IEB2013
Telefonoak eta Informatika

LABURPENAK

1. Paperean banatzeko laburpena.

- Zure hitzaldi/komunikazioaren abstract bat,
- 10 lerrotakoa, arial-10 letra motarekin 1000 karaktere inguru,
- Irudirik ez sartu.

Itzulpenak egiteko OmegaT itzulpen-plataformari hiru hobekuntza egin dizkiogu euskaraz eta Wikipediarekin aritu ahal izateko: (1) Matxin itzultzailea eta euskarazko zuzentzailea erabiltzeko aukerak, (2) Wikipediako artikuluak inportatzeko eta esportatzeko funtzionalitateak, eta (3) Wikipediako estekak errazago itzultzeko laguntza ere bai.

Itzulpen automatikoa ikertu duen OpenMT-2 proiektuaren barruan eta Euskal Wikipediaren laguntzarekin burutu dira hobekuntzok. Itzulpen automatikoaren bidez Wiikipedian gehitu diren 100 sarrera berriekin 100.000 hitzeko corpus berri bat sortu da, eskuz posteditatutako itzulpenak biltzen dituena. Horri ezker gai horrekin dabilen programa itzultzaile berri bat sortu da, postedizio automatikoaren bidez %10eko hobekuntza lortzen duena.

OmegaT programa libre eta irekia hobetzeaz gain guztira beste bost baliabide berri sortu ditugu proiektu honetan, aurrerago Hizkuntza-Teknologian berrerabili ahal izango direnak.

2. Egunean eskuratzeko laburpena.

- Zure hitzaldi/komunikazioaren laburpena,
- Testua eta irudiak (pantallazoak) sartu daitezke,
- Gehienez 2 orri: arial 10, 7000 karaktere inguru.

(idatzi hemendik aurrera)

UEUren idazkaritza: Irati Iciar
943 82 14 26 – i.iciar@ueu.org

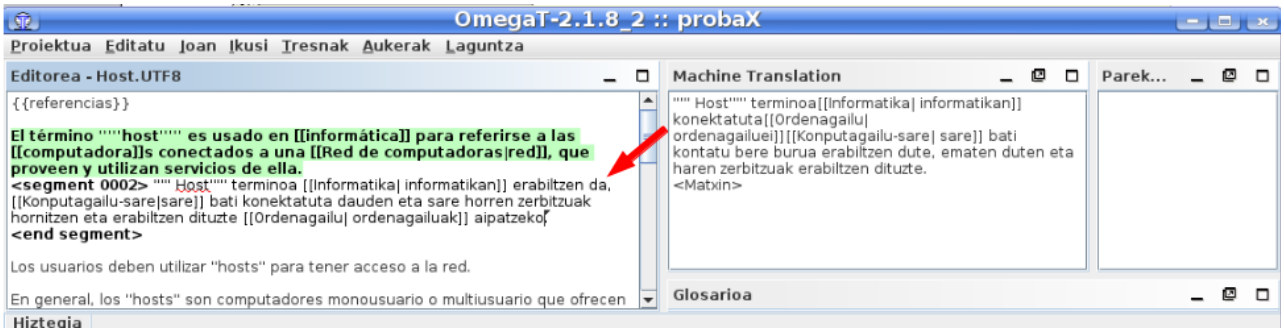
“OpenMT2 eta Euskal Wikipedia” wikiproiektuaren emaitzak.
OmegaT itzulpen-tresna hobetuta euskaraz eta Wikipediarekin aritu ahal izateko.

Iñaki Alegria*, Unai Cabezón*, Unai Fernandez de Betoño**, Galder Gonzalez**,
Mikel Iturbe**, Gorka Labaka*, Kepa Sarasola*, Arkaitz Zubiaga**
* Ixa Taldea https://ixa.si.ehu.es
** Euskal Wikipedia http://eu.wikipedia.org

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 (Wikipediako definizioa da).

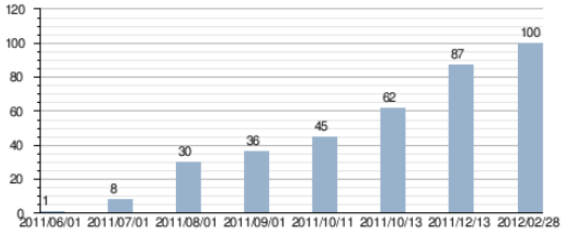
Guk egin dugun ikerketa honen barruan zenbait tresna eta baliabide sortu ditugu OmegaT programa hori hobetzeko, euskaraz eta Wikipediarekin aritu ahal izateko, hain zuzen ere. Hobekuntza horietako batzuk beste hizkuntzetako erabiltzaileek ere erabili ahal izango dituzte. Hiru dira gure ekarpenak:

- Matxin itzultzailea eta euskarazko zuzentzailea OmegaT-n erabiltzeko aukerak sortu ditugu.
- Wikipediako artikuluak inportatzeko eta esportatzeko funtzionalitateak gehitu ditugu.
- Wikipediako estekak itzultzeko laguntza ere bai. Adibidez, erdarazko *[[gravedad|gravedad]]* esteka *[[grabitazio|larritasuna]]* itzultzen du gure itzultzaileak, non *larritasuna* itzultzaile automatikoak lortzen duen itzulpena den (kasu honetan ez da ordain egokia) eta *grabitazioa* gure beste programa batek lortzen duen itzulpena, hizkuntzen arteko Wikipedia barruko informazioa erabilita (erdarazko Wikipediako *gravedad* artikuluak euskarazko Wikipedian duen baliokidea *grabitazio* artikulua da). Posteditore gizakiarentzat laguntza ederra da automatikoki lortzea Euskal Wikipediako lotura (*grabitazio*) eta gainera hori ikusita erraz zuzendu dezake Matxin-ek aukeratu duen ordaina testuinguru horretan egokia ez bada (*larritasuna-->grabitazioa*).


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

OmegaT programarako hobekuntza horiek luze probatu ditugu, 2009tik 2012ra bitartean itzulpen automatikoa ikertu duen OpenMT-2 proiektuaren barruan eta, beti ere, Euskal Wikipediaren laguntzarekin. 36 boluntarioren lankidetzaren bitartez 100 artikulu berri gehitu ditugu Euskal Wikipedian informatikako gaiei buruz, guztira 50.000 hitz izan dira. OmegaT programa aberastua eta Matxin itzultzailea erabili dituzte boluntarioek hasieran itzulpen-zirriborroak sortzeko, espainierazko Wikipediako testuak itzulita, eta ondoren zirriborro horiek zuzendu eta Wikipedian bertan zuzenean argitaratzeko.

Proiektuan itzulitako wikipediako artikuluak.



Lan horri esker 100.000 hitzeko corpus berri bat sortu dugu, eskuz posteditatutako itzulpen horiek biltzen dituena. Corpus hori baliatuta eta teknika estatistikoak erabiliz Matxin sistemarekin lortutako itzulpenak automatikoki posteditatzen dituen programa berri bat sortu da, eta emaitzek erakutsi dute sistema berri honek %10 hobekuntza lortzen duela, Matxin sistema soilarekin konparatuz gero (Alegria et al., 2013).

UEUren idazkaritza: Irati Iciar
943 82 14 26 – i.iciar@ueu.org

Guztira hauek dira sortu eta modu irekian plazaratu ditugun produktuak:

- OmegaT programaren bertsio hobetua, eta berau erabiltzeko eskuliburua.
- Matxin itzultzailearen bertsio berria. Informatikako gaiak itzultzeko egokitu duguna. SOAP zerbitzu moduan definituta dago.
- Euskal Wikipedian sortu diren 100 artikulu berriak.
- Espainiera/euskara corpus paralelo bat. Mozilla softwarearen lokalizazioan sortu denaren bertsio berria. Eskerrak Elhuyarri eta Julen Ruiz-i.
- Testu itzuli eta zuzenduen corpus bat. Espainierazko Wikipediako 100 artikulu horiek Matxin itzultzailearekin sortutako itzulpenak dituena, gure kolaboratzaileek egin dituzten zuzenketekin, noski.
- wikigaiak4koa.pl perl scripta. Wikipediako kategoria batean dauden artikuluen lista ematen dituen perl programa. Artikulu bakoitzarekin beste datu interesgarri batzuk ere ematen dira: ea beste hiru hizkuntzetako wikipedietan dagoen eta zer luzera duen hizkuntza horietan. Script hau baliagarria da euskarazko Wikipedian dauden hutsuneak identifikatzeko. Konkretuki guk erabili dugu katalanezko Wikipedian Informàtica kategoriako artikuluak bilatzeko, artikulu horietako bat gaztelaniaz eta ingelesez bai baina euskaraz ez bazegoen, eta artikulu laburra bazen (10-20 lerro), artikulu hori gehitzen genuen boluntarioei eskaintzen genien Euskal Wikipedian sartzeko proposamenenen artean.

Etorkizunean Wikipediako esteketan eta bere barne antolakuntzan dagoen informazioa era sakonagoan erabiltzea aurreikusten dugu, itzulpen-sistemaren lexikoa aberasteko eta domeinuaren arabera ordain egokiagoak hautatzeko. Informatika ez den beste arlo batera ere zabal genezake gure sistema, baina boluntario kopuru minimo bat dagoela ziurtatu beharko litzateke aurretik. Boluntario-lana lortzea zaila izan da gurean. Bagenekien hasieratik euskara bezalako hizkuntza minoritario batean normala izango zela hori gertatzea, baina lortu dugu gure helburua, eta horrexegatik boluntarioei gure eskerrik beroena emanez bukatu nahi dugu, eurak izan baitira proiektuaren emaitza arrakastatsu hauek lortzea ahalbidetu dutenak.

Erreferentziak

1. Aduriz I., Alegria I., Artola X., Díaz de Ilarraza A., Sarasola K. 2011 Teknologia garatzeko estrategiak baliabide urriko hizkuntzetarako: euskararen eta Ixa taldearen adibidea.*Linguamatica — ISSN: 1647–0818, Vol. 3 Núm. 1 - Junho 2011 - Pág. 13–31*
2. Alegria I., Arregi X., Díaz de Ilarraza A., Labaka G., Lersundi M., Mayor A., Sarasola K. 2008. Strategies for sustainable MT for Basque: incremental design, reusability, standardization and open-source. Proceedings of the IJCNLP-08, pp: 235-243. Hyderabad, India.
3. Alegria I., Aranzabe M., Arregi X., Artola X., Díaz de Ilarraza A., Mayor A., Sarasola K. 2011 Valuable Language Resources and Applications Supporting the Use of Basque *Z. Vetulani (Ed.): LTC 2009, Lecture Notes in Artifitial Intelligence LNAI 6562, pp. 327--338. Springer, Heidelberg. ISBN:978-3-642-20094-6, DOI: 10.1007/978-3-642-20095-3, http://www.springerlink.com/content/c8608h56n4201312/*
4. Iñaki Alegria, Unai Cabezon, Unai Fernandez de Betoño, Gorka Labaka, Aingeru Mayor, Kepa Sarasola and Arkaitz Zubiaga. 2013. “Reciprocal Enrichment between Basque Wikipedia and Machine Translators”. To be published in "The People’s Web Meets NLP: Collaboratively Constructed Language Resources", book edited by Iryna Gurevych and Jungi Kim, Springer, Book series "Theory and Applications of Natural Language Processing", E. Hovy, M. Johnson and G. Hirst (eds.). *ISBN-10: 3642350844 | ISBN-13: 978-3642350849*
5. *Díaz de Ilarraza A., Labaka G., Sarasola K. 2008. Statistical Post-Editing: A Valuable Method in Domain Adaptation of RBMT Systems MATMT2008 workshop: Mixing Approaches to Machine Translation. pp.35-40.*
6. Mikel Iturbe, Unai Fdz. de Betoño, Galder Gonzalez, Arkaitz Zubiaga, Iñaki Alegria, Gorka Labaka, Kepa Sarasola 2010. Reciprocal Enrichment between Wikipedia and Machine Translators *Wikimania 2010 Gdańsk, Poland July 9-11, 2010, http://wikimania2010.wikimedia.org/wiki/Submissions/Reciprocal_Enrichment_between_Wikipedia_and_Machine_Translators*
7. *Simard, M., Ueffing, N., Isabelle, P., and Kuhn, R. 2007. Rule-based translation with statistical phrase-based post-editing Proceedings of the Second Workshop on Statistical Machine Translation. pp:203-206. Prague, Czech Republic.*

UEUren idazkaritza: Irati Iciar
943 82 14 26 – i.iciar@ueu.org