



Semantikan oinarritutako bilaketak: *Kyoto* proiektua

Iñaki Alegria, German Rigau

Ixa taldea

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





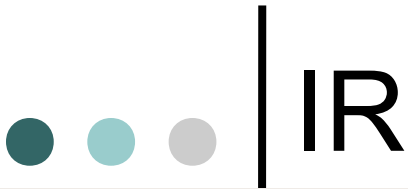
Aurkezpena

- Sarrera
- Informazioa eskuratzea eta erauztea.
Hizkuntza-ingeniaritzaren laguntza.
 - Oinarrizko tresna linguistikoak
 - CLIR (eleaniztasuna)
 - QA (galdera-erantzunak)
- Semantikan oinarritutako eskuratzea
 - Oinarria
 - Ebaluazioa
 - Baliabideak
- Kyoto proiektua



IR: Informazioa eskuratzea

- Bilaketen teknologia
 - Internet: *Google, Yahoo, Elebila, ...*
 - Intranet: Hemerotekak, liburutegi digitalak, ...
- Teknologia
 - Tresna orokorrak (plug-in/gehigarriak):
 - Jabedunak: *Autonomy, Google...*
 - Libreak: *Lucene, ...*
 - Neurrira egindakoak
- Aldaerak
 - CLIR, IR multimodala,
 - (CL)QA
 - Informazioaren erauzketa: datu-egituratuak lortzea



leioa - fitxategi-arakatzaillea

hemerote...

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 Inudiak 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.unirioja.es/servlet/articulo?codigo=1291113 - 11k - [Katxean](#) - [Antzeko orriak](#)

IV Col·loqui Aula d'Historia del Periodisme Diari de Barcelona ...
Hemeroteques digitals: experiències. Autores: Lluís Codina Bonilla; Localització: Treballs de comunicació, ISSN 1131-5687, N.º 16, 2001 , pags. ...
dialnet.unirioja.es/servlet/articulo?codigo=271331 - 10k - [Katxean](#) - [Antzeko orriak](#)
dialnet.unirioja.es domeinuan emaitza gehiago >

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 - 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. ...
montcadaconfidencial.blogspot.com/2007/05/segueixen-les-hemeroteques.html - 167k - [Katxean](#) - [Antzeko orriak](#)
montcadaconfidencial.blogspot.com domeinuan emaitza gehiago >

[PDF] **HEMEROTEQUES. ESTAT ACTUAL I PERSPECTIVES**
Formatua: PDF/Adobe Acrobat - [HTML bertsiog](#)
HEMEROTEQUES. ESTAT ACTUAL I PERSPECTIVES. PROYECTO DE DIGITALIZACIÓN DE, PrensA EN LA BIBLIOTECA NACIONAL, Dolores Rodríguez, ...
eprints.rclis.org/archive/00004465/01/nacional.pdf - [Antzeko orriak](#)

[PDF] **HEMEROTEQUES. ESTAT ACTUAL I PERSPECTIVES**
Formatua: PDF/Adobe Acrobat - [HTML bertsiog](#)
HEMEROTEQUES. ESTAT ACTUAL I PERSPECTIVES. HEMEROTECA DE LA BIBLIOTECA PÚBLICA. DEL ESTADO AZORÍN DE ALICANTE. FICHA TÉCNICA: ...
eprints.rclis.org/archive/00004495/01/alicante.pdf - [Antzeko orriak](#)
eprints.rclis.org domeinuan emaitza gehiago >

[PDF] **HEMEROTEQUES. ESTAT ACTUAL I PERSPECTIVES**
Formatua: PDF/Adobe Acrobat - [HTML bertsiog](#)
HEMEROTEQUES. ESTAT ACTUAL I PERSPECTIVES. LA HEMEROTECA MUNICIPAL DE

Elebila - Mozilla Firefox

Fitxategia Editatu Ikusi Historia Laster-markak Tresnak Hiztegiak Laguntza

http://www.elebila.eu/search/?bilatu=hemerotekek&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 hemerotekek **BILATU** Bilaketa aurreratu Hobespenak

Euskarazko web orrietan Edozein hizkuntzatan

Emaitzak: 36700 orri

Erabilitako analisia: *hemeroteka* 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 Katxean

Untitled Document
...
http://www.uztaro.com/berria/laburpena.cfm?
labur=fitxategiak\Uztaro60\5_Lizaso.txt Katxean

Ameriketako euskaldunei buruzko webgunea - Website on Basques ...
Hemeroteka: Berri gehiago : Laburrak: Agurrak: d: dd ... Buffalon XX mende hasierakoa bada ere, kluba ez zen 1980 ... Bertan hautatzen da urterako lehendakaria eta zuzendaritza taldea.
http://www.basqueheritage.com/berria/eusk/index.php?id=fit&idf=56&PHPSESSID=bca39099488c1b0614d9273df93fc617 Katxean

Kritikak (hemeroteka)
...
http://www.susa-literatura.com/kritikak/argia/krit0262.htm Katxean

Kritikak (hemeroteka)
...
http://www.susa-literatura.com/kritikak/krit0399.htm Katxean

UZEI - Terminologia eta Lexikografia Zentroa
Hemeroteka ... hizkuntzaren egungo bizkortasunaren seinalea da, eta etorkizunean irauteko bermea. Hori 1977an horrela esaten ez bazen ere ...
http://www.uzei.com/antbuspre.asp?nombre=1682&cod=1682&sesion=14 Katxean

UZEI - Terminologia eta Lexikografia Zentroa
...
http://www.uzei.com/antcatalogo.asp?nombre=1701&hoja=0&sesion=14 Katxean

Bide Ertzean: Hemeroteka
Hemeroteka Mondo Sonoro aldizkaria. 06 ... baita beste toki askotan ere. "Non dira"

Opentrad
Itzulpen automatikoa kode irekian

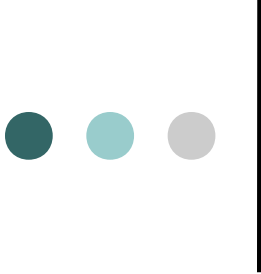
XUXENweb

[Zure enpresa ager daiteke hemen](#)
Zure enpresa hemen ikusi nahi baduzu idatzi info@eleka.netera
<http://www.eleka.net>

[Zure enpresa ager daiteke hemen](#)
Zure enpresa hemen ikusi nahi baduzu idatzi info@eleka.netera
<http://www.eleka.net>

[Zure enpresa ager daiteke hemen](#)
Zure enpresa hemen ikusi nahi baduzu idatzi info@eleka.netera
<http://www.eleka.net>

[08 - So Far Away.mp3] leioa - fitxategi-araka... hemeroteka - OpenO... hemeroteka2 - Open... IXA2007Aurkezpena ... hemeroteques - Goo... Elebila - Mozilla Firefox



Hizkuntza-ingeniaritza

- morfologia/lematizazioa/*stemming*
 - **estaldura** handitzeko (hizkuntzen arabera) **doitasuna** galtzeke
 - galderen akatsak zuzentzeko
- sintaxia + estatistikak
 - terminologia, pertsona/toki/erakunde izenak
 - informazio interesgarriena eskaintzeko
- semantika
 - informazioa erlazionatzeko
 - estaldura handitzeko (epaile/magistratu)
- eleaniztasuna (hiztegiak, itzulpen automatikoa)
 - dokumentazio eleanitzak / erabiltzaile eleanitzak
 - hizkuntza ez-ezagunak
- modu multimodala (hizketaren ezagutza, irudien sailkapena)
 - irudi, hizketa, bideo gainean bilatzea

Morfologian oinarritutako aplikazioak

- Lematizazioa/*stemming*
 - lan egiteko modua:
 - 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
lingüística

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

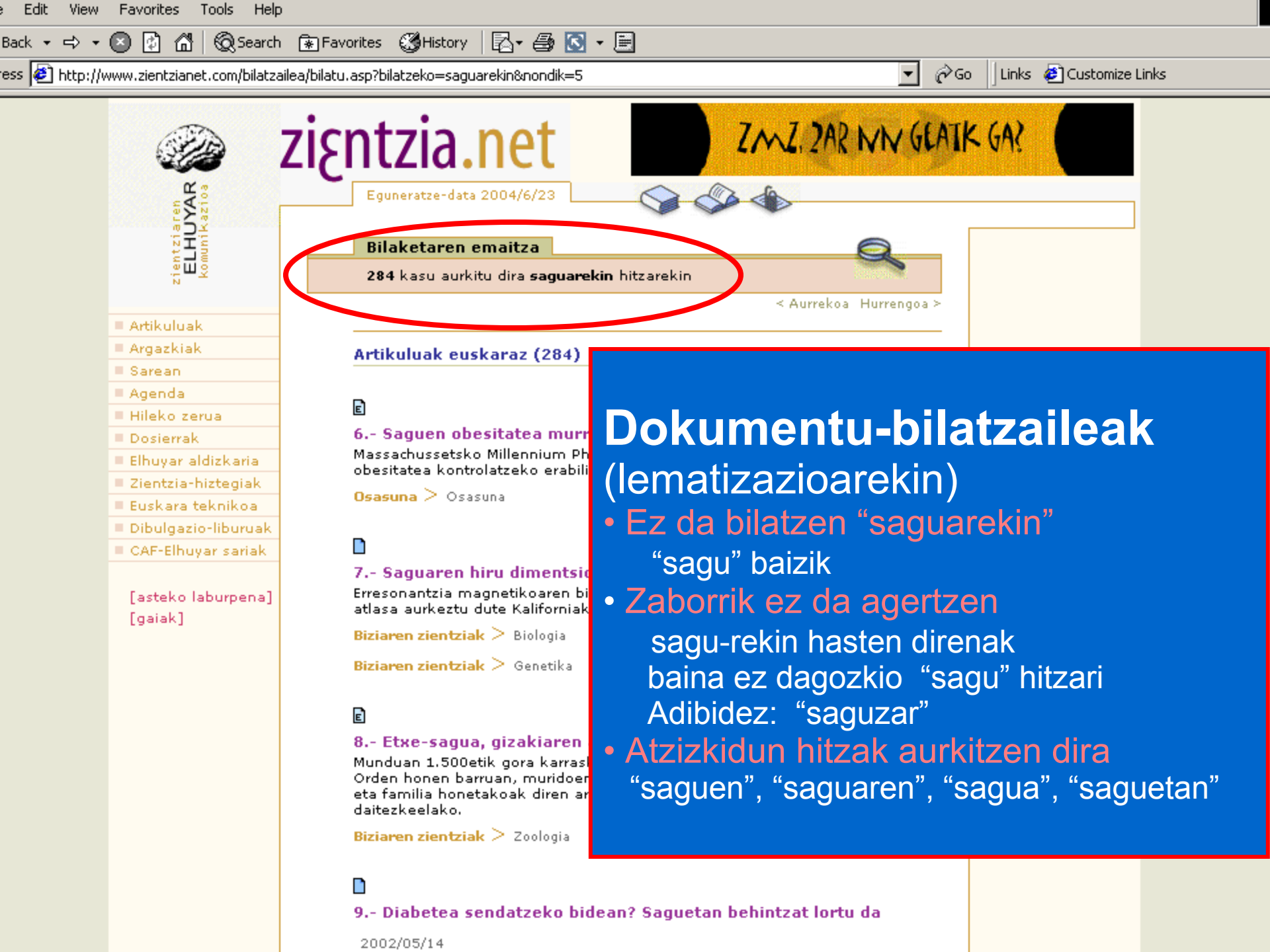
Search

Scholar All articles - [Recent articles](#)

Did you mean: [diachronic linguistics](#)

[Arabic Phonology](#)

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



Artikuluak euskaraz (284)



6.- Saguen obesitatea murr...

Massachussetsko Millennium Ph...
obesitatea kontrolatzeko erabili...

Osasuna > Osasuna



7.- Saguaren hiru dimentsio...

Erresonantzia magnetikoaren bi...
atlasa aurkeztu dute Kaliforniak...

Biziaren zientziak > Biologia

Biziaren zientziak > Genetika



8.- Etxe-sagua, gizakiaren...

Munduan 1.500etik gora karras...
Orden honen barruan, muridoer...
eta familia honetakoak diren ar...
daitezkeelako.

Biziaren zientziak > Zoologia



9.- Diabetea sendatzeko bidean? Saguetan behintzat lortu da

2002/05/14

Dokumentu-bilatzaileak (lematizazioarekin)

- Ez da bilatzen “saguarekin”
“sagu” baizik
- Zaborrik ez da agertzen
sagu-rekin hasten direnak
baina ez dagozkie “sagu” hitzari
Adibidez: “saguzar”
- Atzizkidun hitzak aurkitzen dira
“saguen”, “saguaren”, “sagua”, “saguetan”



Sintaxian oinarritutako aplikazioak

- Informazio esanguratsua:
 - izen-sintagmak, terminoak eta izen propioak
- Funtzioa:
 - bilatzea
 - estekatzea
 - multzokatzea (*clustering*)
 - galdera fintzea
- Aplikazioak:
 - bistaratze bereziak
 - dokumentuen multzokatzea

IR aurreratua

The image displays three overlapping browser windows illustrating advanced search results for the query "Clinton Obama".

- Left Window (Clusty Search):** Shows search results categorized into clusters. The top cluster is "Hillary Clinton and Barack Obama" (28 results), with sub-clusters for "Debate" (28), "Pennsylvania" (22), "McCain" (24), "Bill Clinton" (12), "Obama, Edwards" (12), "Blog" (9), "Bush" (6), "Clinton-Obama ticket" (5), and "Clinton Answers" (4). A "remix" button and a "find in clusters" search box are also visible.
- Middle Window (Quintura):** Displays a word cloud of related terms such as "mocks", "feud", "clash", "bill", "complain", "campaign", "comparing", "ticket", "barack", "clinton", "obama", "rodham", "lead", "returns", "attacks", "senator", "camp", "questions", "debate", and "trade". A list of search results is visible on the right side of the window.
- Right Window (KartOO Metamotor):** Shows a semantic map (concept map) of related terms and sources. The map includes nodes for "report", "politics", "election", "news", "mccain", "vote", "barack", "points", "shows", "chart", "wordpress.com", "www.usnews.com", "www.nytimes.com", "www.washingtonpost.com", "www.cnn.com", "www.aol.com", "www.factcheck.org", "www.politifact.com", "www.truthdig.com", "www.nytimes.com", "www.usnews.com", "www.cnn.com", "www.aol.com", "www.factcheck.org", "www.politifact.com", "www.truthdig.com", "www.nytimes.com", "www.usnews.com", "www.cnn.com", "www.aol.com", "www.factcheck.org". A "Topics" sidebar lists related terms like "hillary clinton", "rival barack obama", "clinton and barack obama", "john mccain", "percentage points", "hillary rodham clinton", "york times", "hillary", "points", "rival", "barack", "chart", "election", "politics", "news", "contest", and "ticker". A "Sponsor" section mentions "Democrats Face Off".

Itzulpen automatikoa: *Matxin*

demo - OpenTrad - Mozilla Firefox

Eitxategia Editatu Ikusi Historia Laster-markak Iresnak 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 España

Portadan Espainia Ir Generada automáticamente hace 15 minutos

Portadan

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

☑ Berrien alertatuak

Atom RSS | Feeds-ei buruz

Higikorrenzat berriak

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), gertatuko den martxoaren 13an ekintza. Latinamerika-erangan 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 - 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, bonba 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 bilaketen

Orrialde hau izena aipatzea

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

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

Garbajosa-k soilqune uzten du ahal badu, Europarra jolastuko da
Marka - [eta 17 gai erlazionatu](#) »

José Coronado espainiar Grissom izango da Telecinco-n
ABC - [eta 23 gai erlazionatu](#) »

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

Berria da:

Ian Paisley	Isabel II
Martin Mcguinness	Dow Jones
Diego Maradona	Néstor Kirchner
Roger Federer	Auzitegi Gorena
San Luis Potosí	George W. Bush

http://www.opentrad.org/demo/libs/nabigatzailea.php?markatu=&norantza=es-eu&inurl=http://www.lagaceta.com.ar/vernotae.asp?id_nota=217363&titulo=El%20pa%C3%ADs%20m%C3%A1s%20...

Inicio Microsoft Mis doc... demo - ... sisx04... Posta E... Bandej... matxin... 14:12



CLIR

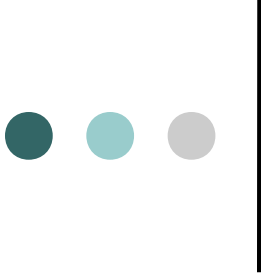
- IR hainbat hizkuntzetan
- Aukerak
 - galderak itzultzea
 - dokumentuak itzultzea
 - semantika bidez proiektatzea
- Gure aukera:
 - MCR
 - Etorkizunean itzulpen automatikoa
- Adibideak: *EFE (Meaning), ArgazkiPress*

CLIR (EFE): *Meaning* proiektua

The screenshot shows a Microsoft Internet Explorer browser window with the address bar containing `http://efe.irion.nl/efe_D/web/init.do?queryLg=en`. The search results are displayed in a table-like format with the following entries:

Relevance	ID	Category	Summary	Flag
75.0%	20040521	CATEGORÍAS SUPLEMENTARIAS: JUSTICIA-INTERIOR-SUCESOS/SUCESOS PALABRAS CLAVE: JUSTICE,ACCIDENTS CRIME INCENDIOS / INCENDIO EN FÁBRICA QUÍMICA, VALENCIA 2004. FUEGO / HUMO NEGRO / CARRETERA / COCHES CT	ACCIDENTS CRIME INCENDIOS/ INCENDIO EN F BRICA QU MICA , VALENCIA 2004 . FUEGO / HUMO NEGRO/ CARRETERA/ COCHES CT INCENDIO FABRICA : V. 11 2004 . Una inmensa columna de humo sale del incendio que se ha declarado esta tarde en una fábrica química dedicada al tratamiento del mármol en la localidad de San Antonio de Benagéber , a	
75.0%	20040521	CATEGORÍAS SUPLEMENTARIAS: JUSTICIA-INTERIOR-SUCESOS/SUCESOS PALABRAS CLAVE: JUSTICE,ACCIDENTS CRIME INCENDIOS / INCENDIO EN FÁBRICA QUÍMICA, VALENCIA 2004. FUEGO / HUMO NEGRO / TENDIDO ELÉCTRICO / CURIOSOS CT	ACCIDENTS CRIME INCENDIOS/ INCENDIO EN F BRICA QU MICA , VALENCIA 2004 . FUEGO / HUMO NEGRO/ TENDIDO EL CTRICO/ CURIOSOS CT INCENDIO FABRICA 2004 . Varios residentes de la urbanización adyacente al incendio que se ha declarado esta tarde en una fábrica química dedicada al tratamiento del mármol en la localidad de San Antonio de Benagéber , a	
58.0%	20040428	CATEGORÍAS SUPLEMENTARIAS: EMERGENCY PLANNING TERRORISMO SIMULACRO DE ATAQUE TERRORISTA CON ARMAS QUIMICAS EN NEWCASTLE BOMBEROS POLICIA JGB NO VENIDER EN DETINO UNIDO AIT IRLANDA		

At the bottom of the browser window, the status bar shows "Internet".

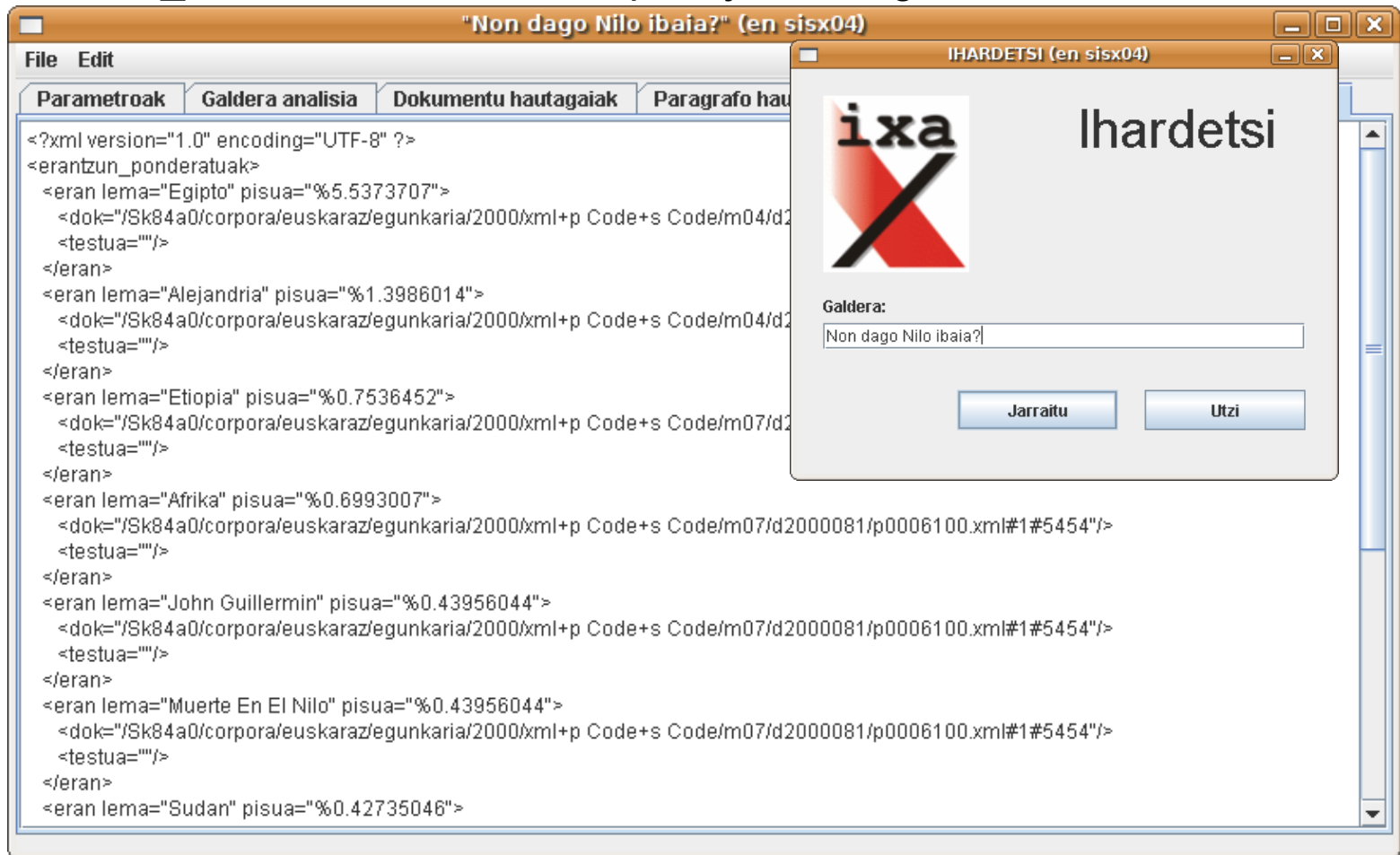


QA (galdera-erantzunak)

- Eraitza ez da dokumentua, erantzuna baizik
- Galdera motak
 - Faktoideak (nork, zer, non, noiz)
 - Definizioak
- IR + patroien erauzketa
 - galderen analisisa
 - pasarten berreskurapena ohiko IRz
 - erantzunaren bilaketa
- Arrakasta mugatua
 - *Ihardetsi* prototipoa

Ihardetsi - Anhitz


- http://teknopolis.elhuyar.org/ikusi.asp?Multi_Kodea=191&atala=erreportajeak&lang=EU



The screenshot displays a web application interface with two windows. The main window, titled "Non dago Nilo ibaia?" (en sisx04), shows a menu with "File" and "Edit", and tabs for "Parametroak", "Galdera analisia", "Dokumentu hautagaiak", and "Paragrafo hau". The content area displays XML data with the following structure:

```
<?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 Code/m04/d2
    <testua=""/>
  </eran>
  <eran lema="Alejandria" pisua="%1.3986014">
    <dok="/Sk84a0/corpora/euskaraz/egunkaria/2000/xml+p Code+s Code/m04/d2
    <testua=""/>
  </eran>
  <eran lema="Etiopia" pisua="%0.7536452">
    <dok="/Sk84a0/corpora/euskaraz/egunkaria/2000/xml+p Code+s Code/m07/d2
    <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">
```

The second window, titled "IHARDETSI (en sisx04)", features the Ixa logo (a red 'X' with a black diagonal line) and the text "Ihardetsi". Below the logo, it includes a search field labeled "Galdera:" containing the text "Non dago Nilo ibaia?". At the bottom of the window are two buttons: "Jarraitu" and "Utzi".



Semantikan oinarritutako eskuratzea (IR)

- Bi kontzeptu:
 - *web semantikoa*. Internet 2.0rekin lotuta
 - **ezagutza semantikoa ustiatzea**
 - testuak (gehiago) ulertzea
- Oinarriak
 - 1 hitz: n adiera (kontzeptu) / hiztegia
 - n adiera: 1 adiera / testuingurua
 - teknikak: terminologi erauzketa, analisi sintaktikoa, distantzia semantikoa, dentsitate semantikoa, inferentzia
 - baliabideak: hiztegiak, thesaurus-ak, wikipedia, DB semantikoak, ontologiak, ezagutza-baseak...



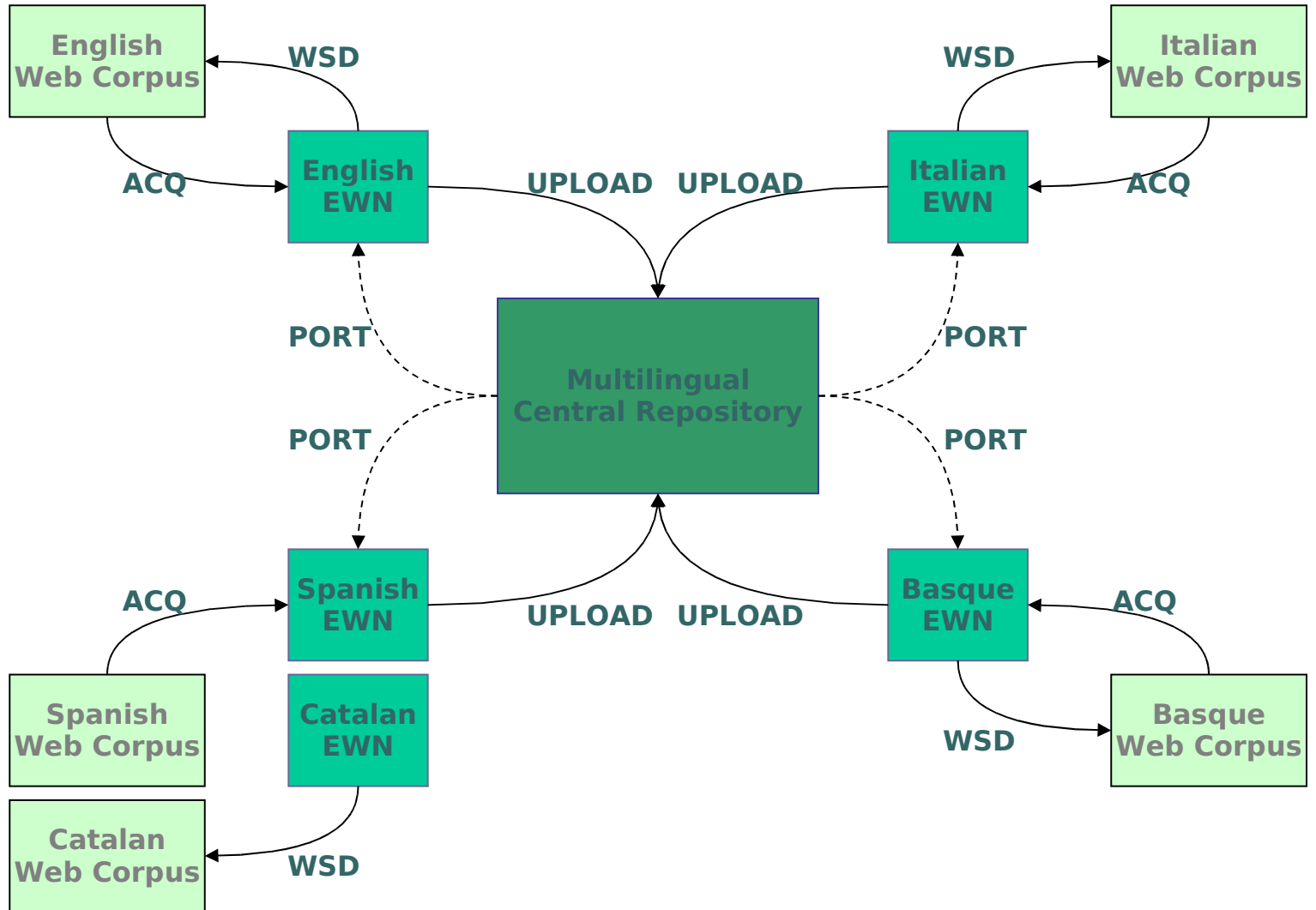
Baliabideak: MCR

(Multilingual Central Repository: Meaning proiektua)

- Hizkuntza desberdinetako kontzeptuak MCRn daude gordeta
- Erlazioa hizkuntzen artean eta kontzeptuen artean
- Kontzeptuen artean erlazio konplexuak lortzea edo inferitzea
 - *Zer egiten da kafearekin? Edan*



MCR



WordNet

Web MCR Interface MEANING - Microsoft Internet Explorer

Fitxategia Editatu Ikusi Gogokoak Tresnak Laguntza

Atzera Bilatu Gogokoak

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

Google eneko agirre

Word: java


Nouns: English_1.6


Synonyms: near_synonym English_1.6

Gloss
 Score
 Rels
 Full


English 1.6
 Spanish 1.6
 Catalan 1.6
 Basque 1.6
 Italian_1.6


English_1.7
 English_1.7.1
 English_2.0
 Catalan_1.5
 Spanish_1.5


 **meaning** Multilingual Central Repository


05948884n  11 coffee_1 java_2


[alimention](#) food

[Beverage+](#) 05948884n  17 café_1 a beverage consisting of an infusion of ground coffee beans: *he ordered a cup of coffee;*

[Comestible+](#) 05948884n  17 café_1 Bebida estimulante que se obtiene de los granos del cafeto

[Liquid+](#) 05948884n  12 kafe_2 Beguda estimulants que s'obté dels grans de l'arbre del cafè

[Natural+](#) 05948884n  12 caffè_1

[Substance+](#) 05948884n  12 caffè_1

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

Kyoto proiektua



- Izenburua: *Yielding Ontologies for Transition-Based Organizations*
- Helburuak:
 - Ezagutza partekatzea hizkuntzen eta kulturen artean
 - Bilaketa semantiko sakona eta testu-erazketa intentsiboa
 - Wiki-ingurunea adituak haien ezagutza sartzeko, informatikarien laguntzarik gabe
- Iraupena:
 - 2008ko martxoa – 2011ko martxoa
- Ekimena:
 - 30 pertsona-urte
 - Partaideak
 - Herbehereak, Italia, Alemania, Euskal Herria
 - Unibertsitateak, enpresak, administrazioa eta ingurumen-taldeak
 - Taiwan and Japonia bertatik finantzatuta

Kyoto (ICT-211423)



- Hizkuntzak:
 - English, Dutch, Italian, Spanish, Basque, Chinese, Japanese
- Domeinua:
 - Environmental domain, BUT usable in any domain
- Esparrua:
 - Both European and non-European languages
- Eskuragarritasuna:
 - Free: as open source system and data
- Etorkizunerako pentsatua:
 - Content standardization that supports world wide communication
 - Global Wordnet Grid

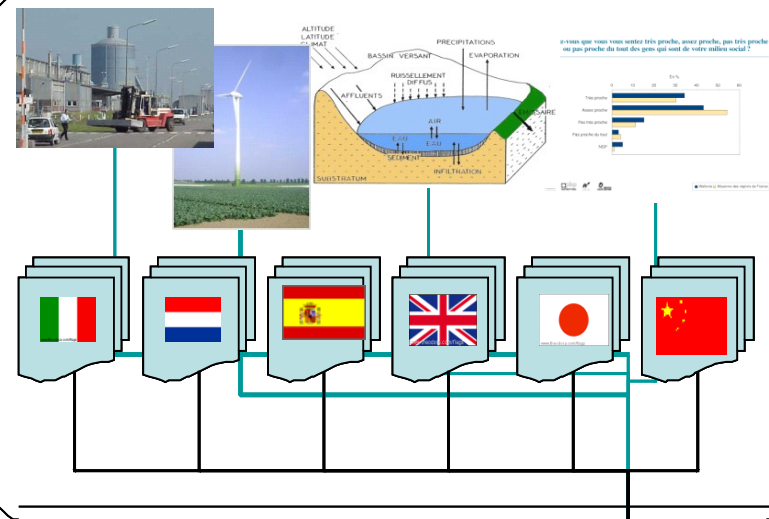
Kyoto: arkitektura



GREENPEACE



Environmental organizations

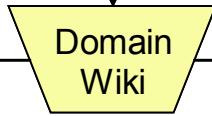


Governors Companies

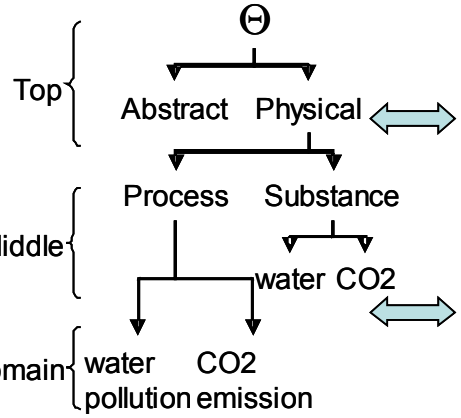
Environmental organizations



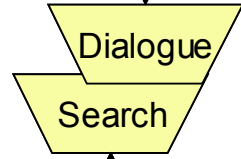
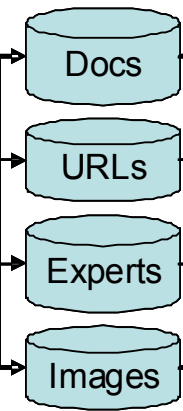
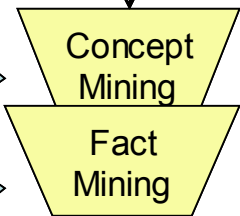
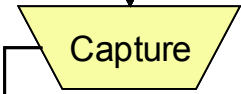
GREENPEACE



Universal Ontology



Wordnets



Informazio-beharrak: galderak



- *What companies produce a lot of damaging substances?*
 - Zein enpresek sortzen dituzte substantzia kaltegarriak?
- *Air pollution by traffic*
 - Trafikoak eragindako aire poluzioa
- *Sick because of air pollution*
 - Aire poluzioarengatik gaixotutakoak
- *Cause of air pollution*
 - Aire-poluzioaren kausak
- *Air pollution from the Ruhr area*
 - Ruhr inguruko aire poluzioa



Kyoto: osagaiak

- **Wikyoto:** komunitatetan oinarritutako wiki sistema domeinuko Wordnet-ak eta ontologiak eruzteko
 - Tybot-ak: terminoak eruzteko robotak (*Term Mining*)
 - Kybot-ak: gertaerak eruzteko robotak (*Fact Mining*)
- Baliabideak
 - *Wordnet, FrameNet*
 - DOLCE, BLC, TCO, SUMO, SIMPLE
 - Wikipedia, DBpedia, Geo databases, etc.
- Estandarrak
 - TMF, LMF, KAF, OWL, ...

Tybot-ak

The screenshot displays the TaxEdit application window. The title bar reads "TaxEdit" and the menu bar includes "File", "View", and "Help". The address bar shows the file path: "C:\Irrion\Twentyone 3.0\Index 3.1\Projects\Kyoto\Kyoto_3\1\classes_en.xml". Below the address bar, a status bar indicates "Bio:0.82#Fishing:0.8# Tops:206, Nodes:476, Docs:397". A toolbar contains buttons for "Import", "View", "Edit", "Save", "Sort", and "Clear". The main workspace shows a hierarchical tree structure of taxonomic terms. The root node is "external threats:2:1::3", which is highlighted in yellow. It branches into "Managing External Threats:2:1::1", "key external threats:1:1::1", and "threats of oil extraction:1:1::1". The "threats of oil extraction" node further branches into "reef:0::10", "largest known cold water coral reef:1:1::1", "important reefs:1:1::1", "Great Barrier Reef:1:1::1", and "Barrier Reef:6:2::2". The "reef" node branches into "coral reef:4:1::5" and "water coral reef:2:1::2". The "coral reef" node branches into "unique coral reefs:1:1::1" and "difficult coral reef:2:1::1". The "water coral reef" node branches into "cold water coral reef:2:1::1". The "habitats:1:1::9" node branches into "sensitive habitats:2:1::2", "affecting habitats:2:1::3", "marine habitats:2:1::2", and "multi-habitat:1:1::1". The "sensitive habitats" node branches into "harm sensitive habitats:2:1::1". The "affecting habitats" node branches into "activities affecting habitats:2:1::2" and "human activities affecting habitats:2:1::1". The "marine habitats" node branches into "sensitive marine habitats:1:1::1". The "species:8:2::7" node branches into "fish species:2:1::2", "multi-species:2:1::1", "one-species:2:1::1", and "marine species:2:1::1". The "fish species" node branches into "key fish species:1:1::1". The status bar at the bottom of the window displays "Done loading thesaurus..." and "100%".



Kybot-ak

- Kybot-ak aplikatzen dira analizatutako testuen gainean (KAF fitxategiak). Osagaiak
 - *Expression Rules (XPath-like)*
 - *Semantic Conditions (DEB API calls)*
 - *Output Templates (enrich KAF)*
- Prozesua:
 - *IF*
 - *Expression Rules match and*
 - *Semantic Conditions hold*
 - *THEN*
 - *generate the Output Template*



Wikybot

- [Home](#)
- [Upload documents](#)
- [Manage kybots](#)
- [Manage snippets](#)

Hello rigau!

Logout

Click on a text portion of a snippet, then select the role to associate. Click on a labelled snippet text portion to delete its role association.

Role colors : Agent Patient Location Time Topic

Document : *Living planet english*

--- The other index in this report, the *Living Planet Index*, shows a rapid and continuing loss of biodiversity - populations of vertebrate species have declined by about one third since 1970.

--- populations of terrestrial species declined by about 30 per cent on average between 1970 and 2003.

--- The rapid rate of population decline in tropical species is mirrored by the loss of natural habitat to cropland or pasture in the tropics between 1950 and 1990 , agricultural conversion being the main driver.



Semantikan oinarritutako bilaketak: *Kyoto* proiektua

Iñaki Alegria, German Rigau

Ixa taldea

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

