“OpenMT2 eta Euskal Wikipedia” wikiproiektuaren emaitzak

OmegaT itzulpen-tresna hobetuta euskaraz eta Wikipediarekin aritu ahal izateko.

Euskal Herriko Unibertsitatea
IXA TALDEA
HIZKUNTZAREN PROZESAMENDUA

Iñaki Alegria, Unai Cabezón, Gorka Labaka, Aingeru Mayor, Kepa Sarasola

Ixa Taldea https://ixa.si.ehu.es

Unai Fernandez de Betoño, Galder Gonzalez, Mikel Iturbe, Arkaitz Zubiaga

Euskal Wikipedia http://eu.wikipedia.org

Helburua

"OpenMT2 eta Euskal Wikipedia" wikiproiektua

http://eu.wikipedia.org/wiki/Wikiproiektu:OpenMT2_eta_Euskal_Wikipedia

Hiru elementu hauek integratzea:

- **OmegaT** itzulpenak egiteko ingurunea
- **Matxin**: Itzulpen automatikoa
- **Euskal Wikipedia**

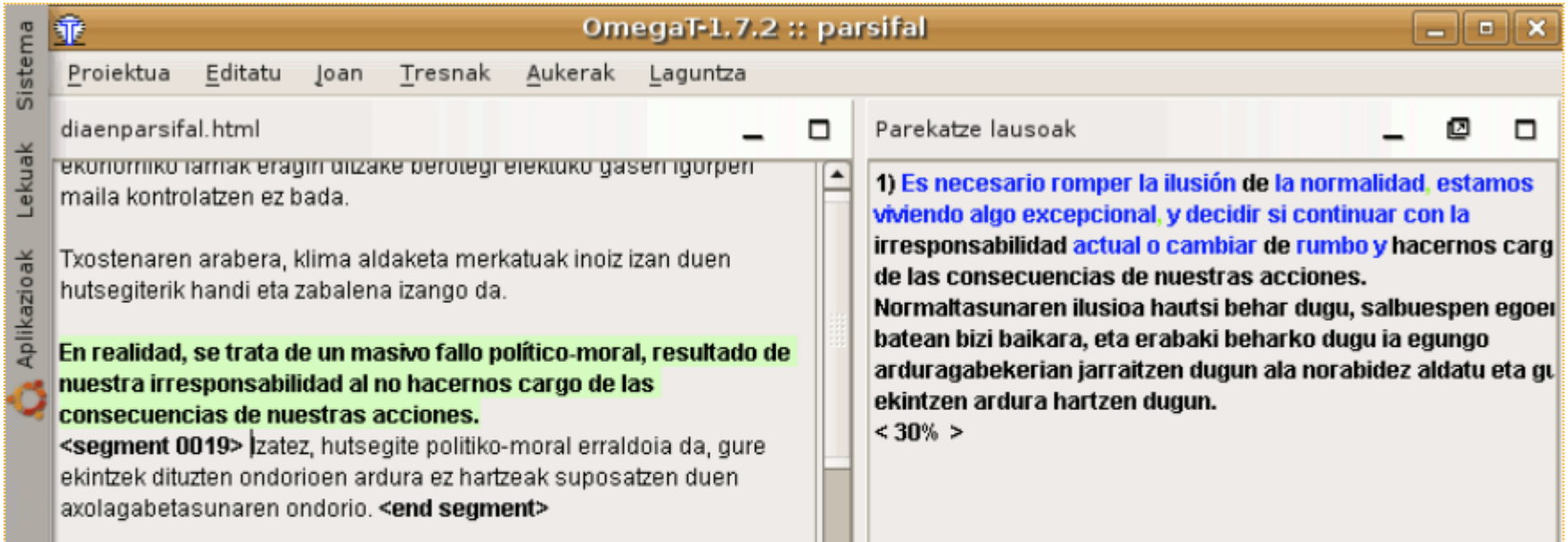
Bide batez, hiru elementuotan **hobekuntzak** egin nahi genituen.

Hiru hankak

- **OmegaT** itzulpenak egiteko ingurunea
- **Matxin**: Itzulpen automatikoa
- **Euskal Wikipedia**

OmegaT

- **OmegaT** Ordenagailuz Lagundutako Itzulpen-sistema plataforma anitza da, kode irekikoa, parekatze lausoz, itzulpen-memoriaz, hitz gakoen bilaketaz eta glosarioz hornitutakoa, eta egindako itzulpenak proiektu eguneratuetan aprobetxatzea ahalbidetzen duena.


Sistema
Lekuak
Aplikazioak
OmegaT-1.7.2 :: parsifal
Proiektua Editatu Joan Tresnak Aukerak Laguntza
diaenparsifal.html
maila kontrolatzen ez bada.
Txostenaren arabera, klima aldaketa merkatuak inoiz izan duen hutsegiterik handi eta zabalena izango da.
En realidad, se trata de un masivo fallo político-moral, resultado de nuestra irresponsabilidad al no hacernos cargo de las consecuencias de nuestras acciones.
<segment 0019> Izatez, hutsegite politiko-moral erraldoia da, gure ekintzek dituzten ondorioen ardura ez hartzeak suposatzen duen axolagabetasunaren ondorio. <end segment>
Parekatze lausoak
1) Es necesario romper la ilusión de la normalidad, estamos viviendo algo excepcional, y decidir si continuar con la irresponsabilidad actual o cambiar de rumbo y hacernos carg
de las consecuencias de nuestras acciones.
Normaltasunaren ilusioa hautsi behar dugu, salbuespen egoe
batean bizi baikara, eta erabaki beharko dugu ia egungo
arduragabekerian jarraitzen dugun ala norabidez aldatu eta gu
ekintzen ardura hartzen dugun.
< 30% >

Matxin

opentrad-matxin

espainiera-euskara itzulpen librea

Opentrad

- Hiru sistema automatiko erabil daitezke espainieratik euskarara itzultzeko:
  - Matxin
    http://www.opentrad.com/
  - Itzultzailea (Eusko Jaurlaritza)
    http://www.itzultzailea.euskadi.net/traductor/portalExterno/text.do
  - Google Translate
    http://translate.google.com/#es/eu/
- Matxin
  - Lehenengoa (2006tik erabilgarri)
  - Software librea

Wikipedia

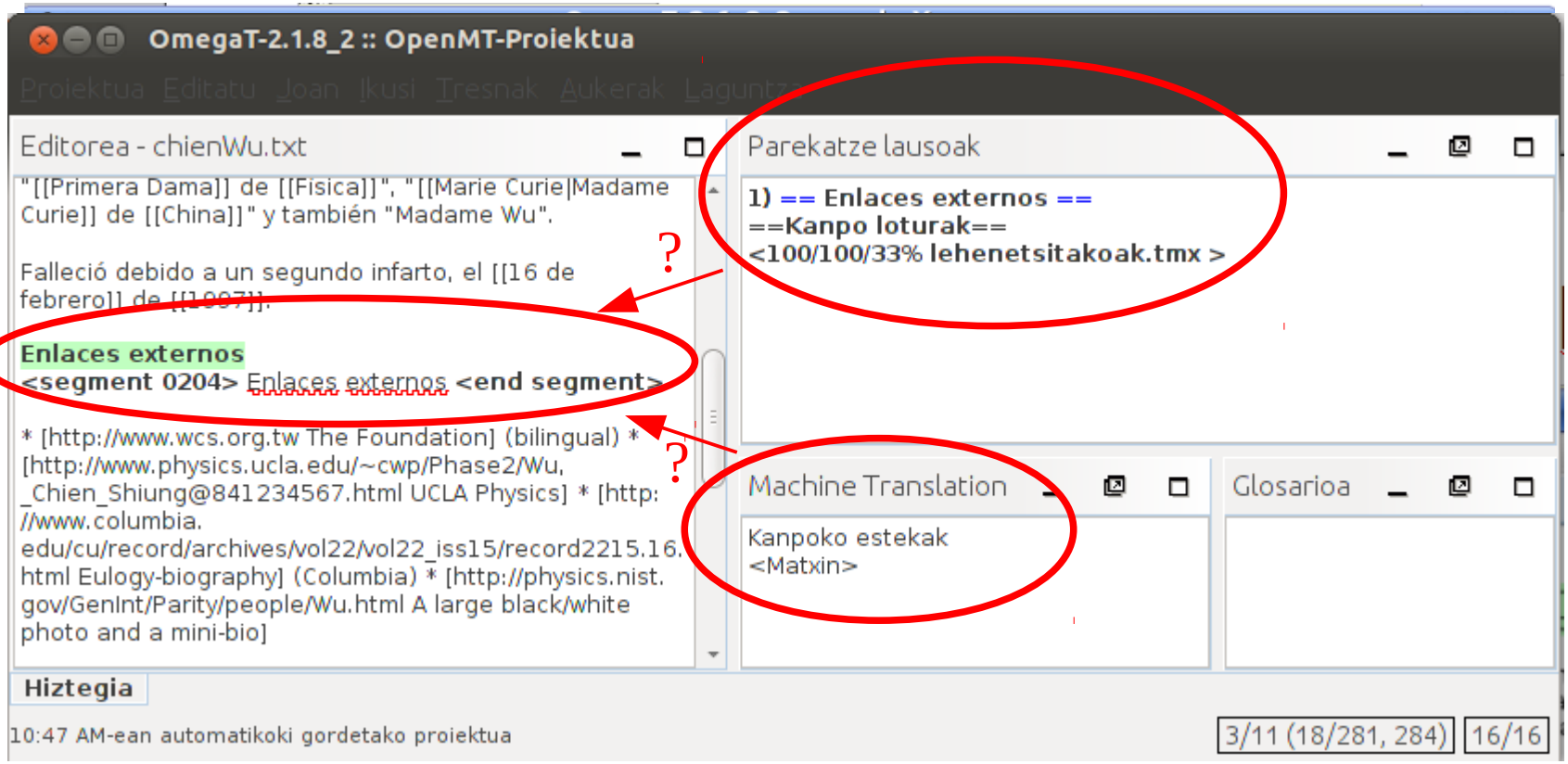
- Eduki askeko **entziklopedia**, lankidetzaz editatua, eleanitza, Interneten argitaratua
- Barne- eta kanpo-**esteka ugari** ditu.
- Programa informatikoek **“hodeian” erabil dezakete**
- Guztira 23 milioi artikulu, **hainbat hizkuntzatan**:
  - Ingelesezkoa: 4.000.000 artikulu
  - Espaineraz: 1.000.000
  - Katalanez: 400.000
  - **Euskaraz: 150.000** (35. hizkuntza; txikia HPko aplikazioetarako)
  - Galegoz: 100.000
  - ...

Hobekuntzak

- **OmegaT**-n
- **Matxin**-en
- **Euskal Wikipedia**-n

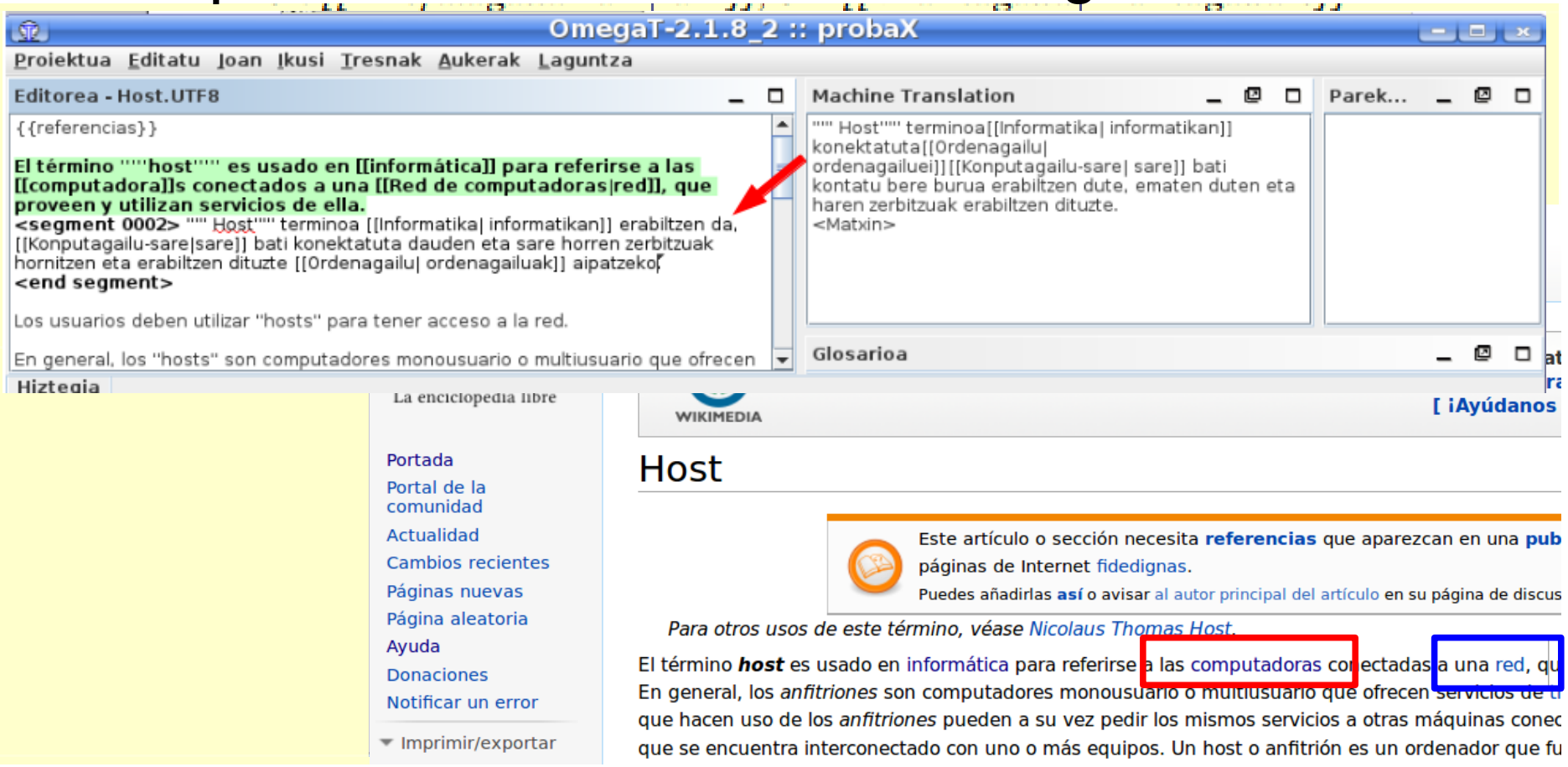
Hobekuntzak OmegaT-n (I)

- Matxin itzultzailea eta Xuxen erabili ahal izatea.


OmegaT-2.1.8_2 :: OpenMT-Proiektua
Proiektua Editatu Joan Ikusi Tresnak Aukerak Laguntza
Editorea - chienWu.txt
"[[Primera Dama]] de [[Fisica]]", "[[Marie Curie|Madame Curie]] de [[China]]" y también "Madame Wu".
Falleció debido a un segundo infarto, el [[16 de febrero]] de [[1997]].
Enlaces externos
<segment 0204> Enlaces externos <end segment>
* [http://www.wcs.org.tw The Foundation] (bilingual) * [http://www.physics.ucla.edu/~cwp/Phase2/Wu,_Chien_Shiung@841234567.html UCLA Physics] * [http://www.columbia.edu/cu/record/archives/vol22/vol22_iss15/record2215.16.html Eulogy-biography] (Columbia) * [http://physics.nist.gov/GenInt/Parity/people/Wu.html A large black/white photo and a mini-bio]
Parekatze lausoak
1) == Enlaces externos ==
==Kanpo loturak==
<100/100/33% lehenetsitakoak.tmx >
Machine Translation
Kanpoko estekak
<Matxin>
Glosarioa
Hiztegia
10:47 AM-ean automatikoki gordetako proiektua
3/11 (18/281, 284)
16/16

Hobekuntzak OmegaT-n (II)

Wikipediako estekak itzultzeko laguntza


OmegaT-2.1.8_2 :: probaX
Proiektua Editatu Joan Ikusi Tresnak Aukerak Laguntza
Editorea - Host.UTF8
{{referencias}}
El término '''''host''''' es usado en [[informática]] para referirse a las [[computadora]]s conectados a una [[Red de computadoras|red]], que proveen y utilizan servicios de ella.
<segment 0002> ''''' Host''''' terminoa [[Informatika| informatikan]] erabiltzen da, [[Konputagailu-sare|sare]] bati konektatuta dauden eta sare horren zerbitzuak hornitzen eta erabiltzen dituzte [[Ordenagailu| ordenagailuak]] aipatzeko
<end segment>
Los usuarios deben utilizar "hosts" para tener acceso a la red.
En general, los "hosts" son computadores monousuario o multiusuario que ofrecen
Machine Translation
''''' Host''''' terminoa[[Informatika| informatikan]] konektatuta[[Ordenagailu| ordenagailuei]][[Konputagailu-sare| sare]] bati kontatu bere burua erabiltzen dute, ematen duten eta haren zerbitzuak erabiltzen dituzte.
<Matxin>
Parek...
Glosarioa
Hiztegia
La enciclopedia libre
Portada
Portal de la comunidad
Actualidad
Cambios recientes
Páginas nuevas
Página aleatoria
Ayuda
Donaciones
Notificar un error
Imprimir/exportar
WIKIMEDIA
[¡Ayúdanos
Host
Este artículo o sección necesita referencias que aparezcan en una pub
páginas de Internet fidedignas.
Puedes añadirlas así o avisar al autor principal del artículo en su página de discus
Para otros usos de este término, véase Nicolaus Thomas Host.
El término host es usado en informática para referirse a las computadoras conectadas a una red, qu
En general, los anfitriones son computadores monousuario o multiusuario que ofrecen servicios de
que hacen uso de los anfitriones pueden a su vez pedir los mismos servicios a otras máquinas conec
que se encuentra interconectado con uno o más equipos. Un host o anfitrión es un ordenador que fu

Hobekuntzak OmegaT-n (III)

- Matxin itzultzailea eta Xuxen erabili ahal izatea.
- Wikipediako estekak itzultzeko laguntza:
  - [[red de computadoras|red]] ---> [[konputagailu-sare|sare]]
  - [[gravedad|gravedad]] ---> [[grabitazio|larritasuna]]

    larritasuna edo grabitazioa??
- Wikipediako artikuluak inportatzeko eta esportatzeko funtzionalitate berriak

OmegaT programaren bertsio hobetua:
http://ixa2.si.ehu.es/glabaka/OmegaT/OpenMT-OmegaT.zip

Hobekuntzak Euskal Wikipedian

Euskal Wikipedian sortu diren 100 artikulu berriak.

(50.000 hitz)
http://eu.wikipedia.org/w/index.php?title=Berezi:ZerkLotzenDuHona/Txantiloi:OpenMT-2&limit=250

wikigaiak4koa.pl perl programa.

Wikipediako kategoria bateko artikuluen lista.
Artikuluaren luzera beste lau hizkuntzatan.
Honela erabili dugu guk:

- Euskal Wikipediako hutsuneak identifikatzeko.
- Katalanezko Wikipedian Informàtica kategoriako artikuluak, gaztelaniaz eta ingelesez bai baina euskaraz ez zeudenak bilatzeko.

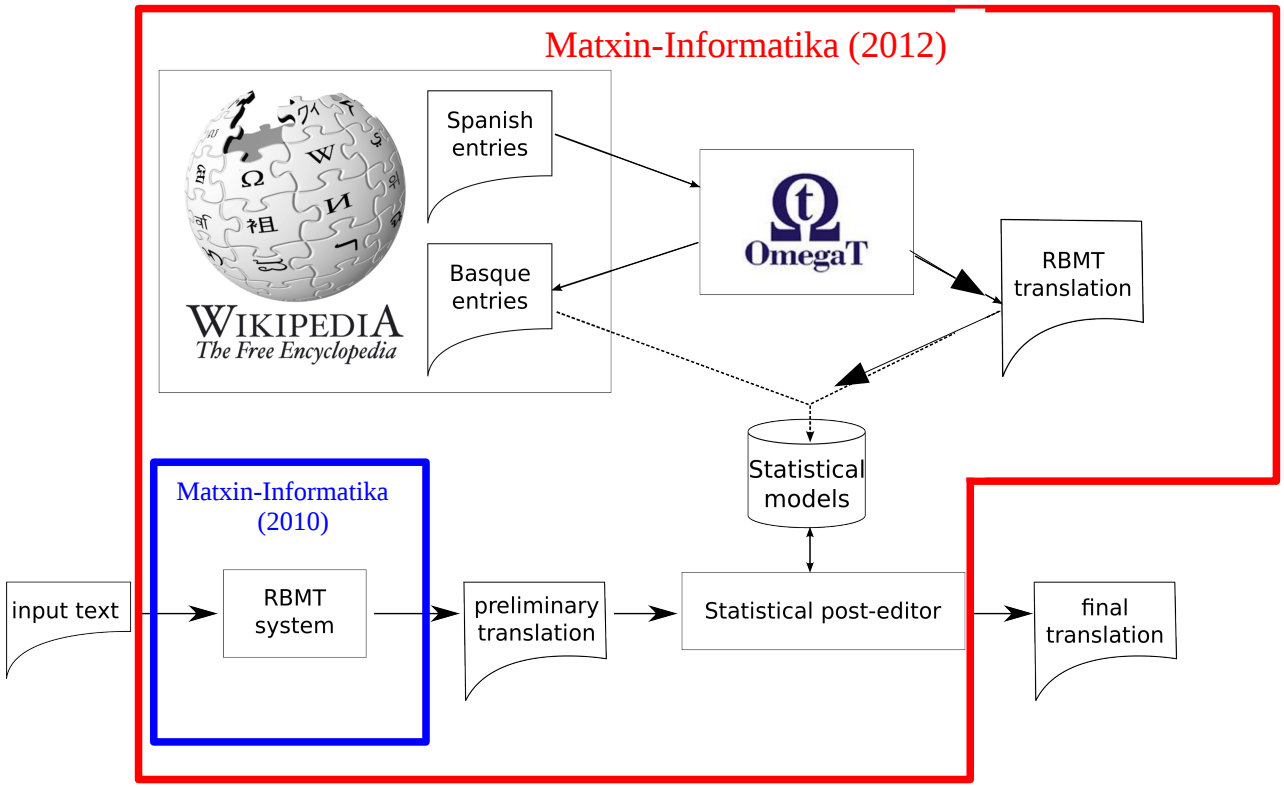
Hobekuntzak Matxin itzultzailean

- Matxinen bertsioa informatika gaietarako (2010)
  - Informatikako gaietarako egokitua
  - Hodeian erabiltzekoa (SOAP zerbitzua)
- Espainiera/euskara corpus paralelo bat
  - Mozilla softwarearena. (Elhuyar eta Julen Ruiz).
- Testu itzuliak eta horien eskuzko zuzenketak biltzen dituen corpus bat
  - Espainierazko Wikipediako 100 artikulu horiek Matxin itzultzailearekin sortutako itzulpenak dituena, gure kolaboratzaileek egin dituzten zuzenketekin, noski. (50.000 hitz)

Hobekuntzak Matxin itzultzailean (II)

- Matxinen bertsio berri bat postedizio estatistikoaren bitartez hobetua (%10)

'Reciprocal Enrichment between Basque Wikipedia and Machine Translators."
The People's Web Meets NLP: Collaboratively Constructed Language Resources, Springer, 2013



Etorkizunerako lanak

- Wikipediako estekak eta Wikipediako barne-datuak hobeto erabiltzea postedizioan
  - Itzulpen-sistemaren lexikoa aberasteko
  - Domeinuaren arabera ordain egokiagoak hautatzeko

- Wikipediako informazio hori sakonki erabiltzea
  - Itzulpengintza automatikoan (beste modutara ere)
    Proiektu berriak (Tacardi eta QTLeap)
  - Hizkuntzaren prozesamenduan, orokorrean

Etorkizunerako lanak

- Informatika ez den beste arlo batean errepikatzea?

  ... baina boluntario kopuru minimo bat behar da!

  - Boluntario-lana lortzea zaila izan da gurean
    - 100.000 hitzeko itzulpen/zuzenketa egin nahi genuen
    - 50.000 hitzekin moldatu behar izan gara.
  - baina azkenean lortu ditugu gure helburuak :-)

OpenMT2 eta Euskal Wikipediatik...
boluntarioei gure eskerrik beroena!

- Acprisip
- Ana Zelaia Jauregi
- Glabaka
- Irasgo
- Izaskun Etxeberria
- Jiparlaa
- Juanan
- Kgojenola
- Natxo
- Olatzarregi
- Txelo
- UnaiBeleko
- Unai Fdz. de Betoño
- Xsarasola
- Xartola
- Asoraluze
- Anderintxa
- Txibi72
- Joseba Izagirre
- Iñaki.Alegria
- NoraAranberri
- oarbelaitz
- Roberto
- Manex Agirrezabal
- e-gor
- bertolarrieta
- joseba.makazaga
- utolotu
- alexgabi
- Koru
- aingeru
- JesusIbañez
- Jesus Aparizio
- a.ilarraza
- ccplaore