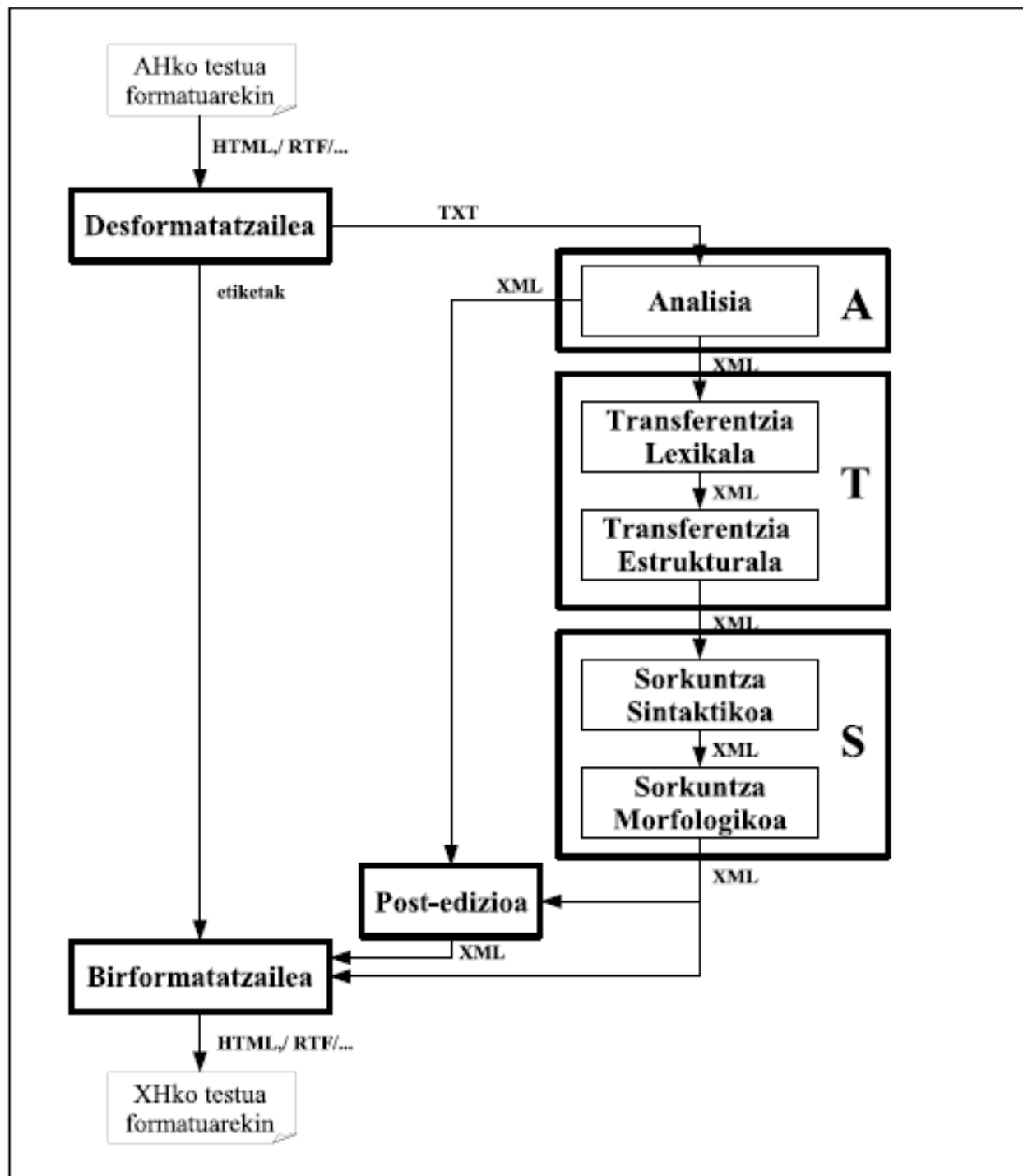
<a href="#">Ian Paisley</a>	<a href="#">Isabel II</a>
<a href="#">Martin McQuinness</a>	<a href="#">Dow Jones</a>
<a href="#">Diego Maradona</a>	<a href="#">Néstor Kirchner</a>
<a href="#">Roger Federer</a>	<a href="#">Auzitegi Go...</a>
<a href="#">San Luis Potosí</a>	<a href="#">George W. B...</a>

http://www.opentrad.org/demo/libs/nabigatzailea.php?markatu=&norantza=es-eu&inurl=http://www.lagaceta.com.ar/vernota.asp?id\_nota=217363&titulo=El%20pa%C3%

Bilatu: baby

Inicio Microso... Mis doc... demo - ... sisx04... Posta E... Bande... matxin...

# Matxin





# Bidean dauden hobekuntzak

- Integrazioa itzulpen-memoriekin
- eu-es, en-eu...
- memoriak zabaldu SMT on bat lortzeko
- RBMT/SMT konbinazioa/hibridazioa
- Domeinurako egokitzapena
  - testu administratiboak, informatika...
- Aurre-edizioa
  - erroreak eta sinplifikazioa
- Lankidetzan egindako postedizioetatik ikastea (wikipedia)



# Euskara eta MT

- Lantegi zaila:

- linguistikoki: antzekotasun-eza, hurrenkera librea, deklinatua...
- baliabideak: itzulpen-memoria gehiago behar da

- Egitasmoa nagusiak:

- OpenTrad ([www.opentrad.com](http://www.opentrad.com)) Soft librea
- Google ([translate.google.com](http://translate.google.com))
- Lucy (EJrako)
- AutomaticTrans (<http://oesi.cervantes.es/traduccionAutomatica.html>)



Netscape - [Members Directory of TAAI]

File Edit View Go Bookmarks Options Directory Window Help

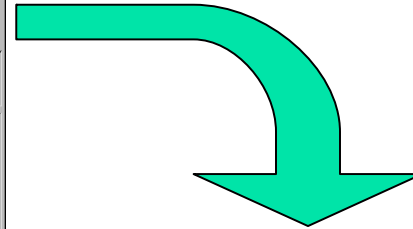
Back Forward Home Reload Images Open Print Find Stop Netscape

Location: <http://taii.et.nit.edu.tw/taii/ai-name-e.html>

What's New What's Cool Net Search News Software Support NETCOM

### Members

1. Wen-Lian Hsu (PhD, OR, Cornell U., USA; Research Fellow, Inst. Info. Sci., Academia Sinica)
2. Cheng-Seen Ho, Lifetime Member (PhD, EE, N. Taiwan U.; Prof., EE, N. Taiwan Inst. Tech.)
3. Hahn-Ming Lee (PhD, CSIE, N. Taiwan U.; Assoc. Prof., EE, N. Taiwan Inst. Tech.)
4. Chyouthwa Chen (PhD, State U. New York at Stony Brook, USA; Assoc. Prof., EE, N. Taiwan Inst. Tech.)
5. Cheng-Chih Wu (吳正己) (PhD, 德州大學電腦科學教育, USA; Assoc. Prof., Info. & Computer Edu., N. Taiwan Normal U.)
6. Ming-Chao (PhD, U. Pittsburgh, USA; Dean, 文理學院, Feng Chia U.)
7. C.J. Liao (廖純中) (PhD, CSIE, N. Taiwan U.; Assoc. Research Fellow, Inst. Info. Sci., Academia Sinica)
8. Wen-Kui Chang (PhD, Tam Kang U.; Assoc. Prof., Info. Sci. & Computer, Tunghai U.)
9. Shie-Jue Lee (PhD, CS, U. North Carolina at Chapel Hill, USA; Prof., EE, N. Sun Yat-Sen U.)
10. Rondon Tsai (PhD, Southern Methodist U., USA; Researcher, Advanced Tech. Center, Industrial Tech. Research Inst.)
11. Mark Liao (PhD, EE, Northwestern U., USA; Assoc. Research Fellow, Inst. Info. Sci., Academia Sinica)
12. B.C. Yang (楊博清) (Ms., EE, N. Taiwan Inst. Tech.; Technician, EE, N. Taiwan Inst. Tech.)
13. Hung-Chan Jang (PhD, CS, U. Illinois at Chicago, USA; Assoc. Prof., CS, N. Chengchi U.)
14. Sun-Min Tsai (Ms., EE, N. Chiao Tung U.; Assoc. Prof., EE, N. Taiwan Int. Tech.)
15. Shyi-Ming Chen (PhD, EE, N. Taiwan U.; Prof., Computer & Info. Sci., N. Chiao Tung U.)
16. W.C. Wang (王文卿) (MS, CIS, Syracuse U., USA; Engineer, Inst. Info. Industry)
17. Y.H. Hsu (徐永興) (海洋大學船機系, 現役軍人)



Name	Degree	School	Aff.
WL Hsu	PhD	Cornell	IIS
CS Ho	PhD	NTU	EE
C.Chen	PhD	SUNY	EE
C.Wu	PhD	Utexas	Cedu
Mark Liao	PhD	NWU	IIS
CJ Liao	PhD	NTU	IIS
WK Cheng	PhD	TKU	Tung
WC Wang	MS	Syracus	FIT
...			

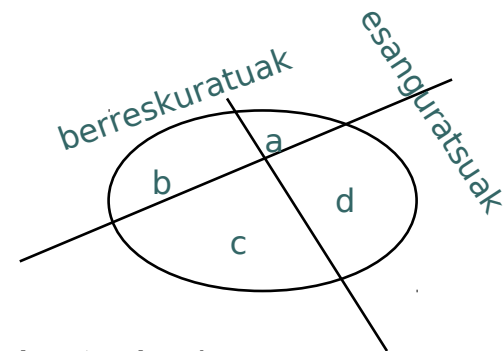


# Edukien kudeaketa HTz

- IR: *Information Retrieval*
  - biltegiratu, bilatu, aurkeztu
- IE: *Information Extraction*
  - biltegiratu, analizatu + datu-basea sortu, aurkeztu
  - HTrako probaleku egokiagoa
- Bestelakoak:
  - QA (galdera-erantzunak), CLIR, sailkapena, laburpen automatikoa...
- Adibide bat: Kyoto proiektua

# IR

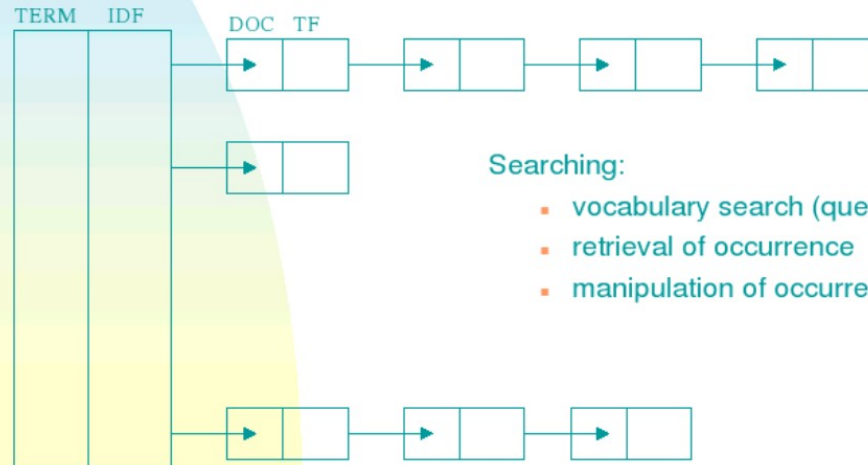
- Bilaketen teknologia
  - Internet: *Google, Yahoo, Elebila, ...*
  - Intranet: Dokumentazio-guneak, liburutegi digitalak, ...
- Teknologia: Robota+indexatzailea+bilatzailea (*page rank*)
  - *stemming+stop-list*
  - Tresna orokorrak (plug-in/gehigarriak):
    - Jabedunak: *Autonomy, Google Search Appliance...*
    - Libreak: *Lucene, Fedora Commons...*
- Ebaluazioa:
  - *Precision* (doitasuna): lortutakoaren kalitatea
  - *Recall* (estaldura): lortu ez dena
- Aldaerak
  - Sailkapena, bideratzea (*routing*), multzokatzea (*clustering*)...
    - Adib. *doku-sare* (Elhuyar)
  - CLIR, IR multimodala, QA



# IR

## Indexing: Inverted file

Word-oriented mechanism for indexing document collections to speed up searching

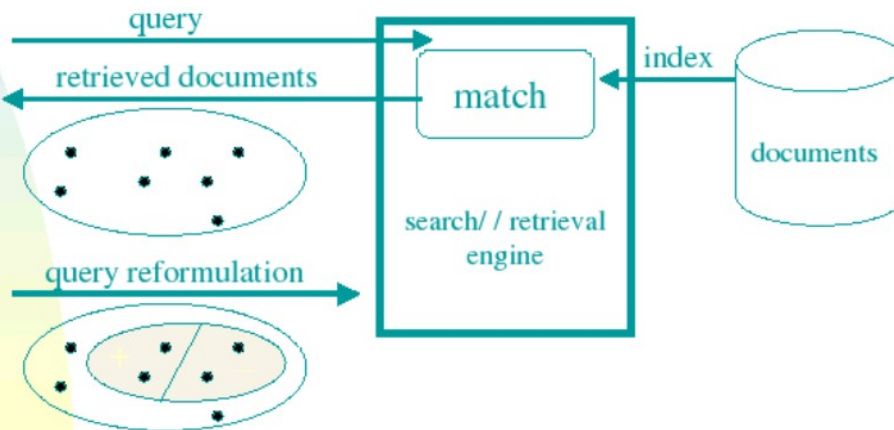


Searching:

- vocabulary search (query terms)
- retrieval of occurrence
- manipulation of occurrence

## Introduction: The (Ba

information need





leioa... http-arakatzaillea... hemeroteq...

Fitxategia Editatu Ikusi Historia Laster-markak Tresnak Hiztegiak Laguntza

http://www.google.com/search?num=100&hl=eu&q=hemeroteques&btnG...

bidaiak moodle eurOp kongresuak EHU TR tekniker udako ikast izaskun

Aukerak ES EU Euskalterm 3000 Elhuyar Hiztegi Batua Itzul

Webgunea [ludiak](#) [Maps](#) [Taldeak](#) [Gida](#)

# Google

hemeroteques Bilatu [Bilaketa Aurreratu](#) [Hobespenak](#)

Webgunea

**PROMOTORA de SOMNIS | Les pancartes i les hemeroteques**  
 Les pancartes i les **hemeroteques**. promotora | 11 Març, 2007 22:51. "La política no es cosa de agitación, sino de ideas. No es cosa de pancartas, ...  
[promotoradesomnis.baleaerweb.net/post/31007](#) - 15k - [Katxean](#) - [Antzeko orriak](#)

**Hemeroteques. Estat actual i perspectives - Dialnet**  
 Información del artículo **Hemeroteques**. Estat actual i perspectives.  
[dialnet.unioja.es/servlet/articulo?codigo=1291113](#) - 11k - [Katxean](#) - [Antzeko orriak](#)

**IV Col·loqui Aula d'Historia del Periodisme Diari de Barcelona...**  
**Hemeroteques** digitals: experiències. Autors: Lluís Codina Bonilla; Localización: Treballs de comunicació, ISSN 1131-5687, Nº. 16, 2001, pags. ...  
[dialnet.unioja.es/servlet/articulo?codigo=271331](#) - 10k - [Katxean](#) - [Antzeko orriak](#)  
[dialnet.unioja.es/domeinuan/emaizta/gehiago](#)

Elebila - Mozilla Firefox

Fitxategia Editatu Ikusi Historia Laster-markak Tresnak Hiztegiak Laguntza

http://www.elebila.eu/search?bilatu=hemerrotekak&bot\_bilatu=Bilatu&...

bidaiak moodle eurOp kongresuak EHU TR tekniker udako ikast izaskun MAMT apache2 clef

Aukerak ES EU Euskalterm 3000 Elhuyar Hiztegi Batua Itzul Harluxet Mokoroa ZT Corpora

eu es Txertatu nabigatzailearen tresna-barran **Berria!** | Laguntza | Honi buruz

# elebila

hemerrotekak **BILATU** Bilaketa aurreratu Hobespenak

Euskarazko web orrietan Edozein hizkuntzatan

Emaitzak: 36700 orri

Erabilitako analisia: *hemerroteka* izena.

**Untitled Document**  
**HEMEROTEKA** ... bat adostea ez da erraza, parte-hartze maila eta ... daudelako eta garapenerako estrategia desberdinak ere badirelako. Testuaren helburua, beraz, ez da ...  
[http://www.uztaro.com/berria/laburpena.cfm?labur=fitxategiak/Uztaro63/Uztaro63\\_6.Villalba.txt](http://www.uztaro.com/berria/laburpena.cfm?labur=fitxategiak/Uztaro63/Uztaro63_6.Villalba.txt) Kabxean

**Untitled Document**  
 ...  
[http://www.uztaro.com/berria/laburpena.cfm?labur=fitxategiak/Uztaro63/Uztaro63\\_6.Villalba.txt](http://www.uztaro.com/berria/laburpena.cfm?labur=fitxategiak/Uztaro63/Uztaro63_6.Villalba.txt) Kabxean

**Opentrad**  
 Itzulpen automatikoa kode irekian

**XUXENweb**  
 Zure enpresa ager daiteke hemen

Fitxategia Editatu Ikusi Historia Deliculous Laster-markak Tresnak Hiztegiak Laguntza

http://www.montcadaconfidencial.blogspot.com/2007/05/hemeroteques.html - 182k - [Katxean](#) - [Antzeko orriak](#)

**MONTCADA CONFIDENCIAL: HEMEROTEQUES...**  
**HEMEROTEQUES** ... (PER VEURE LA IMATGE MÉS GRAN, "CLICA" SOBRE LA MATEIXA). RECOMANEM LLEGIR ATENTAMENT AQUEST ARTICLE PUBLICAT A LA VEU (L'ANY 1999) I QUE ...  
[montcadaconfidencial.blogspot.com/2007/05/hemeroteques.html](http://www.montcadaconfidencial.blogspot.com/2007/05/hemeroteques.html) - 182k - [Katxean](#) - [Antzeko orriak](#)

**MONTCADA CONFIDENCIAL: SEGUEIXEN... LES HEMEROTEQUES...**  
 LES **HEMEROTEQUES** ... (per veure més gran la imatge, clica sobre la mateixa) CARTA ADREÇADA PER JOAN MARESMÀ ALS ELECTORS, ANY 1995. ...

Fitxategia Editatu Ikusi Historia Deliculous Laster-markak Tresnak Hiztegiak Laguntza

http://translate.google.com/translate\_s?h...

Translate English -> Euskara Sidebar

Moscow opera concerts - Google...

# Google translate

Fitxategia Editatu Ikusi Historia Deliculous Laster-markak Tresnak Hiztegiak Laguntza

http://www.quintura.com/?request=ixa+h...

Translate English -> Euskara Sidebar

Quintura - visual search engine

see & find

ixa taldea

Web Images Video

Site search Search for kids

aholab eta ehu ixaren ikerkuntza telematikoak

university euskal unibertsitateko slides elhuyar fundazioa grupo fakultateko ehuko

ixa taldea

ixa informatikari euskaldunen

Documents found - 16900

- IXA taldea**  
 Ehuko Informatika fakultateko euskaldun koadrila, hizkuntzalari batzuk lagun dituztela, euskararen prozesamenduan jo ta ke dihardutenak.  
<http://ixa.s1.ehu.es/>
- IXA taldea**  
 IXA Group. Natural Language Processing. You are here: [Ixa](#) > Products ... A Question-Answering system for the area of Science and Technology. **IXA taldea** 2009 ...  
<http://ixa.s1.ehu.es/Ixa/Produktuak>

**Search websites in other languages**

Moscow opera concerts **Translate and Search**

Search pages written in: **Russian** Translated to: **концерты Москвы опера** - [Not quite right? Edit](#)

My language: **Spanish**

**Translated results from Russian web** Results 1 - 10 of about 10,700,000 for **концерты Москвы опера** pages

**Spanish translation**

[Las entradas para el teatro, teatros conciertos en Moscú: cartel y repertorio...](#)  
 Compra entradas para el concierto en el Kremlin y otros teatros de Moscú no puede ahora...el cual se encuentran el ballet, la ópera y la música pop - todos los musicales...  
[www.portalbilet.ru/](http://www.portalbilet.ru/) - 68k - [En caché](#)

**Original Russian - [Hide Russian results](#)**

[Билеты в театры, на концерты Москвы: афиша и репертуар театров ...](#)  
 Заказать билеты на концерт в Кремле и другие театры **Москвы** теперь можно не ... участниками которого являются балет, опера и эстрада – все это мюзиклы. ...  
[www.portalbilet.ru/](http://www.portalbilet.ru/) - 68k - [Сохраненная копия](#)

[Las entradas para el teatro de Moscú, entradas de teatro, cartelera...](#)  
 ópera potencial, combinado clásico y moderno... conciertos en Moscú. alentar a todos los móviles de oficina, la oficina de la agencia en movimiento teatral Arbat

[Заказ билетов в театры Москвы, билеты в театр, театральная афиша ...](#)  
 потенциал, соединяются классика и современная опера-та ... концерты в москве . рекомендуем всем офисный переезд, переезд офиса театрального агентства Арбат



# IR

- Bizitza *google* baino lehen
  - emaitza eskasagoak / abiadura motela
  - galdera zehatzagoak
  - ontologiaren erabilpena
- Egoera korapilatsuak (HTrako aukerak)
  - eleaniztasuna
  - erantzunik ez/gutxi (estaldura handitu behar)
  - erantzun gehiegi (galdera findu behar)
    - *relevance-feedback*
- HTen ekarpena
  - sinonimia/hiperonimia: estaldura handitzeko
  - desanbiguazio semantikoa: doitasuna handitzeko



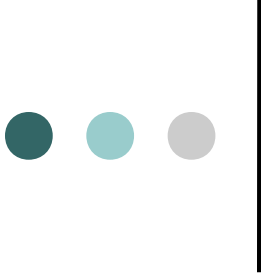
# CLIR

- IR hainbat hizkuntzetan
- Aukerak
  - galderak itzultzea
  - dokumentuak itzultzea
  - semantika bidez proiektatzea
- Gure aukera:
  - MCR (WordNet-en multzoa)
  - Itzulpen automatikoa
- Adibideak:
  - *EFE (Meaning), ArgazkiPress, Dokusare (Elhuyar)*



# IR multimodala

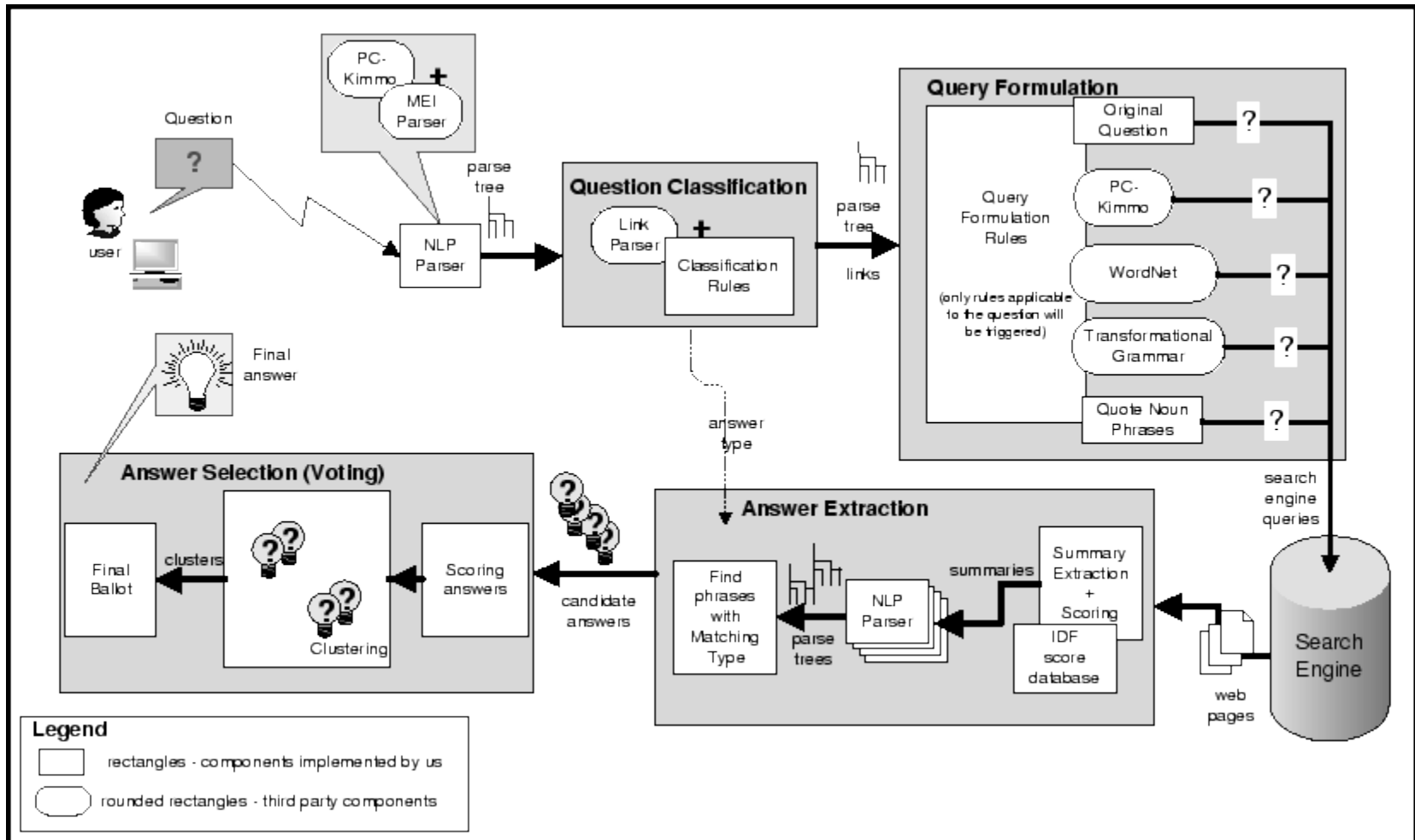
- Soinuan, irudietan eta bideotan bilaketak egitea
- Oinarrizko tresnak
  - hizketa-ezagutzaileak
    - doitasuna: %60tik gora
  - irudien sailkapen automatikoa
- Oinarrizko estrategia:
  - sailkapenak --> metadatuak
  - hizketa --> testua
  - elementuak: estekak
  - ohiko bilatzailea



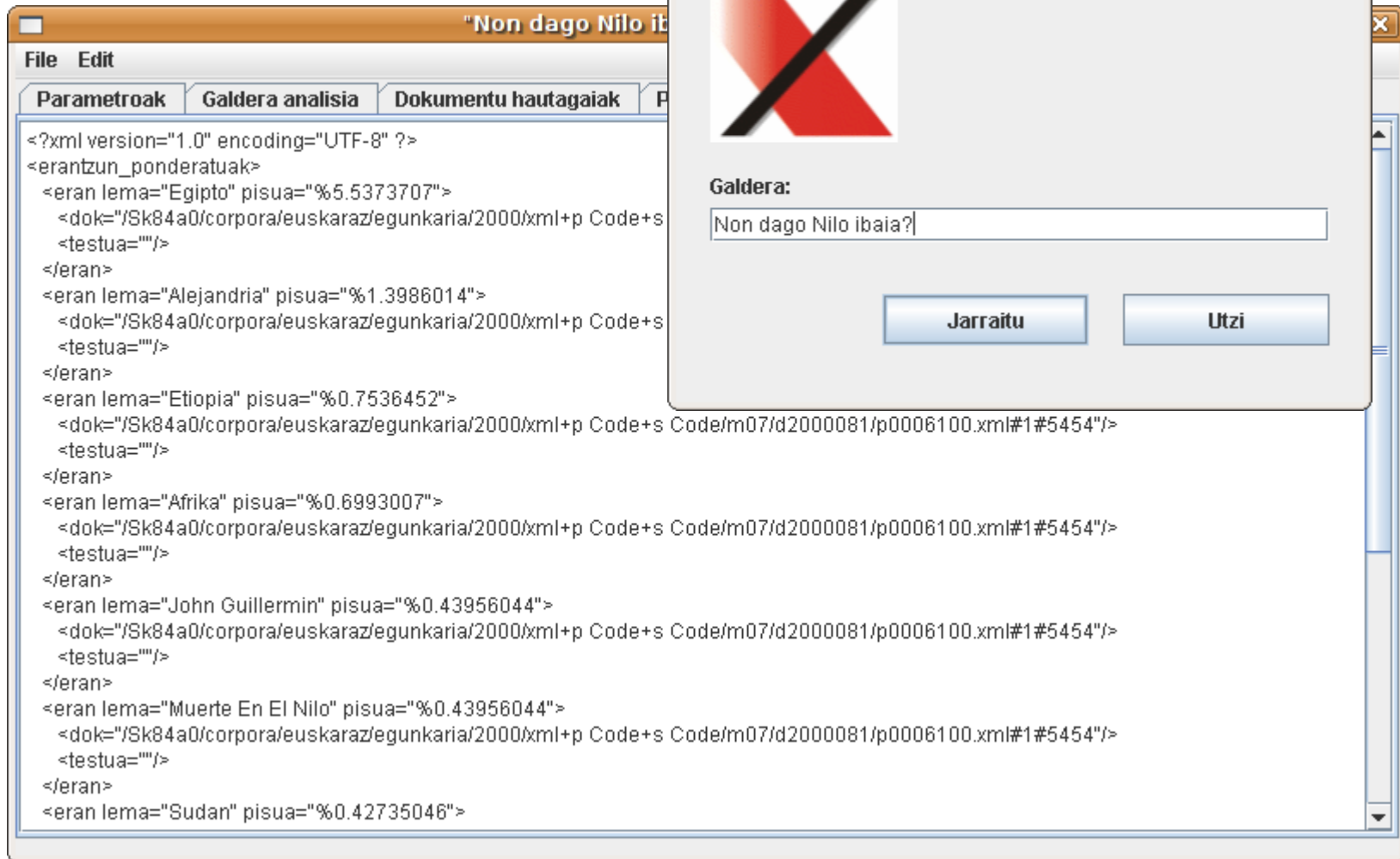
# QA (galdera-erantzunak)

- Eraitza ez da dokumentua, erantzuna baizik
- Galdera motak
  - Faktoideak (nork, zer, non, noiz)
  - Definizioak
- IR + patroien erauzketa (wikipedia gero eta gehiago)
  - galderen analisia
  - pasarteen berreskurapena ohiko IRz
  - erantzunaren bilaketa (IE)
- Arrakasta mugatua
  - Adibideak:
    - *Ihardetsi* prototipoa
    - Anhitz: **Teknopolis**

# QA: arkitektura



# Ihardetsi



Non dago Nilo ibaia

File Edit

Parametroak Galdera analisisa Dokumentu hautagaiak

```
<?xml version="1.0" encoding="UTF-8" ?>
<erantzun_ponderatuak>
  <eran lema="Egipto" pisua="%5.5373707">
    <dok="/Sk84a0/corpora/euskaraz/egunkaria/2000/xml+p Code+s
    <testua=""/>
  </eran>
  <eran lema="Alejandria" pisua="%1.3986014">
    <dok="/Sk84a0/corpora/euskaraz/egunkaria/2000/xml+p Code+s
    <testua=""/>
  </eran>
  <eran lema="Etiopia" pisua="%0.7536452">
    <dok="/Sk84a0/corpora/euskaraz/egunkaria/2000/xml+p Code+s Code/m07/d2000081/p0006100.xml#1#5454"/>
    <testua=""/>
  </eran>
  <eran lema="Afrika" pisua="%0.6993007">
    <dok="/Sk84a0/corpora/euskaraz/egunkaria/2000/xml+p Code+s Code/m07/d2000081/p0006100.xml#1#5454"/>
    <testua=""/>
  </eran>
  <eran lema="John Guillermin" pisua="%0.43956044">
    <dok="/Sk84a0/corpora/euskaraz/egunkaria/2000/xml+p Code+s Code/m07/d2000081/p0006100.xml#1#5454"/>
    <testua=""/>
  </eran>
  <eran lema="Muerte En El Nilo" pisua="%0.43956044">
    <dok="/Sk84a0/corpora/euskaraz/egunkaria/2000/xml+p Code+s Code/m07/d2000081/p0006100.xml#1#5454"/>
    <testua=""/>
  </eran>
  <eran lema="Sudan" pisua="%0.42735046">
```

Ihardetsi

Galdera:

Non dago Nilo ibaia?

Jarraitu Utzi



# IE

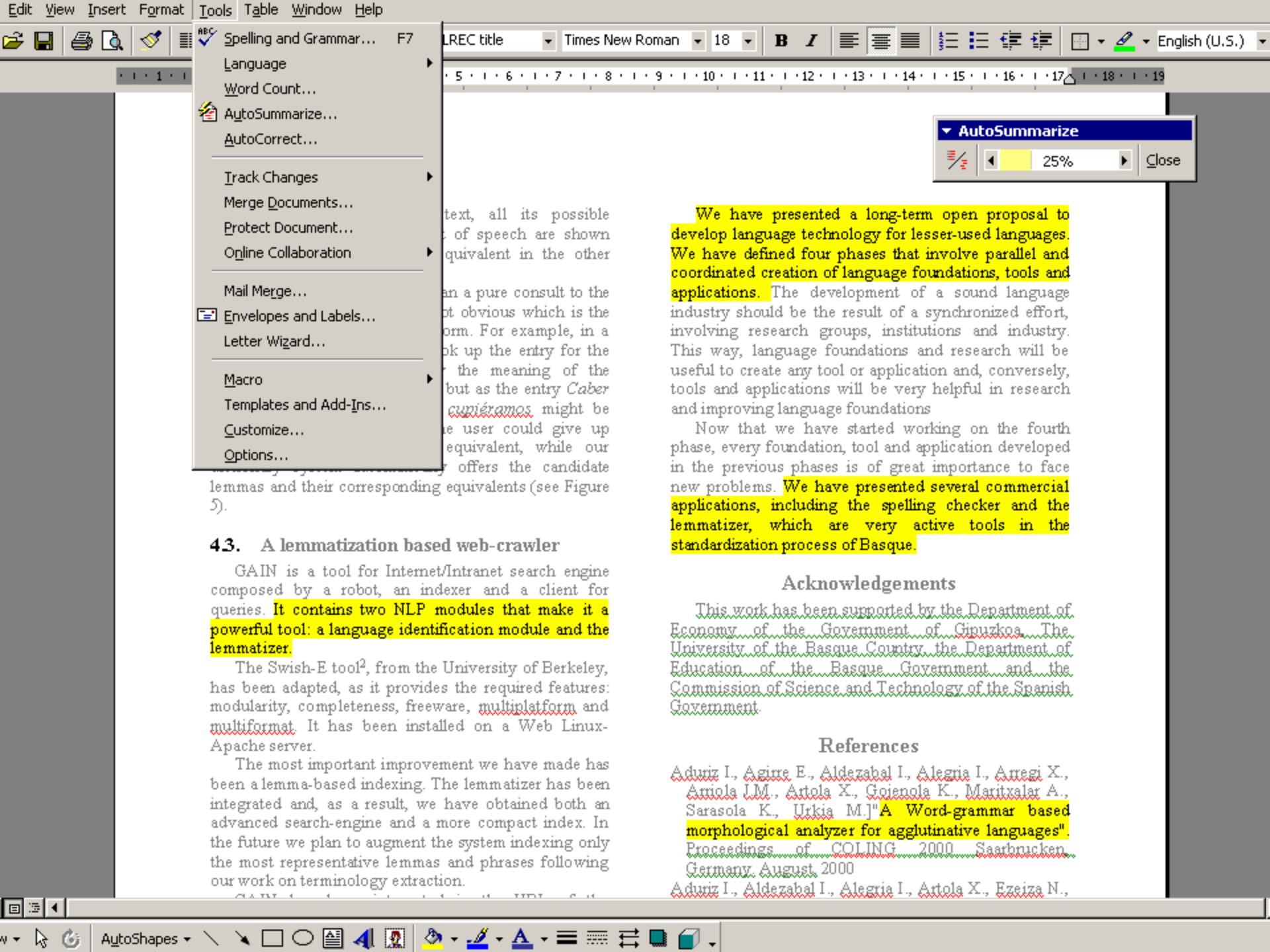
- Bilaketa sintaktiko-semantikoa
  - EZ web semantikoa
  - Tresna sintaktiko-semantikotan oinarrituta
  - Intranet mailan
- Helburua: entitate, erlazio edota gertaerei buruzko informazioa erauztea domeinu mugatu baten dokumentuen artean.
  - Funtza: informazio esanguratsuaren detekzioa
    - DBen erauzketa automatikoa, informazioaren integrazioa...
  - Domeinuak: enpresa-erabakiak, administrazioa, medikuntza, salerosketak, militarra
  - Adibideak: iritzi-meatzea (*opinion mining*), zaintza teknologikoa
- Ebaluazioa:
  - *Precision* (doitasuna): lortutakoaren kalitatea
  - *Recall* (estaldura): lortu ez dena





# IE: teknologia

- HTetan oinarrituta
- Urratsak
  - Iragazketa
  - Morfologia eta *tagging*
  - Entitateen/terminoen ezagutza eta sailkapena
  - Azaleko sintaxia
    - azkarra, azalekoa edo partziala gehienetan
  - Patroien gauzatzea: lan handia
    - Adib: *Pertsona sal-eros-aditza kopurua izen-unitateak*
    - Domeinua aldatzeko zaila: *Adaptive IE*
  - Semantika, Inferentzia, pragmatikaren beharra
    - korreferentzia, anafora (bera, hura, horrek)...
    - ontologiaren erabilpena
  - Datu-basea osatzea
- Wikipediaren laguntza gero eta gehiago
- Laburpena: tresna laguntzailea



- Spelling and Grammar... F7
- Language
- Word Count...
- AutoSummarize...
- AutoCorrect...
- Track Changes
- Merge Documents...
- Protect Document...
- Online Collaboration
- Mail Merge...
- Envelopes and Labels...
- Letter Wizard...
- Macro
- Templates and Add-Ins...
- Customize...
- Options...

**AutoSummarize**

25%

Close

text, all its possible  
of speech are shown  
equivalent in the other  
an a pure consult to the  
not obvious which is the  
orm. For example, in a  
ok up the entry for the  
the meaning of the  
but as the entry *Caber*  
*cupiéramos* might be  
the user could give up  
equivalent, while our  
offers the candidate  
lemmas and their corresponding equivalents (see Figure 5).

We have presented a long-term open proposal to develop language technology for lesser-used languages. We have defined four phases that involve parallel and coordinated creation of language foundations, tools and applications. The development of a sound language industry should be the result of a synchronized effort, involving research groups, institutions and industry. This way, language foundations and research will be useful to create any tool or application and, conversely, tools and applications will be very helpful in research and improving language foundations

Now that we have started working on the fourth phase, every foundation, tool and application developed in the previous phases is of great importance to face new problems. We have presented several commercial applications, including the spelling checker and the lemmatizer, which are very active tools in the standardization process of Basque.

### 4.3. A lemmatization based web-crawler

GAIN is a tool for Internet/Intranet search engine composed by a robot, an indexer and a client for queries. It contains two NLP modules that make it a powerful tool: a language identification module and the lemmatizer.

The Swish-E tool<sup>2</sup>, from the University of Berkeley, has been adapted, as it provides the required features: modularity, completeness, freeware, multiplatform and multiformat. It has been installed on a Web Linux-Apache server.

The most important improvement we have made has been a lemma-based indexing. The lemmatizer has been integrated and, as a result, we have obtained both an advanced search-engine and a more compact index. In the future we plan to augment the system indexing only the most representative lemmas and phrases following our work on terminology extraction.

### Acknowledgements

This work has been supported by the Department of Economy of the Government of Gipuzkoa, The University of the Basque Country, the Department of Education of the Basque Government and the Commission of Science and Technology of the Spanish Government.

### References

Aduriz I., Agirre E., Aldezabal I., Alegria I., Arregi X., Arriola JM., Artola X., Goienola K., Maritxalar A., Sarasola K., Urkia M.] "A Word-grammar based morphological analyzer for agglutinative languages". Proceedings of COLING 2000 Saarbrücken, Germany, August, 2000

Aduriz I., Aldezabal I., Alegria I., Artola X., Ezeiza N.,

# Adibidea: Kyoto



- [www.kyoto-project.eu](http://www.kyoto-project.eu)
  - *Yielding Ontologies for Transition-Based Organization*
- Helburuak:
  - Ezagutza partekatzea hizkuntzen eta kulturen artean
  - Bilaketa semantiko sakona eta testu-erazketa intentsiboa
  - Wiki-ingurunea adituak haien ezagutza sar dezaten informatikarien laguntzarik gabe
  - Ingurumenean aplikatua baina zabaltzekoa
- Ezaugarriak
  - Software librea
  - Europar proiektua; partaideak:
    - Herbehereak, Italia, Alemania, Euskal Herria
    - Unibertsitateak, enpresak, administrazioa eta ingurumen-taldeak
    - Taiwan eta Japonia, bertatik finantzatuta



# Kyoto: baliabide semantikoak

- WordNets

- English Wordnet
- English Wordnet Environment
- Japanese Wordnet
- Chinese Wordnet
- Italian Wordnet
- Dutch Wordnet
- Spanish Wordnet
- Basque Wordnet

- Ontologiak

- Sumo
- Onto Wordnet



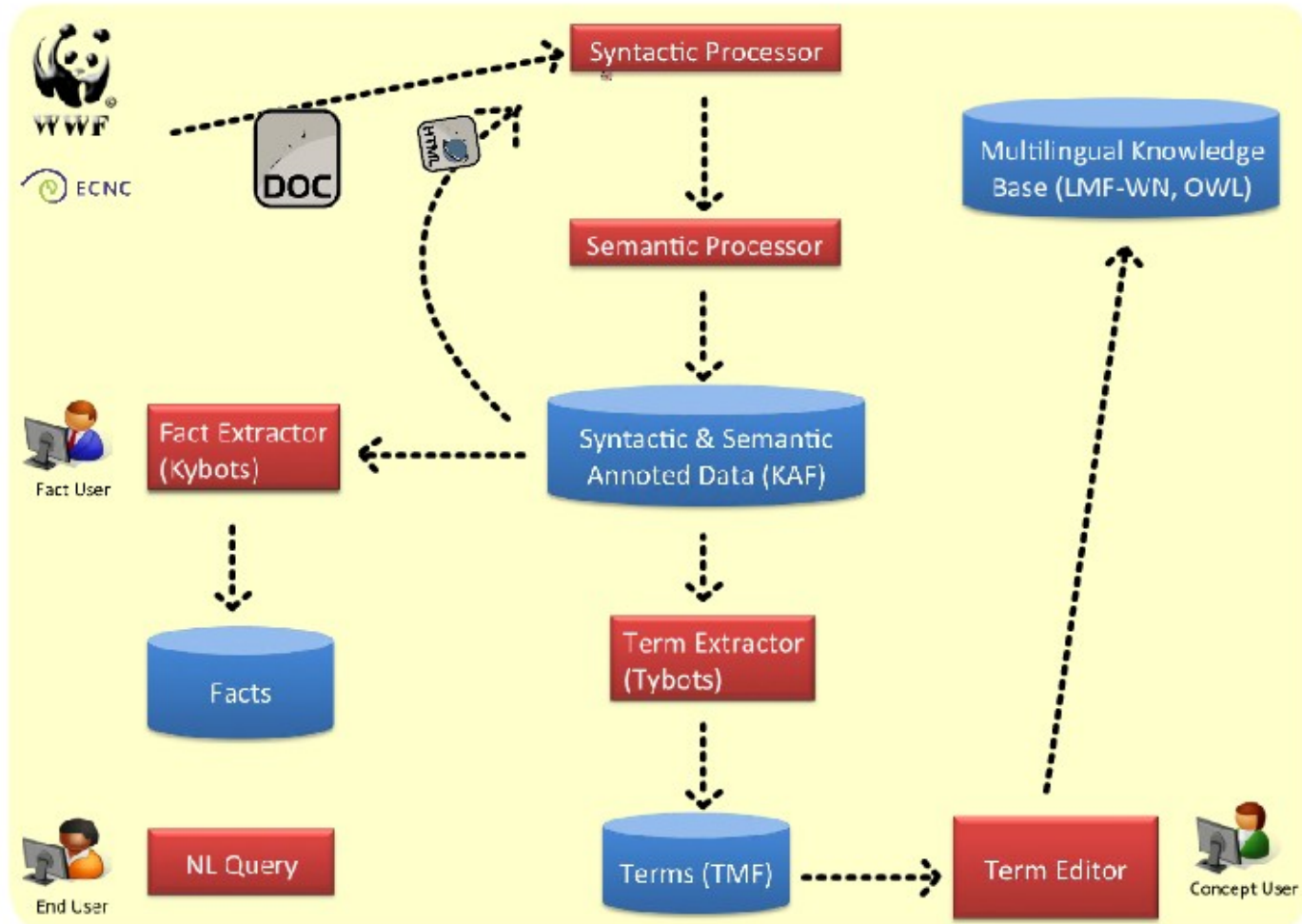
# Kyoto:

## tresna esanguratsuenak

- Tybot-ak: terminoen erauzlea/antolatzailea
  - terminoen editorea
- Kybot-ak: gertaeren erauzlea, sintaxian eta semantikan oinarrituta
- Wikyoto: ezagutzaren editorea, semantikan oinarritua
- PipeT: HTrako arkitektura malgua (eleanitza)
  - morfologia, sintaxia, entitateak
- KAF formatua (XML): HTen integraziorako

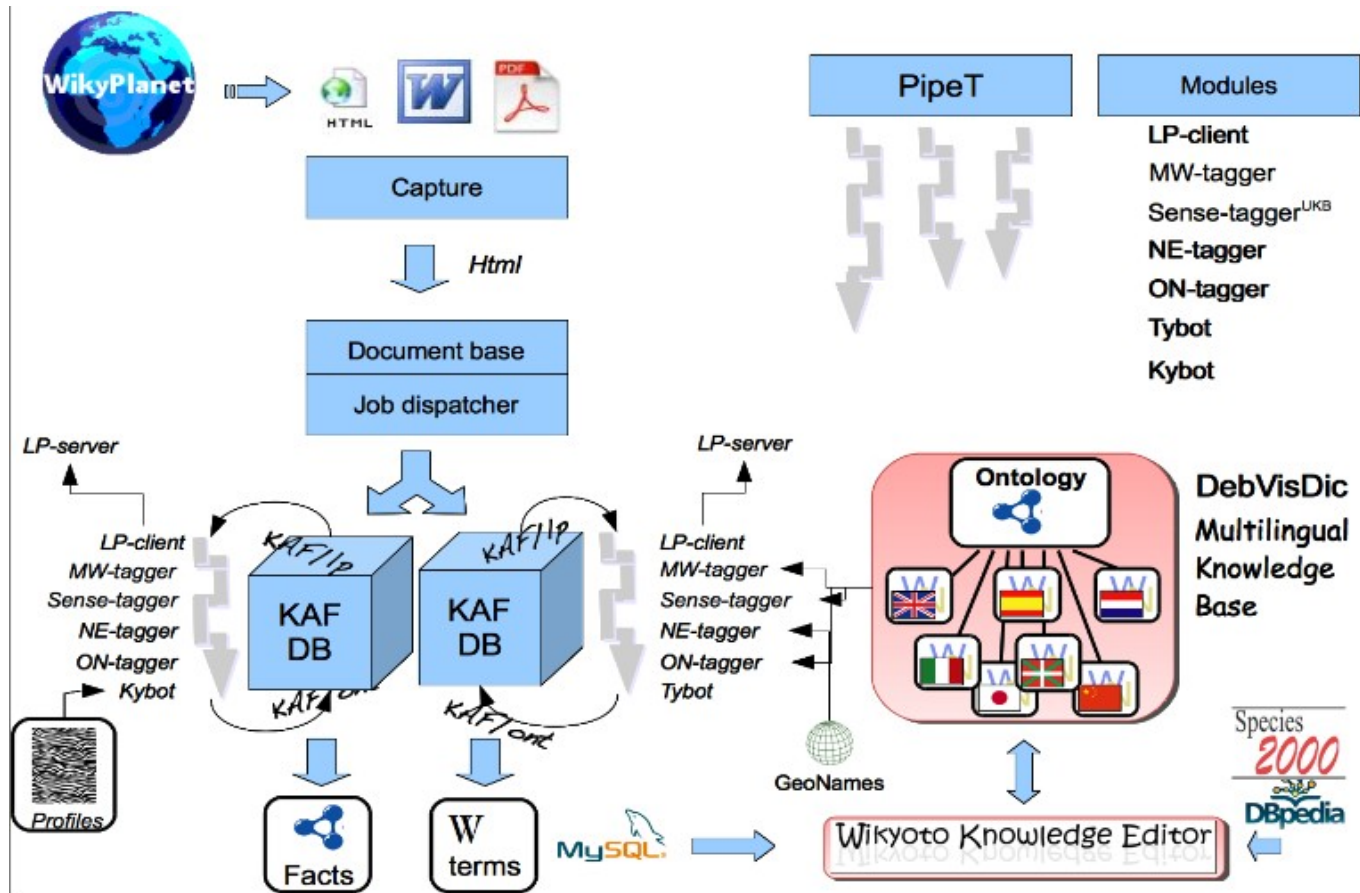
# Kyoto: arkitektura

*Kyoto System Architecture*



# Kyoto: integrazioa

## Bideoa





# Kyoto: bilaketa semantiko

- Aukera batzuk:
  - Tokien identifikazioa (mapan islatua)
  - Tokien sailkapena
  - Datak eta egutegi moduko aurkezpena
  - Propietateak:
    - kopuruak, erlazioak, prozesuak
  - Terminoen hierarkia
  - Eleaniztasuna
  - Zoom konbinatua (datak eta tokiak adib.)





**busuu.com** [Take a Tour!](#) 🇺🇸 🇬🇧 🇫🇷

**The language learning community - join now for free!**

**login**  
E-mail:   
Password:   
Remember me:  **login**  
[Forgot your password?](#)

**Ich möchte Deutsch lernen** **Je veux apprendre l'espagnol** **To parler français avec moi?** **I want to learn English!**

**Connect with native speakers** and learn directly from other members of the busuu.com community!

**Be completely flexible** and learn only what you really need!

**Have fun** by experiencing a new way of learning languages and forget those boring grammar books!

**Join now for free!**



# Irakaskuntza/Ikaskuntza (ICALL)

- Lerro nagusiak
  - Autoikasketa/Laguntza (ikasle zein irakasleentzat)
    - Autoebaluazioa
    - Erroreen detekzioa
  - Komunitateak (2.0)
- Aplikazioak IXA taldean
  - Idazketarako laguntzak:
    - arbela digitala (Idazlagun/ELET)
    - autoebaluazioa (prototipoak eu, es)
  - Ariketen sorkuntza automatikoa: *ArikIturri*
- Oinarriak:
  - Zuzentzaileak
  - Corpusak ikasketa automatikorako: Ikasleen testuak
    - Etiketatuak: Erroreei buruzko informazio linguistikoa eta teknikoa

# Idazlagun/ELET

Apkazioak Lekuak Sistema ?

Berria.info - Bizkaiko Golkoan antxo IdazLagun - Mozilla Firefox

Fitxategia Editatu Ikusi Historia Delicious Laster-markak Tresnak Hiztegiak Laguntza

http://sisx04.si.ehu.es/idazlagun/idazlagun.html

Aukerak EN EU Euskalterm Elhuyar Morris Open-tran Batua OEH Harluxet Wikipedia Mokoroa Intza EPG Klasikoak ZT Corpusa CorpEus

WordNet - Wikipedia, the fr... IdazLagun

Fitxategia Edizioa Formatua Laguntza tresnak

Dokumentu berria 2

Hil honen hasieran Kofradiakoek Europako Erakundeei antxo harrapatzeko kuota bikoiztea proposatu z.

Zuzendu Testu formatu normalera itzuli

Hitzen zuzenketa

- Zuzena da?
- Hitz okerrak zuzentzeko proposamenak

Hitzen deklinabidea

- Deklinabide guztia
- Kasuaren deklinabidea
- Kasuaren eta numeroaren deklinabidea

Hiztegiak

- Elhuyar hiztegia
- Sinonimoak
- Morris hiztegia

Data, ordu eta zenbakien

- Daten idazkera
- Orduen idazkera
- Zenbakien idazkera

Aditza

- Aditz laguntzaileak
- Aditz trinkoak
- Komak zuzen jartzeko
- Esaldi mailako egitura
- Erabilera adibideak lotu
- Euskal WordNet: Sinonimoak

Adizkiak

Adizki laguntzaileak

Bilaketa

Paradigma

Zer-nori-nork

Nor

Hura

Nork

Haiek

Nori

Haiei

Hitanoa

Modua/Denbora

Indikatibozko lehenaldia. A...

Emitza

Adizkia	Nor	Nork	Nori	Modua/D...
zieten	HURA	HAIEK-K	HAIEI	Indikatibo...

Bilatu

Utzi

## Eleka Demo

EUSKARAZ LAN EGITEKO TRESNAK



### Menua:

- Hasiera
- Aditzen jokatzea
- Deklinabidea
- Erabilera-adibideak
- Gramatika
- Hiztegiak
- Nola itzuli dute
- Laguntza

### Datak eta orduak nola idatzi

Urtea: 2010 Hilabetea: 07 Eguna: 26 Ados

Data idazteko modu erabilgarrienak eta gomendagarrienak honakoak dira:  
**2010eko uztailaren 26a**  
**2010eko uztailaren 26an**

Eta toki-izena eranstean bazaie horiei:  
**Donostia, 2010eko uztailaren 26a**  
**Donostian, 2010eko uztailaren 26an**

Informazio gehiago [hemen](#)

Ordua: 0 Minutuak: 00 Ados

### Aditzen jokatzea

Aditz trinkoa

Oin Forma: \*edin  
Nor: NI  
Nork:  
Nori:  
Hitanoa:

### Zenbakiak testuz idatzi

Zenbakia: 7596 Ados

**zazpi mila, bostehun eta laurogeita hemezortzi**

Informazio gehiago [hemen](#)

### Deklinabidea

Hitza: etxe  
Kasua: nor (absolutiboa)  
Mugatasuna: DENAK

Sortu

### Ereduzko Prosa

Nola erabiltzen da...?

Bilatu

### Hiztegiak

Hizkuntza: Gaztelania

Hitza:

# Autoevaluazioa (prototipoa)

Client side

User

Graphic Interface

Linguistic Process

Measure Manager

Database Manager

Communication Interface PHP

Socket

Communication Interface C++

Request Manager

Spanish module

Basque module

Other languages modules

Text analysis

Error Manager

Text analysis

Error Manager

Linguistic process

Linguistic process

freeling.h

liborto

Criteria module

Morphosyntactic analyzer

Constraint grammar

Criteria module

Gráficas

Miércoles, 09 de Julio de 2008

Hola lola!!!

Título: El medio ambiente

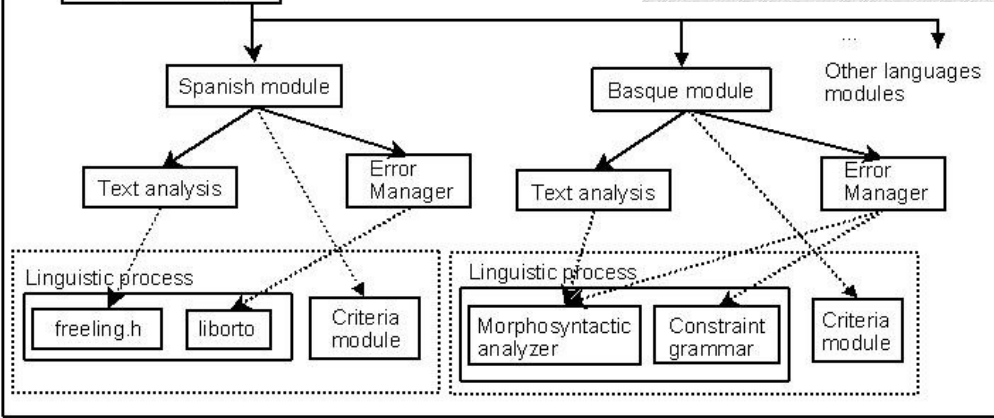
Número de Palabras: 248  
 Número de Vocablos: 115  
 Frases Cortas: 0  
 Frases Largas: 7  
 Riqueza Léxica: No  
 Redundancia en el texto: Si  
 Monotonía en el texto: Si  
 Concurrencia de Conjunciones: No  
 Errores Ortográficos: No  
 Errores de Acentuación: Si


Indicadores

Medio ambiente, conjunto de elementos abióticos (energía solar, suelo, agua y aire) y bióticos (organismos vivos) que integran la delgada capa de la Tierra llamada biosfera, sustento y hogar de los seres vivos. Dentro del medio ambiente tenemos la atmósfera que protege a la tierra del exceso de radiación ultravioleta y permite la existencia de vida es una mezcla gaseosa de nitrógeno, oxígeno, dióxido de carbono vapor de agua, otros elementos y compuestos, y partículas de polvo. Calentada por el sol y la energía radiante de la Tierra, la atmósfera circula en torno al planeta y modifica las diferencias térmicas. También los problemas medio ambientales como son El dióxido de carbono de azufre y otros contaminantes emitidos por las chimeneas de las industrias contribuyen a la contaminación de la atmósfera. El dióxido de carbono contribuye al calentamiento global, y el dióxido de azufre es la principal causa de las lluvias ácidas en el norte y este de Europa y en el noreste de Norteamérica, uno de los impactos del uso de combustible ha producido sobre el medio ambiente ha sido el aumento de la concentración de dióxido de carbono (CO2) en la atmósfera. A pesar de los cambios económicos y políticos, el interés y la preocupación del medio ambiente aún es importante. La cualidad del aire ha mejorado, pero estan pendiente de solución y requieren una acción coordinada los problemas de las lluvias ácidas; los clorofluorocarbonos, la pérdida de la capa de ozono y la enorme contaminación atmosférica.

Categorías Gramaticales:  
 Sustantivo: 72  
 Adjetivo: 21  
 Adverbio: 6  
 Verbo: 33  
 Preposición: 38  
 Conjunción: 19  
 Artículo: 38  
 Interjección: 2  
 Contracción: 7  
 Pronombre  
 Numeral

Volver a Registro  
 Redactar  
 Texto Original  
 Volver a Evaluados





# (Auto)ebaluazioa: Automated Essay Scoring (AES)

- Criterion/e-rater: [www.ets.org/criterion](http://www.ets.org/criterion)
  - erakunde sasipublikoa
  - arrakasta handia
  - inbertsio handia
  - HT eta ikasketa automatiko sakona
- Sariak
  - <http://criterion28.ets.org/news/>
- Azterketak eta testak
  - [www.ets.org/tests\\_products](http://www.ets.org/tests_products)
- Tour

Your essay:

- States your position clearly. Your reasons or examples are right for your argument and are working as pertinent support some of the time.
- Is organized in a generally unified manner, but making stronger connections among the parts would make this essay tighter and more convincing. Those kinds of changes are easy to make with a lot of payback for the effort.
- Sticks pretty much to simple sentences, when varied sentence structures could be very effective, and your word choice is mostly accurate.
- Demonstrates that grammar, spelling, and punctuation are things that you know, and mistakes in these features do not interfere with reader understanding.

**Holistic Score:** 4 out of 6 [View Score Analysis](#)

Look carefully at the *grammar, usage, and mechanics* feedback for this essay and check the "Writer's Handbook" for each feedback message. Working on the grammar, usage, and mechanics in the essay may help you improve it.

**Trait Feedback Analysis**

Grammar:	7 errors	 <a href="#">View Grammar results</a>
Usage:	4 errors	<a href="#">View Usage results</a>
Mechanics:	5 errors	<a href="#">View Mechanics results</a>
Style:	18 comments	<a href="#">View Style comments</a>
Organization & Development:		<a href="#">View Organization &amp; Development comments</a>

Click on each bolded item below to see the corresponding feedback.

**Summary of Grammar Errors**

**Fragment or Missing Comma**

**Run-on Sentences**

Garbled Sentences

Subject-Verb Agreement

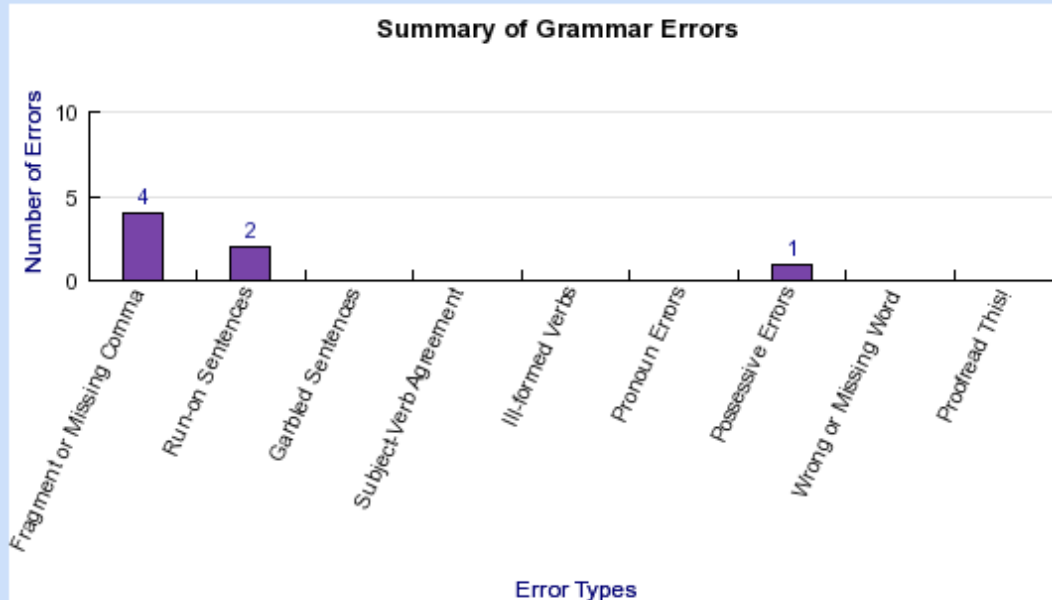
Ill-formed Verbs

Pronoun Errors

**Possessive Errors**

Wrong or Missing Word

Proofread This!



186 words

- View Score Analysis
- Print Expanded Performance Summary Report
- Print Combined Feedback Report...
- Close Report

Remember, for more information, click on the Writer's Handbook link for each feedback message.





# Norberak ikasi sare sozialen bidez

- Oso arrakasta handia azken urteotan
  - Aukera zabala, hizkuntza asko
    - <http://answers.yahoo.com/question/index?qid=20100618122617AAIgdse>
- Adibide batzuk
  - [www.livemocha.com/](http://www.livemocha.com/)
    - [www.livemocha.com/content/learn\\_more/active/en-us/sourceid:](http://www.livemocha.com/content/learn_more/active/en-us/sourceid:)
  - [www.busuu.com/tour](http://www.busuu.com/tour)
  - [www.byki.com/](http://www.byki.com/)
  - [www.babble.com](http://www.babble.com)
- Funtzio nagusiak
  - ariketak-ebaluazioa
  - multimedia (hitz egin, entzun, skype)
  - trukea (irakasle/ikasle)
- Negozio-eredu egokia: doan/ordainpekoa

# Adib.

Fitxategia Editatu Ikusi Historia Delicious Laster-markak Tresnak Hitz...

← → ↻ × 🏠 📄 TAG  🌐

Translate English → Euskara Sideb...

Criterion Writ... Criterion Writ... criterion writi... Languag...

UNIT 4 > Lesson 2 > Time

Required Exercises

- Listen
- Write
- Read
- Speak
- Listen Exercise
- Read
- Listen
- Write
- Read
- Speak

### Days of the Week

Please write one sentence for each day. Please be creative and use the words you have learned.

2016-10-20 10:00:00

0:58 / 2:03 360p

"Speaking" exercises will help you improve your pronunciation and allow you to practice sentence constructions so that you can

### Peer Review (English)

Thank you for your submission. You can help other members of the Livemocha community by rating and commenting on their submissions. Please review this submission.

#### Describing People

Describe six different people, using two descriptive words each: a woman, a man, a boy, a girl, you and me. Ex. The boy is young and tall.

Submitted by **Joana Pedro de Oliveira** Average Score: ★★★★★  
Speaks: Portuguese (Brazil) | Native  
Learning: English | Beginner

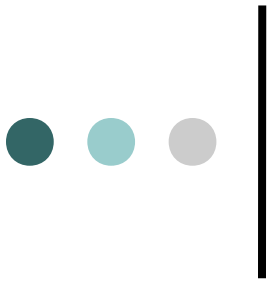
I am a woman  
Submitted: Nov 4, 2009, 2:04 pm

#### Review this submission

Spelling: ★★★★★  
Quality: ★★★★★

In return, you can help others learn your language by providing helpful comments on their exercises.

1:25 / 2:03 360p



Eskerrik asko

[i.alegria@ehu.es](mailto:i.alegria@ehu.es)

[ixa.si.ehu.es](http://ixa.si.ehu.es)

[ixa.si.ehu.es/master](http://ixa.si.ehu.es/master)