Tresna linguistikoak informazioa atzitzeko

Eneko Agirre, Iñaki Alegria

Ixa taldea

http://ixa.si.ehu.es

Aurkezpena

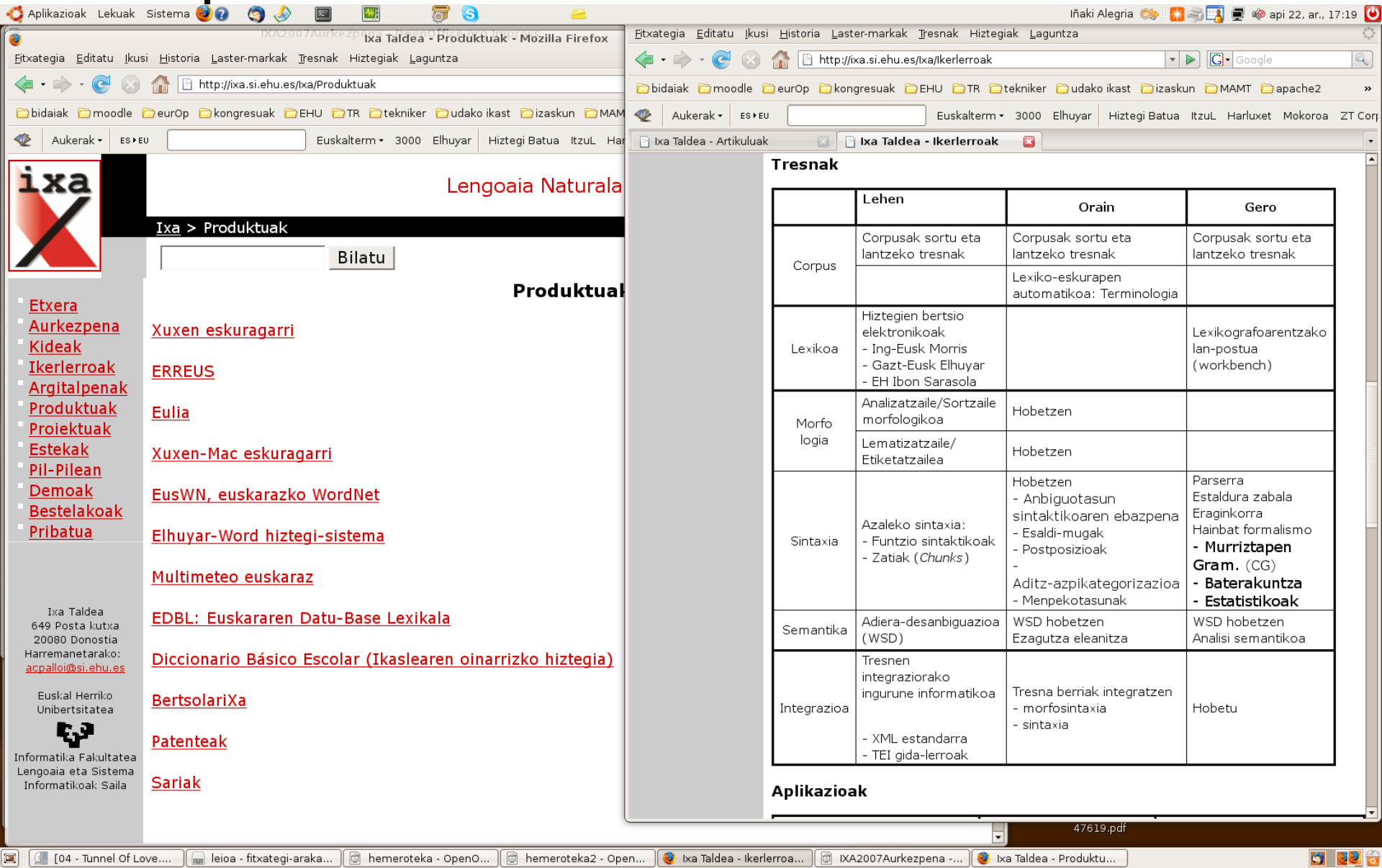
- Sarrera
- IR (informazioaren bilaketa)
- Oinarrizko tresna linguistikoak
- Morfologian oinarritutako aplikazioak
- Sintaxian oinarritutako aplikazioak
- CLIR (eleaniztasuna)
- IR multimodala
- QA (galdera-erantzunak)

Ixa taldea

Ixa taldea

- 6 esplotazio-lizentzia (patente)
- Spin-off enpresa: Eleka
- Hainbat produktu merkatuan (lankidetzan)
  - *Eleka, Elhuyar, Elkar, Euskaltel, ...*
  - *Microsoft, Scansoft, Eatoni, ...*
- Hainbat prototipo aurrekomertzial
- Adibideak:
  - *Xuxen* zuzentzaile ortografikoa: www.xuxen.com
  - Opentrad-Matxin itzultzaile automatikoa: www.opentrad.com
  - Beste asko

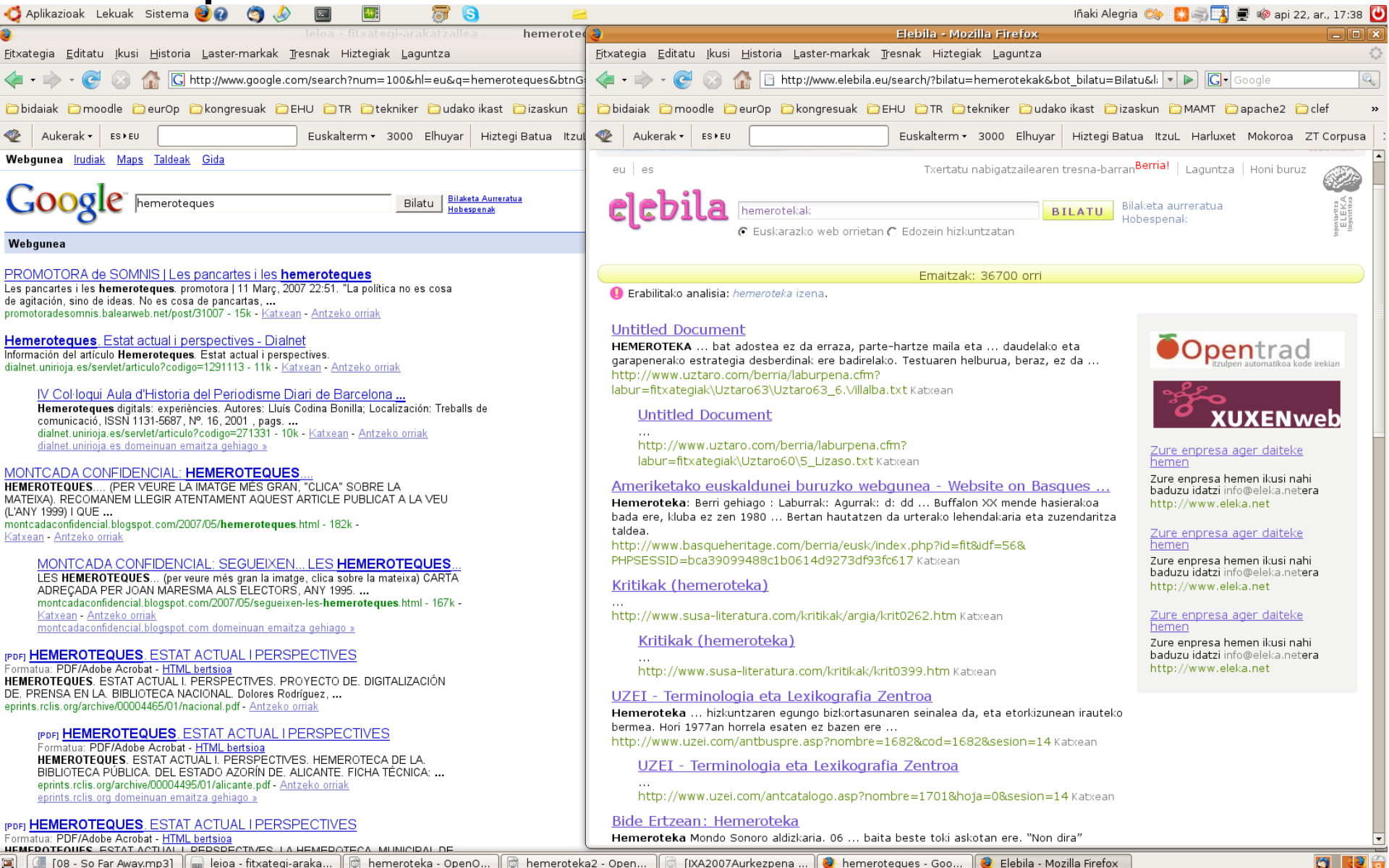
ixa.si.ehu.es


Ixa Taldea - Produktuak - Mozilla Firefox
http://ixa.si.ehu.es/Ixa/Produktuak
Lengoaia Naturala
Ixa > Produktuak
Bilatu
Produktuak
Xuxen eskuragarri
ERREUS
Eulia
Xuxen-Mac eskuragarri
EusWN, euskarazko WordNet
Elhuyar-Word hiztegi-sistema
Multimeteo euskaraz
EDBL: Euskararen Datu-Base Lexikala
Diccionario Básico Escolar (Ikaslearen oinarrizko hiztegia)
BertsolariXa
Patenteak
Sariak
Etxera
Aurkezpena
Kideak
Ikerlerroak
Argitalpenak
Produktuak
Proiektuak
Estekak
Pil-Pilean
Demoak
Bestelakoak
Pribatua
Ixa Taldea
649 Posta kutxa
20080 Donostia
Harremanetarako:
acpalloi@si.ehu.es
Euskal Herriko Unibertsitatea
Informatika Fakultatea
Lengoaia eta Sistema Informatikoak Saila
http://ixa.si.ehu.es/Ixa/Ikerlerroak
Ixa Taldea - Artikuluak
Ixa Taldea - Ikerlerroak
Tresnak
Lehen
Orain
Gero
Corpus
Corpusak sortu eta lantzeko tresnak
Lexiko-eskurapen automatikoa: Terminologia
Lexikoa
Hiztegien bertsio elektronikoak
- Ing-Eusk Morris
- Gazt-Eusk Elhuyar
- EH Ibon Sarasola
Lexikografoarentzako lan-postua (workbench)
Morfologia
Analizatzaile/Sortzaile morfologikoa
Hobetzen
Lematizatzaile/Etiketatzailea
Sintaxia
Azaleko sintaxia:
- Funtzio sintaktikoak
- Zatiak (Chunks)
Hobetzen
- Anbiguotasun sintaktikoaren ebazpena
- Esaldi-mugak
- Postposizioak
- Aditz-azpikategorizazioa
- Menpekotasunak
Parserra
Estaldura zabala
Eraginkorra
Hainbat formalismo
- Murriztapen Gram. (CG)
- Baterakuntza
- Estatistikoak
Semantika
Adiera-desanbiguazioa (WSD)
WSD hobetzen
Ezagutza eleanitza
WSD hobetzen
Analisi semantikoa
Integrazioa
Tresnen integraziorako ingurune informatikoa
- XML estandarra
- TEI gida-lerroak
Tresna berriak integratzen
- morfosintaxia
- sintaxia
Hobetu
Aplikazioak
47619.pdf

IR

- Bilaketen teknologia
  - Internet: *Google, Yahoo, Elebila*, ...
  - Intranet: Hemerotekak, liburutegi digitalak, ...
- Teknologia
  - Tresna orokorrak (plug-in/gehigarriak):
    - Jabedunak: *Autonomy*, ...
    - Libreak: *Lucene*, *Greenstone*, ...
  - Neurrira egindakoak
- Ebaluazioa:
  - *Precision* (doitasuna)
  - *Recall* (estaldura)
- Aldaerak
  - CLIR, IR multimodala, QA

IR


Aplikazioak Lekuak Sistema
Iñaki Alegria
api 22, ar., 17:38
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 Irudiak Maps Taldeak Gida
Google hemeroteques Bilatu Bilaketa Aurreratua 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.balearweb.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; Localización: Treballs de comunicació, ISSN 1131-5687, Nº. 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 MARESMA 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 bertsioa
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 bertsioa
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 bertsioa
Elebila - Mozilla Firefox
Fitxategia Editatu Ikusi Historia Laster-markak Tresnak Hiztegiak Laguntza
http://www.elebila.eu/search/?bilatu=hemerotekak&bot_bilatu=Bilatu&l
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 Corpusa
eu es
Txertatu nabigatzailearen tresna-barran Berria! Laguntza Honi buruz
elebila
hemerotekak
BILATU
Bilaketa aurreratua
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

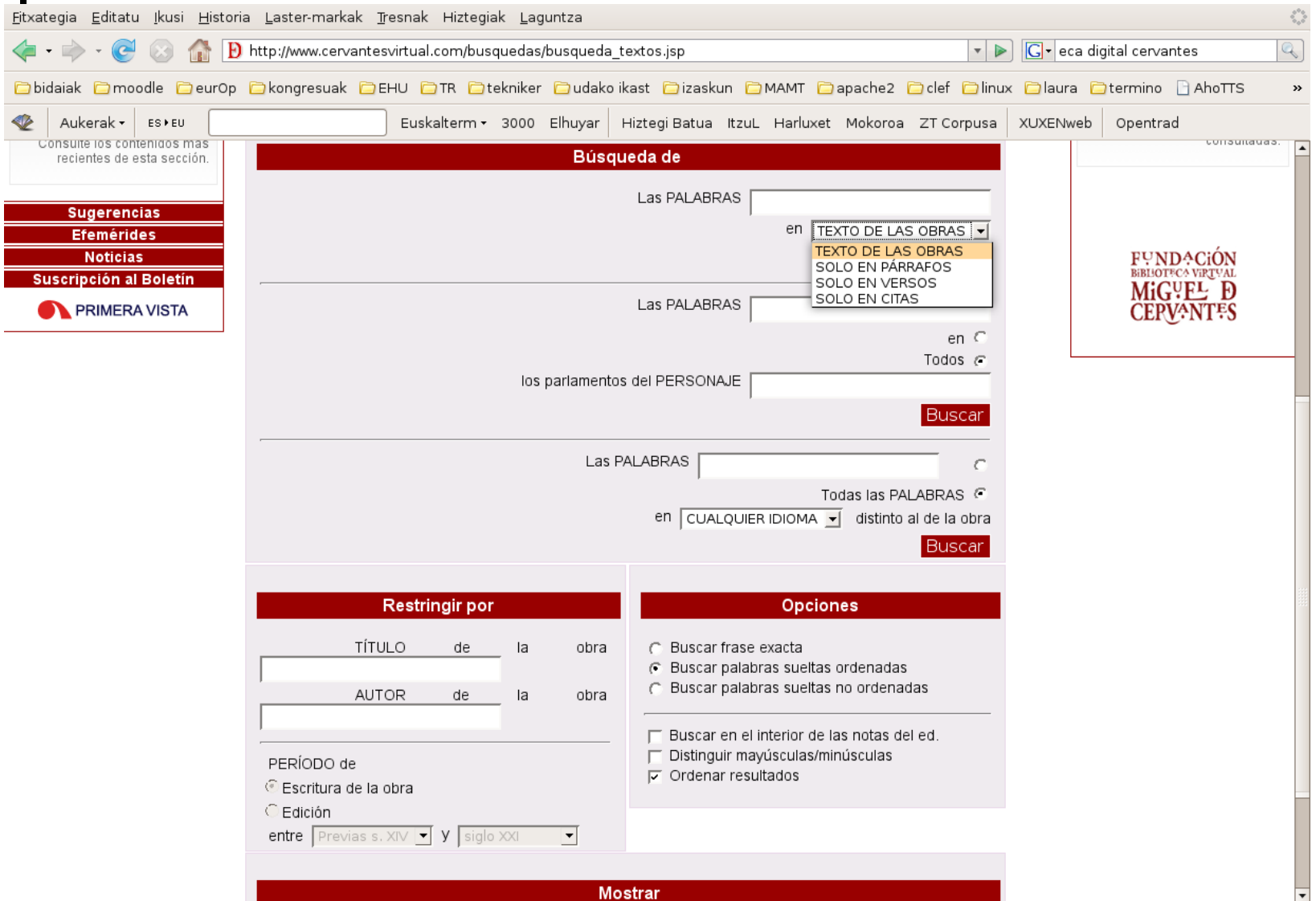
IR

- Bizitza *google* baino lehen
  - emaitza eskasagoak
  - abiadura motela
  - galdera zehatzagoak
  - ontologien erabilpena
- Egoera korapilatsuak
  - erantzunik ez/gutxi (estaldura handitu behar)
  - erantzun gehiegi (galdera findu behar)
    - *relevance-feedback*

Hemerotekak

- Informazio egituratuagoa eta laburragoa
- Metadatuak:
  - bilaketa-estrategia matadatuen arabera
- Informazio multimodala?
- Estandarrak informazio-trukerako
  - *Dublin Core, MARC* (liburutegi digitalak)
- IR ezaugarriak
  - galdera osoagoak?
    - bilaketa aurreratua
  - denbora gehiago?
  - estaldura oso garrantzitsua

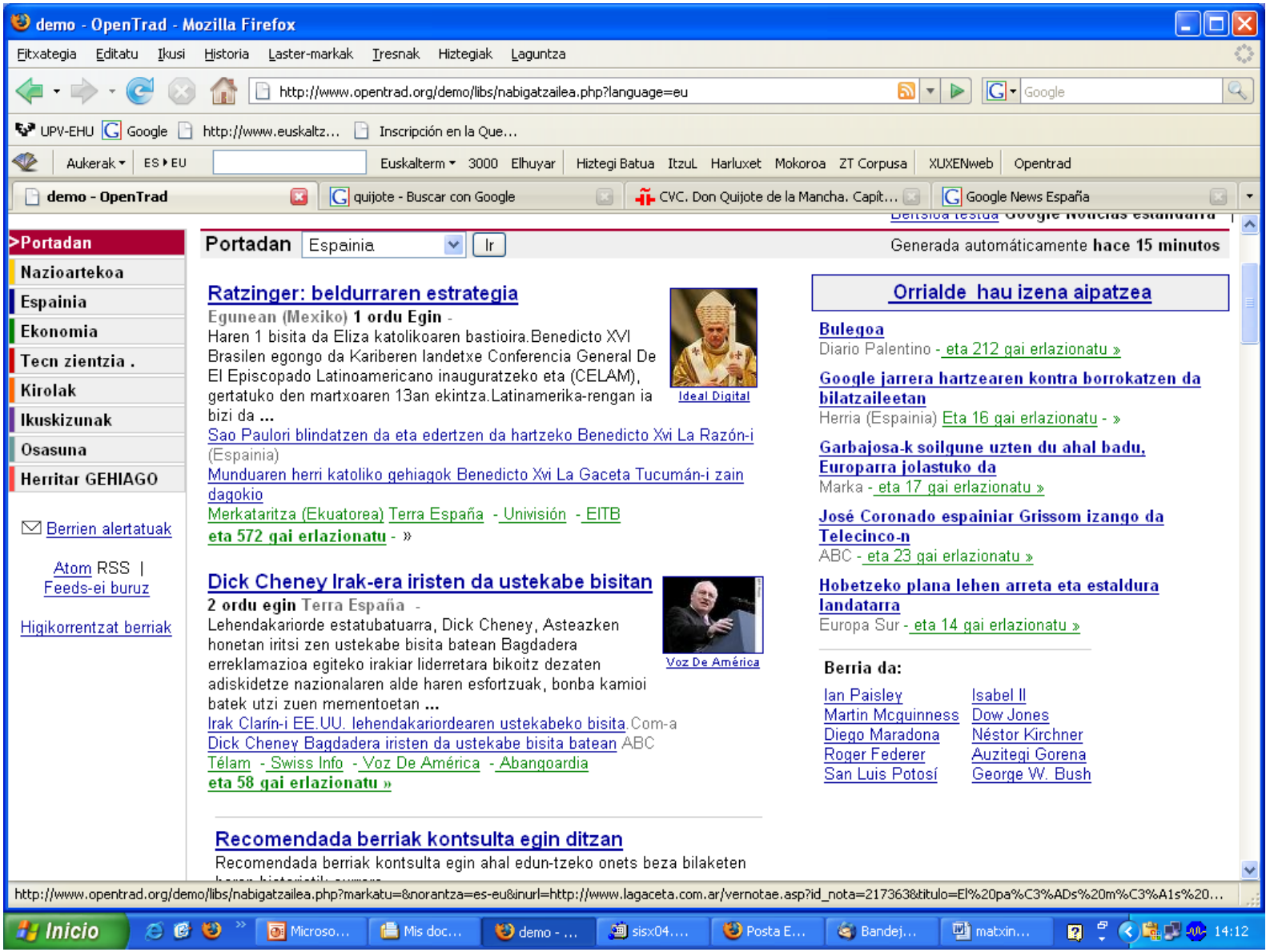
Hemeroteka


Fitxategia Editatu Ikusi Historia Laster-markak Tresnak Hiztegiak Laguntza
http://www.cervantesvirtual.com/busquedas/busqueda_textos.jsp
eca digital cervantes
bidaiak moodle eurOp kongresuak EHU TR tekniker udako ikast izaskun MAMT apache2 clef linux laura termino AhoTTS
Aukerak ES EU Euskalterm 3000 Elhuyar Hiztegi Batua ItzuL Harluxet Mokoroa ZT Corpusa XUXENweb Opentrad
Consulte los contenidos más recientes de esta sección.
Sugerencias
Efemérides
Noticias
Suscripción al Boletín
PRIMERA VISTA
Búsqueda de
Las PALABRAS
en TEXTO DE LAS OBRAS
TEXTO DE LAS OBRAS
SOLO EN PÁRRAFOS
SOLO EN VERSOS
SOLO EN CITAS
Las PALABRAS
en
Todos
los parlamentos del PERSONAJE
Buscar
Las PALABRAS
Todas las PALABRAS
en CUALQUIER IDIOMA distinto al de la obra
Buscar
Restringir por
TÍTULO de la obra
AUTOR de la obra
PERÍODO de
Escritura de la obra
Edición
entre Previas s. XIV y siglo XXI
Opciones
Buscar frase exacta
Buscar palabras sueltas ordenadas
Buscar palabras sueltas no ordenadas
Buscar en el interior de las notas del ed.
Distinguir mayúsculas/minúsculas
Ordenar resultados
Mostrar
FUNDACIÓN BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Tresna linguistikoak

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

Itzulpen automatikoa: *Matxin*

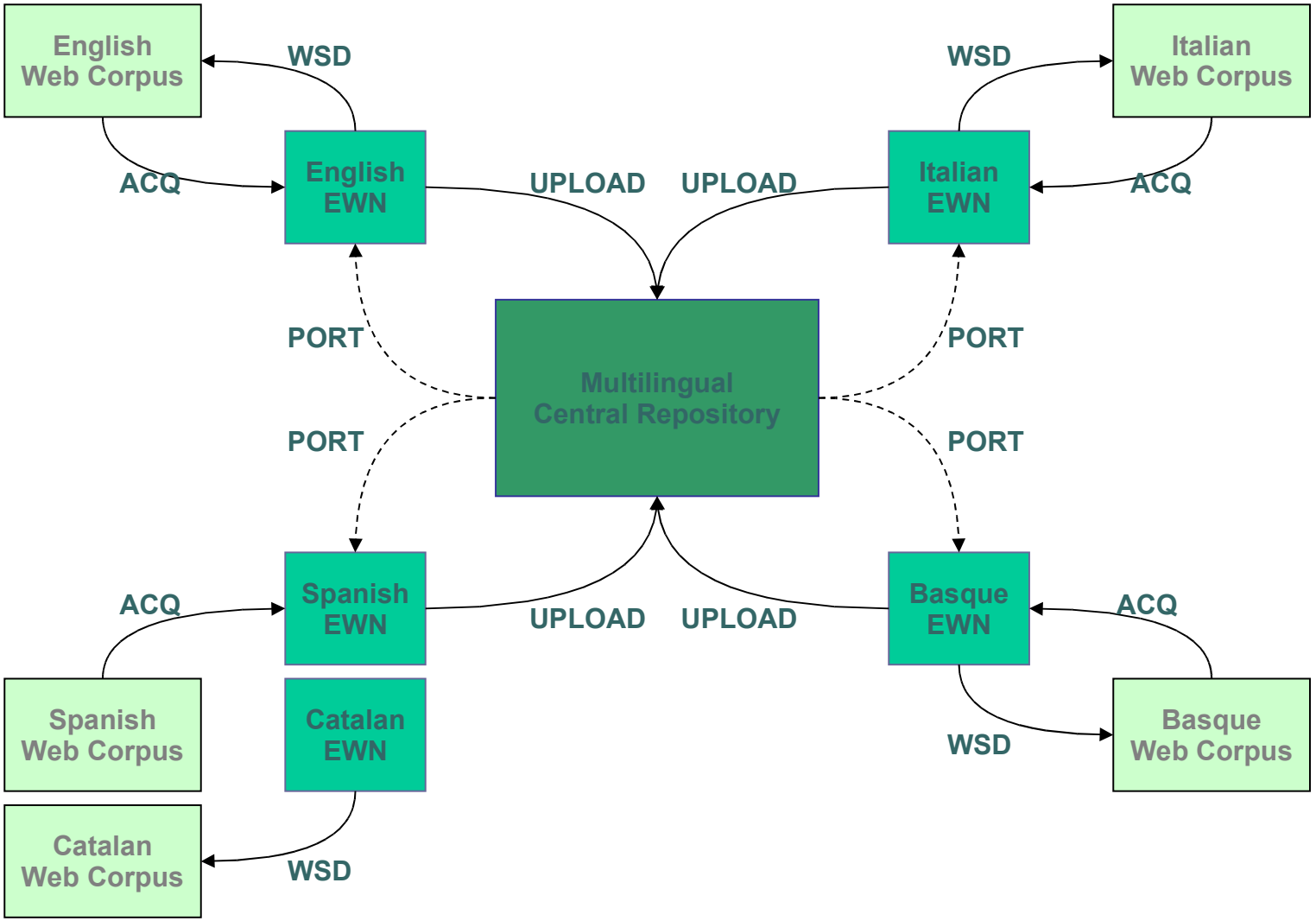


Semantika: MCR

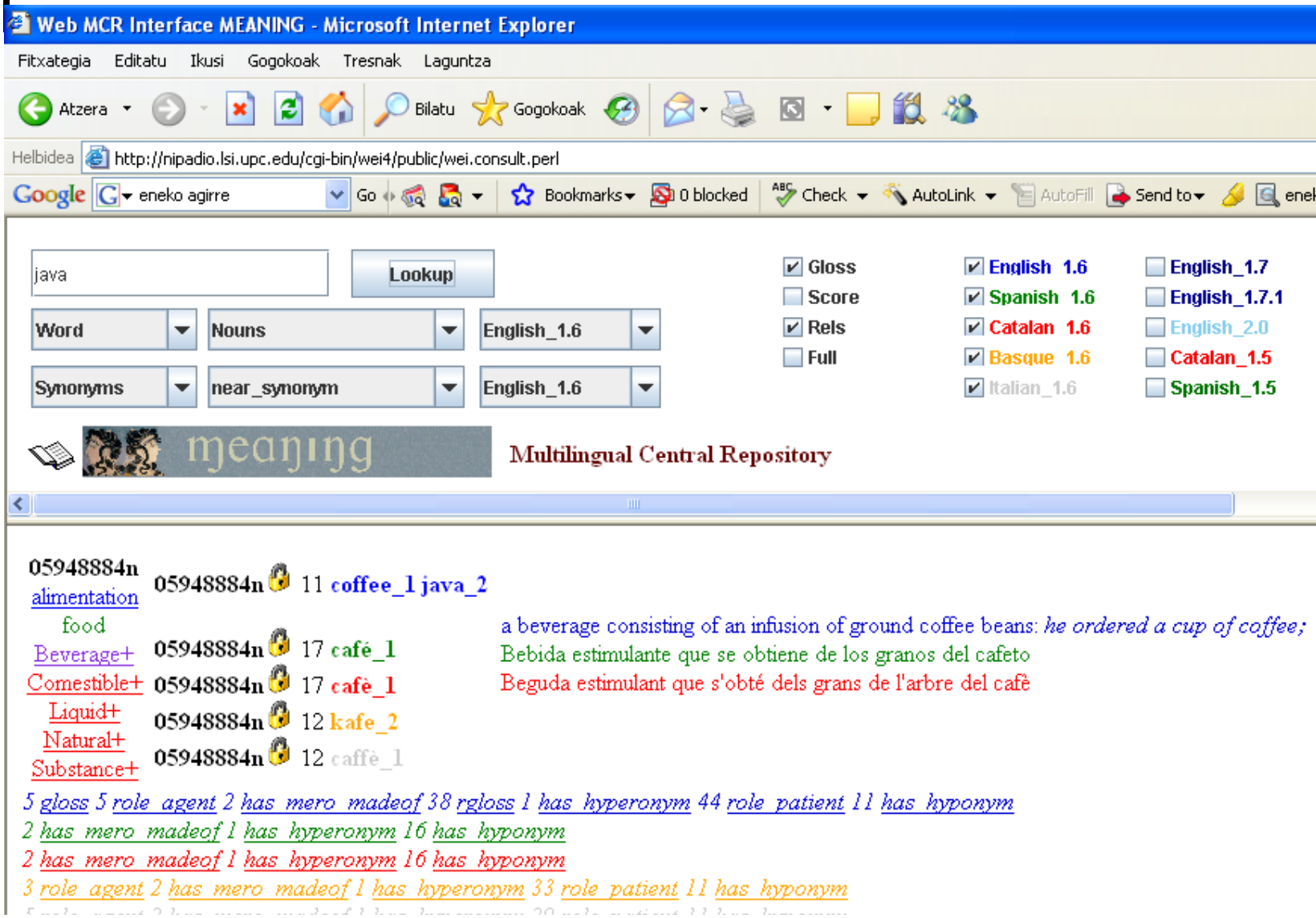
(Multilingual Central Repository)

- 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

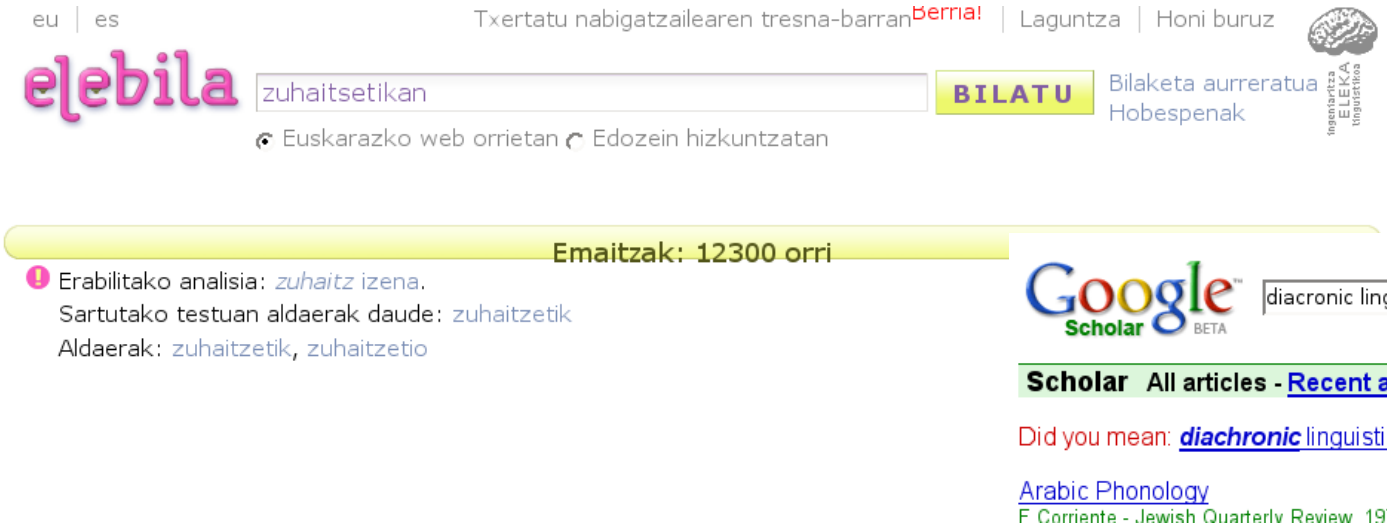

English Web Corpus
WSD
ACQ
English EWN
UPLOAD
UPLOAD
Italian EWN
WSD
Italian Web Corpus
ACQ
PORT
PORT
Multilingual Central Repository
PORT
PORT
Spanish EWN
ACQ
UPLOAD
UPLOAD
Basque EWN
ACQ
Spanish Web Corpus
Catalan EWN
WSD
Basque Web Corpus
Catalan Web Corpus
WSD

WordNet

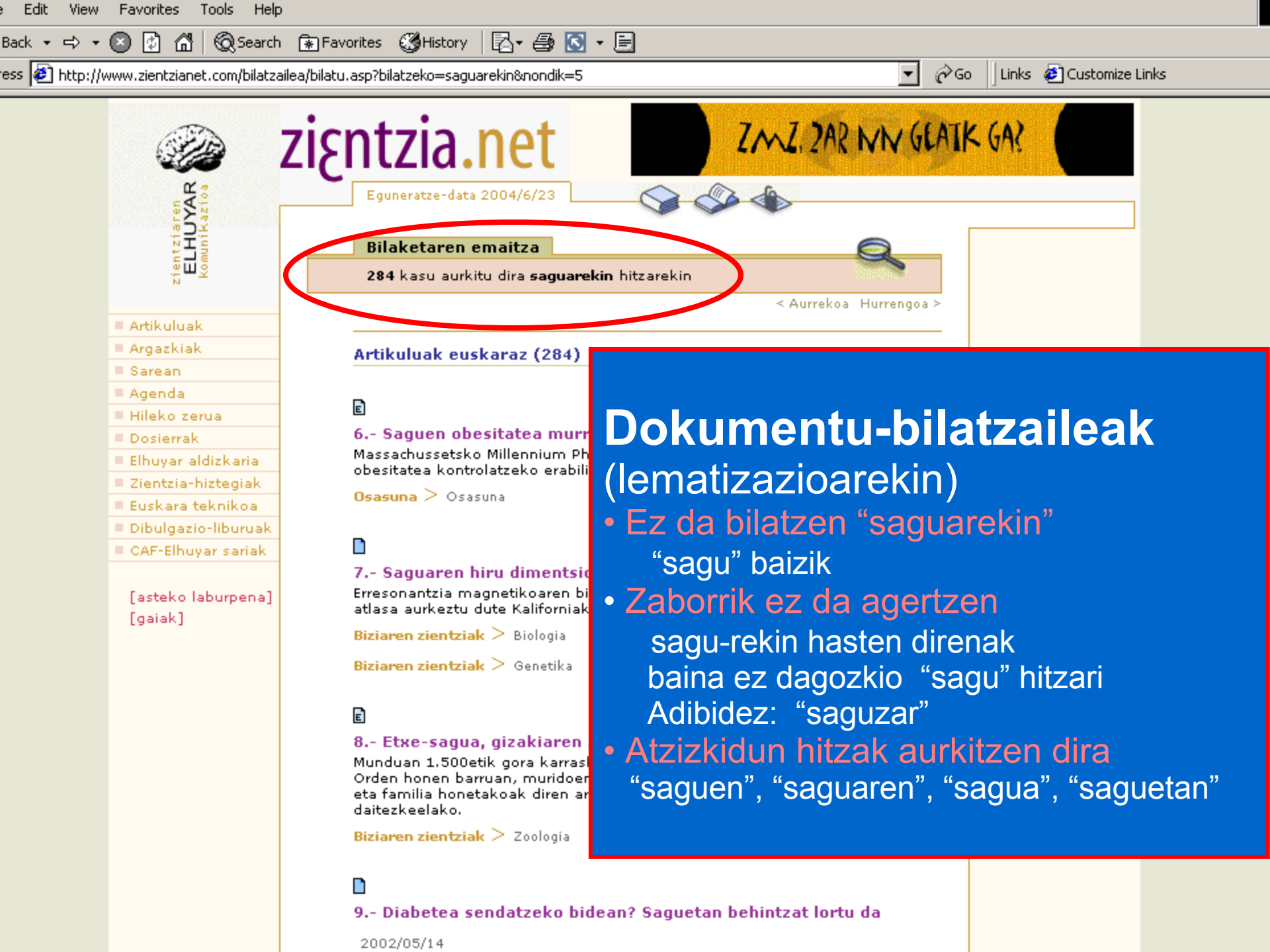

Web MCR Interface MEANING - Microsoft Internet Explorer
Fitxategia Editatu Ikusi Gogokoak Tresnak Laguntza
Atzera Bilatu Gogokoak
Helbidea http://nipadio.lsi.upc.edu/cgi-bin/wei4/public/wei.consult.perl
Google eneko agirre Go Bookmarks 0 blocked Check AutoLink AutoFill Send to
java
Lookup
Word
Nouns
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
alimentation
food
Beverage+
Comestible+
Liquid+
Natural+
Substance+
05948884n 11 coffee_1 java_2
a beverage consisting of an infusion of ground coffee beans: he ordered a cup of coffee;
05948884n 17 café_1
Bebida estimulante que se obtiene de los granos del cafeto
05948884n 17 cafè_1
Beguda estimulant que s'obté dels grans de l'arbre del cafè
05948884n 12 kafe_2
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

Morfologian oinarritutako **aplikazioak**

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




Edit View Favorites Tools Help
Back Search Favorites History
http://www.zientzianet.com/bilatzailea/bilatu.asp?bilatzeko=saguarekin&nondik=5
Go Links Customize Links
zientzia.net
Eguneratze-data 2004/6/23
Bilaketaren emaitza
284 kasu aurkitu dira saguarekin hitzarekin
< Aurrekoa Hurrengoa >
Artikuluak
Argazkiak
Sarean
Agenda
Hileko zerua
Dosierrak
Elhuyar aldizkaria
Zientzia-hiztegiak
Euskara teknikoa
Dibulgazio-liburuak
CAF-Elhuyar sariak
[asteko laburpena]
[gaiak]
Artikuluak euskaraz (284)
6.- Saguen obesitatea murriztu
Osasuna > Osasuna
7.- Saguaren hiru dimentsioko a
Biziaren zientziak > Biologia
Biziaren zientziak > Genetika
8.- Etxe-sagua, gizakiaren aprob
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 dagozkio “sagu” hitzari
Adibidez: “saguzar”
• Atzizkidun hitzak aurkitzen dira
“saguen”, “saguaren”, “sagua”, “saguetan”



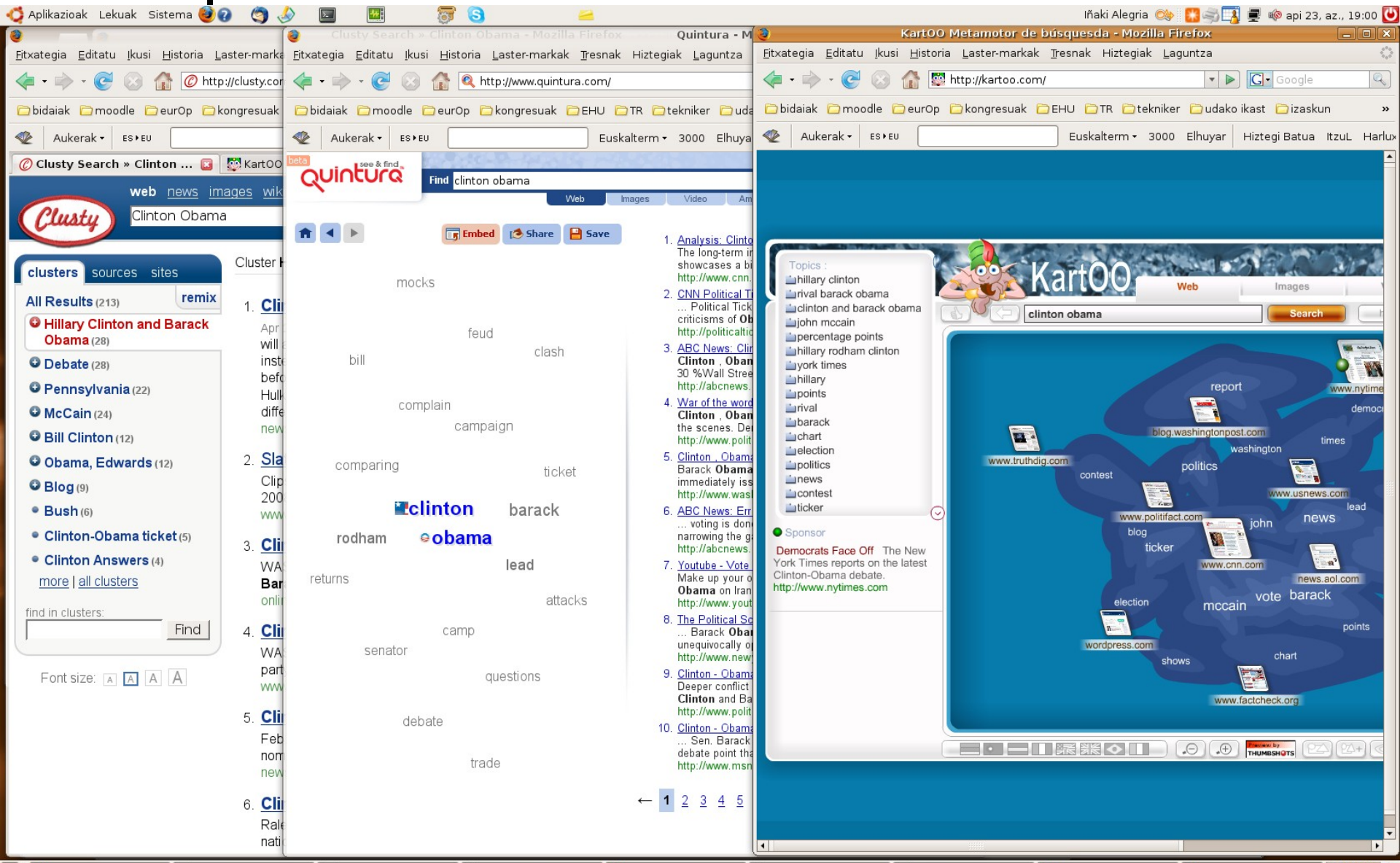
Dokumentu-bilatzaileak
(lematizazioarekin)

- Ez da bilatzen “saguarekin”
  “sagu” baizik
- Zaborrik ez da agertzen
  sagu-rekin hasten direnak
  baina ez dagozkio “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 (Eleka)

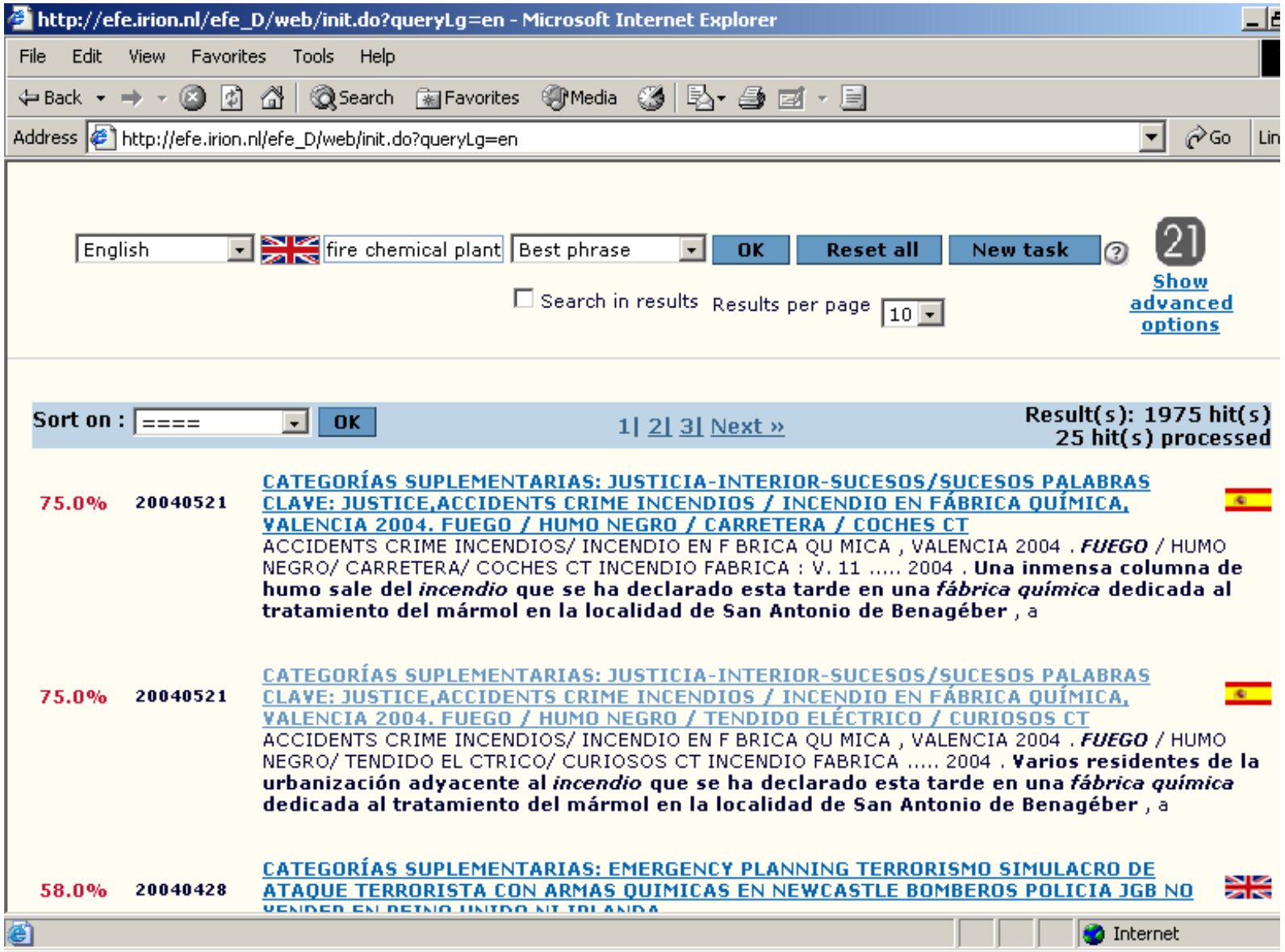
IR aurreratua



CLIR

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

CLIR (EFE)


http://efe.irion.nl/efe_D/web/init.do?queryLg=en - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites Media
Address http://efe.irion.nl/efe_D/web/init.do?queryLg=en
Go
English
fire chemical plant
Best phrase
OK
Reset all
New task
21
Search in results
Results per page 10
Show advanced options
Sort on : ====
OK
1| 2| 3| Next »
Result(s): 1975 hit(s)
25 hit(s) processed
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
Internet

CLIR (EFE)

News Article 10
CONTEXT = Sigue la violencia en Colombia y especialmente en Medellín.
GOAL = Un entierro en Medellín.

QUERY = **entierro medellín**
TEXT = **sepelio medellín**
RESULT = **FH_1205173 20040524**
RESULT = **FH_1205172 20040524**
<entierro #35, sepelio #14, enterramiento #7> = <burial, funeral>

CLIR: ebaluazioa

	Hitza	Hobetua
Actions	295	168
Pictures	20	24

MCR+desanbiguazio semantikoa:

- Ekintza gutxiago
- Argazki gehiago

SemEval-2007 / CLEF-2008 exercise

- Semantikaren eragina CLIR eta Q&A) arlotan

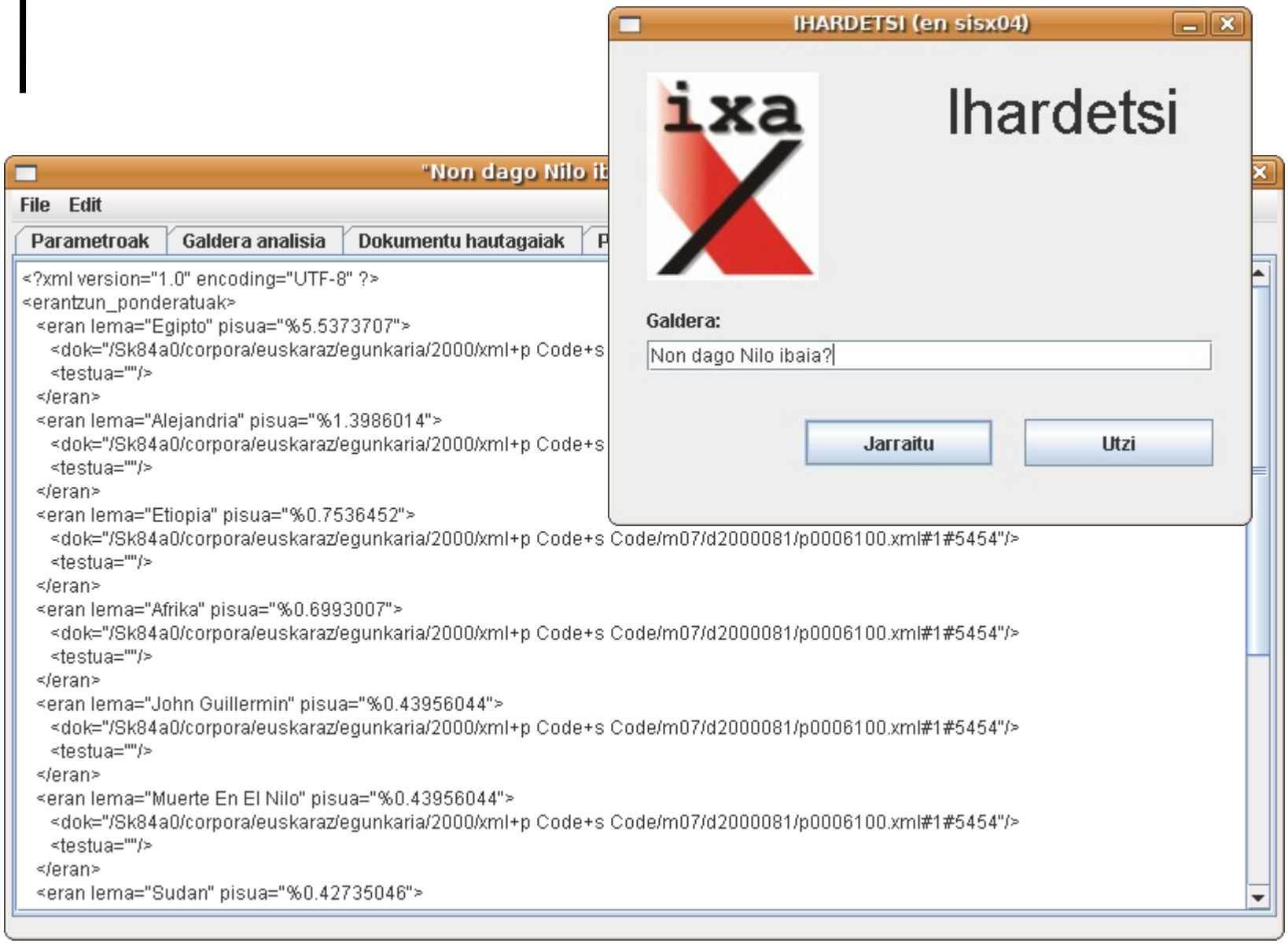
IR multimodala

- 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)

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

Ihardetsi

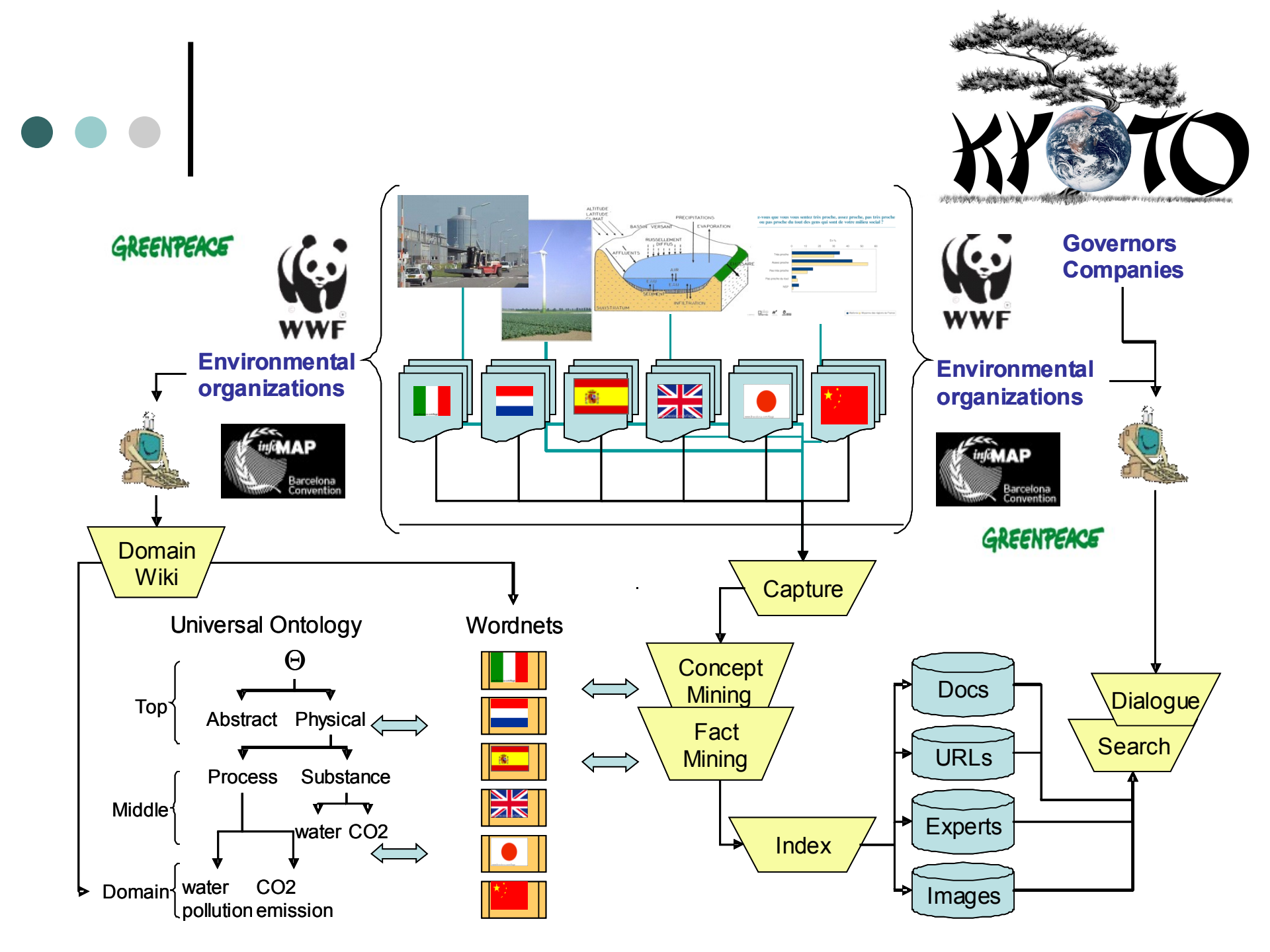

IHARDETSI (en sisx04)
ixa
Ihardetsi
Galdera:
Non dago Nilo ibaia?
Jarraitu
Utzi
"Non dago Nilo ib
File Edit
Parametroak
Galdera analisia
Dokumentu hautagaiak
<?xml version="1.0" encoding="UTF-8" ?>
<erantzun_ponderatuak>
<eran lema="Egipto" pisua="%5.5373707">
<dok="/Sk84a0/corpora/euskaraz/egunkaria/2000/xml+p 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">

IR semantikoa

- **Izenburua: Y**ielding **O**ntologies for **T**ransition-Based **O**rganization
- **Helburuak**:
  - Ezagutza partekatzea hizkuntzen eta kulturen artean
  - Bilaketa semantiko sakona eta testu-erauzketa 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 finantziatuta

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
GREENPEACE
WWF
Environmental organizations
infoMAP
Barcelona Convention
Governors
Companies
WWF
Environmental organizations
infoMAP
Barcelona Convention
GREENPEACE
Domain Wiki
Capture
Universal Ontology
Wordnets
Θ
Top
Abstract
Physical
Middle
Process
Substance
water
CO2
Domain
water pollution
CO2 emission
Concept Mining
Fact Mining
Index
Docs
URLs
Experts
Images
Dialogue
Search

Informazio-beharrak

- *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