

Mejorando los Buscadores usando Ingeniería Lingüística

IXA taldea



Ingeniería Lingüística y Búsqueda

- Identificación de idioma:
 - distinguir entre idiomas
- Lematización
 - identificar palabras de la misma raíz
- Análisis sintáctico automático
 - distinguir quién, qué, a quién, dónde, por qué?
- Análisis semántico automático
 - Hitz erlazionatuak identifikatzeko
- Itzulpen automatikoa
 - Hizkuntza desberdinetan aurkitzeko
- Hizketaren ezagutza/sintesia
 - Multimodala: hizketa bidez bilatzeko / emaitzak jasotzeko



Búsqueda(1): caracteres

AltaVista Búsquedas: Best books review

http://www.altavista.com/web/results?itag=ody&q=Best+books+review&kgs=0&kls=0

altavista Web Imágenes MP3/Audio Vídeo

Filtro familiar: **desactivar** Ayuda

Best books review **ENCONTRAR** Búsqueda avanzada Configuración

BÚSQUEDAS: ☒ En todo el mundo ☐ España RESULTADOS EN: ☒ Todos los Idiomas ☐ Inglés, Español

AltaVista ha encontrado 813.000.000 resultados

[Books - Sunday Book Review - New York Times](#)
Find book reviews & news from the Sunday Book **Review** on new **books** **best-seller** lists, fiction, non-fiction, literature, children's **books**, hardcover & paperbacks.
www.nytimes.com/pages/books/review/index.html • [Translate](#)
[Más páginas de nytimes.com](#)

[5 Best Books Review, Author of 5 Best Books, Book lovers read on ...](#)
click here to respond to this **review**. Add a comment to 5 **Best Books**. Comment: ... Read more **review** on 5 **Best Books** " Icons Help. Your reviews are read by the ...
www.mouthshut.com/review/5_Best_Books-53136-1.html • [Translate](#)
[Más páginas de mouthshut.com](#)

[5 Best Books Review, Author of 5 Best Books, My Best Books - MouthShut.com](#)
5 **Best Books**, My **Best Books**. ... click here to respond to this **review**. Add a comment to 5 **Best Books**. Comment: ... Read more **review** on 5 **Best Books** " Icons Help ...
www.mouthshut.com/review/5_Best_Books-12416-1.html • [Translate](#)
[Más páginas de mouthshut.com](#)

[Complete Review - The Best Books under Review](#)
The **Best Books** under **review** at the Complete **Review** ... The Top Rated **Books** at. the complete **review**. What are the **best books** under **review** ? ...
www.complete-review.com/best/toprate.html • [Translate](#)
[Más páginas de complete-review.com](#)

[The 10 Best Books of 2006 - New York Times](#)
The editors chose **books** by Gary Shteyngart, Amy Hempel, Claire ... This list will run in the Dec. 10 print edition of the Book **Review**. MORE **BEST BOOKS** LISTS ...
www.nytimes.com/ref/books/review/20061210tenbestbooks.html • [Translate](#)
[Más páginas de nytimes.com](#)

Erabiltzaileak bilatu nahi duena

Karaktere-kateen bilaketa hutsa denez, bilatzailea ez da gai "best-seller" termino bakarra dela ulertzeko

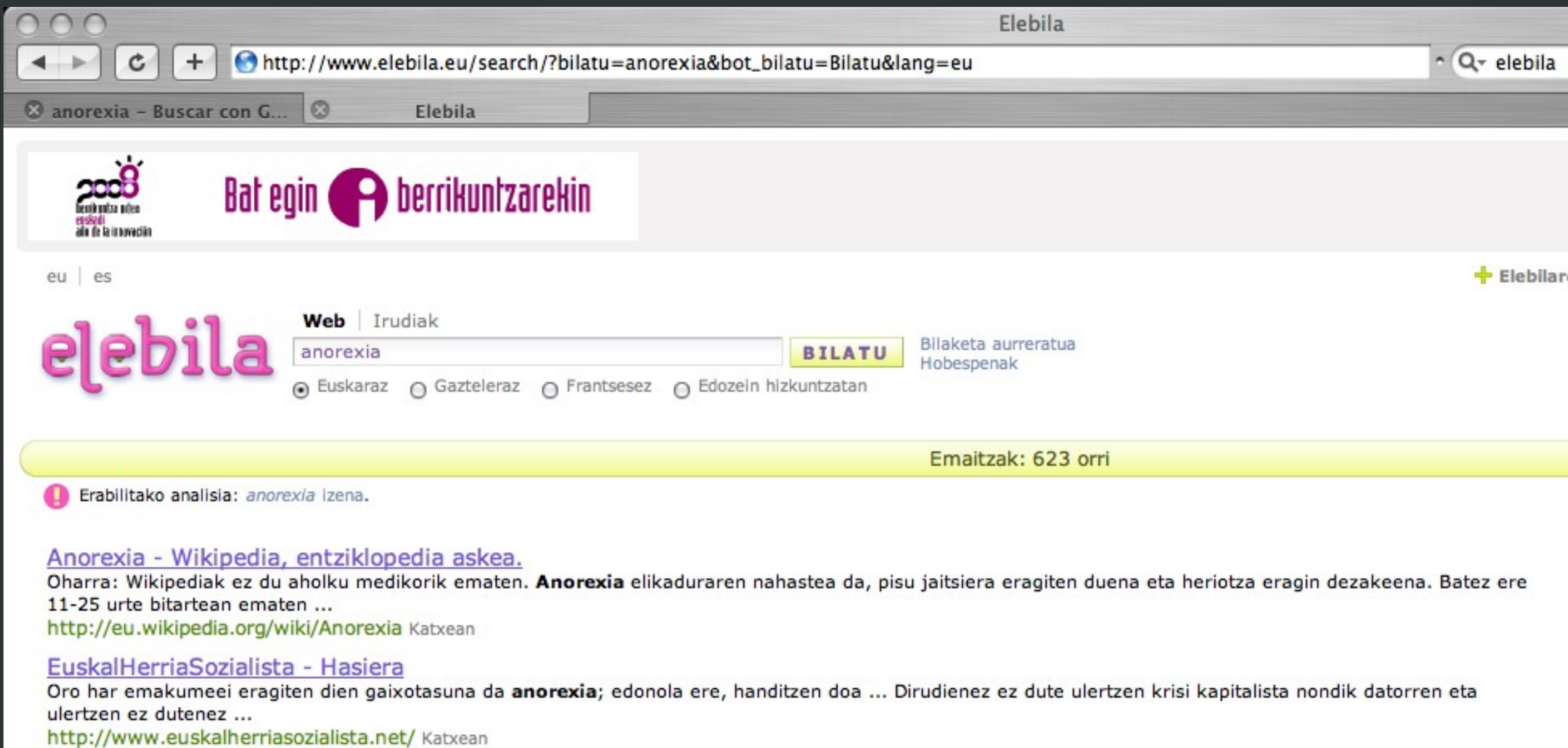
Erabiltzaileak bilatu nahi duena

Identificador de Idioma

- Basado en estadística
- Buenos resultados
- Problemas: idiomas similares y frases cortas



Búsqueda(2): Idioma



Lematización

- Lema a partir de palabra. P. ej.: *etxearen* => *etxe*
- Dos tecnologías lingüísticas:
 - finales de palabra (stemming)
 - morfología (prefijo + lema + sufijos)
- Buenos resultados pero problemas con la ambigüedad.
P. ej. como => *como* vs *comer*
- Resolución de la ambigüedad
 - métodos estadísticos, lingüísticos...



Búsqueda(3): Lematización

Elebilaren gehigarriak | Laguntza | Honi buruz

eu | es

elebila Web | Irudiak

uholdeen ondorioak **BILATU** Bilaketa aurreratua Hobespenak

☒ Euskaraz ☐ Gaztelera ☐ Frantsesez ☐ Edozein hizkuntzatan

Emaizak: 564 orri

! Erabilitako analisia: *uholde* izena, *ondorio* izena.
Aldaerak: *uhaldeen ondorioak*, *ujoldeen ondorioak*
Akaso horren sinonimoak interesatuko zaizkizu:

[Kyotoko Protokoloa negoziua egiteko beste mekanismo bat da»](#)
... **ondorioa** dela esan genezake? Bai, **ondorio** zuzena da. Eta konpontzeko ez ... ere aldatu behar dute. Komunitate txikiekiko justizia gehiago behar da. Herrialde pobreek gehiago sufritzen dute klima aldaketa, ez dutelako **uholde** ...
http://www.carbontradewatch.org/index.php?option=com_content&task=view&id=116&Itemid=36
Katxean

[Tamra Gilbertson - «Kyotoko Protokoloa negoziua egiteko beste ...](#)
... **ondorioa** dela esan genezake? Bai, **ondorio** zuzena da. Eta konpontzeko ez ... ere aldatu behar dute. Komunitate txikiekiko justizia gehiago behar da. Herrialde pobreek gehiago sufritzen dute klima aldaketa, ez dutelako **uholde** ...
http://www.tni.org/detail_page.phtml?act_id=18432&username=guest@tni.org&password=9999&publish=Y Katxean

[GeoNative - Ingria](#)
... hauei ere Ingriar esaten zaie batzuetan, eta nahasketak gerta daitezke **ondorioak**), 17. mendeko ... dira, errusiarren **uholdeak** ... Desagertzen direnean, ez da ...
<http://www.geocities.com/Athens/9479/ingria.html> Katxean

[3ARRANO.com - Mediateka - Trollak sarean](#)

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>

Análisis sintáctico automático

- De la frase al árbol sintáctico
- Tecnologías lingüística y estadística combinadas
- Buenos resultados en frases cortas
- Problemas: Coordinación y ambigüedad
 - P. ej.: *Asistieron los padres de Ander y Ane*



Búsqueda(4): Sintaxis

what books are written by shakespeare? - Powerset

http://www.powerset.com/explore/go/what-books-are-written-by-shakespeare%3F










Zer idatzi? Nork idatzi?

what books are written by shakespeare?

Sistemak erabiltzaileak galdetutakoa prozesatu ondoren eskainitako erantzuna

search

William Shakespeare: Works written Series Written (or Contributed To) Anthony Burgess source: freebase™ (view topic) ?

[A Midsummer Night's Dream](#) [Venus and Adonis](#) [Shakespeare's sonnets](#) [The Rape of Lucrece](#) [The Passionate Pilgrim](#) [The Phoenix and the Turtle](#) [A Lover's Complaint](#) [Hamlet](#) [Cymbeline](#)

Wikipedia Articles: results 1 - 10 of 7970 advanced ?

- ☒ [William Shakespeare](#) In some cases, for example Hamlet, Troilus and Cressida and Othello, Shakespeare could have revised texts between the quarto and folio editions. ... Plays thought to be only partly written by Shakespeare are marked with a dagger (†) below.
- ☒ [Shakespeare's plays](#) Some scholars argue that Shakespeare wrote less than 20% of the text. ... While it is generally accepted in academic circles that Shakespeare's plays were written by Shakespeare of Stratford and not another author, popular interest in the subject, particularly the

Traducción automática

- Tecnologías: lingüística y estadística
- Tecnología lingüística. P. ej.: *babelfish*, *opentrad*
 - Buenos resultados entre lenguas similares
 - Alto costo
 - Problemas: sintaxis, semántica.
- Tecnología estadística. P. ej.: *google*
 - Buenos resultados con textos muy grandes (ONU)
 - Problemas: idiomas minoritarios, de orden libre



Traducción automática

The screenshot displays a Mozilla Firefox browser window with the OpenTrad demo interface. The browser's address bar shows the URL `http://www.opentrad.org/demo/libs/nabigatzailea.php?language=eu`. The OpenTrad interface includes a sidebar with navigation links such as "Portadan", "Nazioartekoa", "Espainia", "Ekonomia", "Tecn zientzia", "Kirolak", "Ikuskizunak", "Osasuna", and "Herriar GEHIAGO". The main content area features news articles in Basque, including "Ratzinger: beldurraren estrategia" and "Dick Cheney Irak-era iristen da ustekabe bisitan".

Below the OpenTrad interface, a SourceForge search results page is visible. The search query is "matxin apertium". The results table shows two projects:

Name	Relevance	Activity	Rank	Registered	Latest File	Downloads
Apertium: machine translation toolbox	<div></div>	99.75%	584	2005-07-14	2008-09-29	36,670
Apertium is a toolbox to build open-source shallow-transfer machine translation systems, especially suitable for related language pairs: it includes the engine, maintenance tools, and open linguistic data for several language pairs. Members (75) Download						
Topic: Other/Nonlisted Topic						
matxin	<div></div>	94.66%	12,445	2006-01-24	2007-07-16	652
Machine translation engine based on a dependency grammar and XML interchange format. The Spanish-Basque (es-eu) translation is ready. Members (3) Download						
Topic: Artificial Intelligence , Text Processing						

The browser's status bar at the bottom shows the current page URL and navigation controls.

Búsqueda(5): Traducción automática

concert of Oskorri - Google Translate

http://translate.google.com/translate_s?hl=en&clss=&q=concert+of+Oskorri&sl=en&tl=es

Google Translate BETA

Text and Web Translated Search Dictionary Tools

Translated Search

Search for: concert of Oskorri Translated to: concierto de Oskorri Not quite right? Edit

My language: English Search pages written in: Spanish

Translate and Search

Jatorrizko hizkuntza Helburuko hizkuntza

Translated results from Spanish web pages

Results 1 - 10 of about 22,200 for

English translation

Upcoming concerts **Oskorri**, discography, photos and videos of...
Upcoming concerts **Oskorri**, discography, photos and videos of **Oskorri**. Receive alerts of concerts, festivals and tours all the artists in your city.
www.nvivo.es/artistas/n/Oskorri/ - 57k - [Cached](#)

Oskorri next concert in Bilbao Aste Nagusia...
Oskorri next concert in Bilbao Aste Nagusia - Abandoibarra, Bilbao (Vizcaya) Tuesday, August 21, 2007, purchase tickets and learn more about concerts...
www.nvivo.es/conciertos/i/2040/ - 38k - [Cached](#)

Original Spanish - [Hide Spanish results](#)

Próximos conciertos de Oskorri, discografía, fotos y videos de...
Próximos conciertos de Oskorri, discografía, fotos y videos de Oskorri. Recibe alertas de conciertos, festivales y giras de todos los artistas en tu ciudad.
www.nvivo.es/artistas/n/Oskorri/ - 57k - [En caché](#)

Próximo concierto de Oskorri en Bilboko Aste Nagusia ...
Próximo concierto de Oskorri en Bilboko Aste Nagusia - Abandoibarra, Bilbao agosto 2007, compra entradas y conoce más sobre conciertos ...
www.nvivo.es/conciertos/i/2040/ - 38k - [En caché](#)

(1) Sistemak jatorrizko hizkuntzan dagoen galdera automatikoki utzuliko du

(2) Sistemak dagokion hizkuntzan dauden testuetan bilatuko du

(3) Lortutako emaitzak gure jatorrizko hizkuntzara itzuliko ditu

Análisis semántico automático

- Relaciones semánticas: sinónimos, hiperónimos, hipónimos... P. ej.:

ave > pájaro > gorrión juez ~ magistrado

- Bases de conocimiento, Ontologías: *WordNet*
- Buenos resultados con pocos documentos interesantes
- Problemas: necesidad de grandes bases de conocimiento / ambigüedad. Adibidez:

caja de ahorro, banco vs banco, asiento



Búsqueda(6): Semántica

Fitxategia Editatu Ikusi Historia Laster-markak Tresnak Hiztegiak Laguntza

http://www.hakia.com/search.aspx?lang=web&q=hillary+clinton uholdeen ondorio:

BETA hakia® Enter a Ask

hillary clinton

All results Credible sites News Images Meet Others Galleries my hakia

Below is the hakia resume for Hillary Clinton. Ask a more detailed question about Hillary Clinton.

The Official Website

[Hillary Rodham Clinton](#), Senator from New York
Official US Senate web site of Senator Hillary Rodham Clinton (D - NY).
<http://clinton.senate.gov/>

[HillaryClinton.com - Welcome](#)
Official Site of Hillary Clinton for President Exploratory Committee.
<http://www.hillaryclinton.com/>

Image Search more images



clinton.senate.gov

On This Page

- [The Official Website](#)
- [Image Search](#)
- [Biography and Timeline](#)
- [Awards and Accomplishments](#)
- [Speeches and Quotes](#)
- [Photographs and Pictures](#)
- [Blogs and Fan Sites](#)
- [News and Interviews](#)
- [Myths and Controversies](#)
- [Criticism, Commentary, and Interpretation](#)
- [Resources, Archives, and Organizations](#)
- [Contact Information](#)
- [Bibliography](#)
- [Filmography](#)

Biography and Timeline

[Biography of Hillary Clinton](#)
As First Lady, her public involvement with many activities sometimes led to controversy. Undeterred by critics, Hillary won many admirers for her staunch support for women around the world and her commitment to children's issues. She was elected United States Senator from New York on November 7, 2000. She is the first First Lady elected to the United States
<http://www.whitehouse.gov/history/firstladies/hc42.html>

[National Women's Hall of Fame - Women of the Hall](#)
Hillary Rodham Clinton, the first First Lady to be elected to the United States [United States Senate](#) and the first woman U.S. Senator from New York State, was raised in Park Ridge, Illinois.

hakia

http://www.hakia.com/search.aspx?lang=web&q=John+Lennon

BETA hakia® Enter a Ask

John Lennon

All results Credible sites News Images Meet Others Galleries my hakia


Below is the hakia resume for John Lennon. Ask a more detailed question about John Lennon. You can also search for: [Yoko Ono](#), [Eric Clapton](#), [The Beatles](#), [Elton John](#), [Paul McCartney](#)

Erlazionatutako galderak proposatzen ditu

Music Profile

[John Lennon](#)
THE OFFICIAL SITE
<http://www.johnlennon.com/>

Image Search more images



www.johnlennon.com

On This Page

- [Music Profile](#)
- [Image Search](#)
- [Headline News](#)
- [Biography and Timeline](#)
- [Famous Songs](#)
- [Discography/Musical Works](#)
- [Movies and Videos](#)
- [Lyrics](#)
- [Awards and Accomplishments](#)
- [Photographs and Pictures](#)
- [Blogs and Fan Sites](#)
- [News and Interviews](#)
- [Speeches and Quotes](#)
- [Myths and Controversies](#)

Headline News more news

[John Lennon: The Life](#)
... off of them have been overharvested to the point of extinction. So why bother inviting Philip Norman's new 851-page biography of [John Lennon into your life](#)? Even if the book is good, you'll just get swept up in how funny, brilliant, troubled, insecure, caustic, and loving Lennon was and...

Entertainment Weekly Online - 10 days ago

Biography and Timeline

Interneten bilatutako dokumentuak erlazionatu eta multzokatzen ditu

Búsqueda(7): Semántica

http://efe.irion.nl/efe_D/web/init.do?queryLg=en - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Media

Address http://efe.irion.nl/efe_D/web/init.do?queryLg=en

English fire chemical plant Best phrase OK

☐ Search in results Results per

Sort on : ==== OK 1| 2| 3| Next >>

75.0% 20040521

CATEGORÍAS SUPLEMENTARIAS: JUSTICIA-INTERIOR
CLAVE: JUSTICE,ACCIDENTS CRIME INCENDIOS / INC
VALENCIA 2004. FUEGO / HUMO NEGRO / CARRETERA

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 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 QUÍMICAS EN NEWCASTLE BOMBEROS POLICIA JGB NO
VENIDER EN DETINO UNIDO NI IRLANDA

05445176n

-factotum-
event

RadiatingLight+
BoundedEvent=

05445176n 2 **conflagration_1 inferno_3**

05445176n 0 **conflagración_1 incendio_1**

05445176n nolex 1

a very intense and uncontrolled fire

05444579n mn

99

-factotum-
event

RadiatingLight+
BoundedEvent=

05444579n 13 **fire_1**

05444579n 11 **fuego_4**

05444579n 7 **sute_1**
sutzar_3 su_11

the event of something burning (often destruc

ustekabeen edo eraginda piztutako su handia,
aurkako sistema automatikoa eta berau antzer

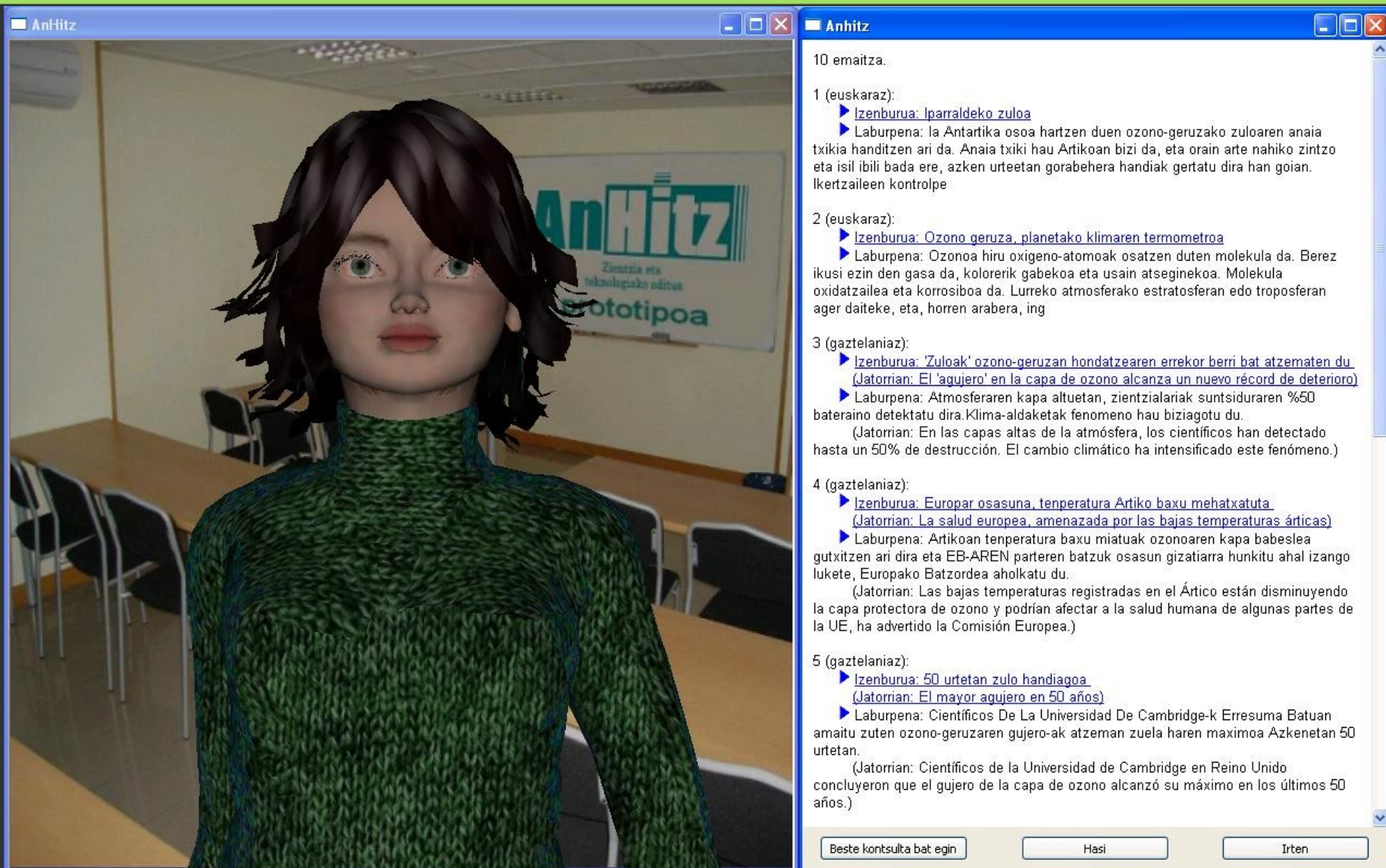


Reconocimiento del habla

- Basada en bases de datos pregrabados y etiquetados
- Buenos resultados con locutores limitados y diccionario restringido
- Problemas:
 - variantes, interrupciones, ruidos...



Búsqueda(8): multimodal



AnHitz

10 emaitza.

1 (euskaraz):

- ▶ izenburua: Iparaldeko zuloa
- ▶ Laburpena: la Antartika osoa hartzen duen ozono-geruzako zuloaren anaia txikia handitzen ari da. Anaia txiki hau Artikoan bizi da, eta orain arte nahiko zintzo eta isil ibili bada ere, azken urteetan gorabehera handiak gertatu dira han goian. Ikertzaileen kontrolpe

2 (euskaraz):

- ▶ izenburua: Ozono geruza, planetako klimaren termometroa
- ▶ Laburpena: Ozonoa hiru oxigeno-atomoak osatzen duten molekula da. Berez ikusi ezin den gasa da, kolorerik gabekoa eta usain atseginekkoa. Molekula oxidatzailea eta korrosiboa da. Lurreko atmosferako estratosferan edo troposferan ager daiteke, eta, horren arabera, ing

3 (gaztelaniaz):

- ▶ izenburua: 'Zuloak' ozono-geruzan hondatzearen errekor berri bat atzematen du
(Jatorrian: El 'agujero' en la capa de ozono alcanza un nuevo récord de deterioro)
- ▶ Laburpena: Atmosferaren kapa altuetan, zientzialariak suntsiduraren %50 bateraino detektatu dira. Klima-aldaketak fenomeno hau biziagotu du.
(Jatorrian: En las capas altas de la atmósfera, los científicos han detectado hasta un 50% de destrucción. El cambio climático ha intensificado este fenómeno.)

4 (gaztelaniaz):

- ▶ izenburua: Europar osasuna, temperatura Artiko baxu mehatxatuta
(Jatorrian: La salud europea, amenazada por las bajas temperaturas árticas)
- ▶ Laburpena: Artikoan temperatura baxu miatuak ozonoaren kapa babeslea gutxitzen ari dira eta EB-AREN parteren batzuk osasun gizatiarra hunkitu ahal izango lukete, Europako Batzordea aholkatu du.
(Jatorrian: Las bajas temperaturas registradas en el Ártico están disminuyendo la capa protectora de ozono y podrían afectar a la salud humana de algunas partes de la UE, ha advertido la Comisión Europea.)

5 (gaztelaniaz):

- ▶ izenburua: 50 urtetan zulo handiagoa
(Jatorrian: El mayor agujero en 50 años)
- ▶ Laburpena: Científicos De La Universidad De Cambridge-k Erresuma Batuan amaitu zuten ozono-geruzaren gujero-ak atzeman zuela haren maximoa Azkenetan 50 urtetan.
(Jatorrian: Científicos de la Universidad de Cambridge en Reino Unido concluyeron que el gujero de la capa de ozono alcanzó su máximo en los últimos 50 años.)

Beste kontsulta bat egin Hasi Irten

Pruebas

- <https://ixa.si.ehu.es/Ixa/Demoak>
- www.elebila.eu
- www.opentrad.com
- www.google.com/language_tools (adib. *concert of Oskorri*)
- www.powerset.com (adib. *what books are written by Shakespeare*)
- www.hakia.com (adib. *Hillary Clinton*)
- <http://demo.irion.nl/kyoto/> (adib. *residuos urbanos*; English %50)

