

Memoria da ferramenta libre (GNU/GPL) para a consulta de terminografía informática «Tiboxe»



Rúa de A Plaza, Figueroa, nº 28

15318 Abegondo

A Coruña

Telf.: 600519089

info@trasno.net

20 de maio de 2010

Táboa de contidos

Historial da entidade.....	3
Obxectivos da actividade	3
Exposición e descrición das características do traballo presentado	4
Medios utilizados, formatos estándares utilizados	5
Detalles técnicos de software e hardware necesarios	5
Hardware servidor	6
Software servidor.....	6
Dominio servidor.....	6
IP público.....	6
Información relevante para un bo desenvolvemento do contido.....	6
Licenzas.....	6
Dereitos.....	7
Outros aspectos inherentes ao proxecto obxecto de valoración competitiva conxunta.....	7
Os URL de acceso e publicación	7

Historial da entidade

A "**Asociación Cultural para a Galeguización de Software Libre Proxecto Trasno**" é unha organización sen ánimo de lucro de ámbito autonómico galego que representa dende o 2009, legal e financeiramente, o **Proxecto Trasno**.

Os fins da asociación, conforme os seus estatutos, son entre outros os seguintes:

- Promover a iniciativa e os grupos de localización (adaptación lingüística-cultural)
- Traducir e localizar software libre ao galego
- Crear e xestionar recursos e ferramentas para os voluntarios que localizan ao galego
- Contribuír á mellora do status da lingua galega mediante a inserción desta no ámbito das novas tecnoloxías así como á fixación do tecnolecto no ámbito da enxeñaría e a sociedade da información
- Representar a comunidade de persoas e grupos do Proxecto Trasno

O Proxecto Trasno é unha comunidade en rede que reúne aos colaboradores en proxectos de localización (isto é, tradución e adaptación) de software libre ao galego. O proxecto nace como unha comunidade “virtual” de persoas dedicadas á actividade de localización de, exclusivamente, software libre no ano 2000 a partir de dous grupos que comezaran a traballar independentemente un ano antes. Durante dez anos de labor ininterrompida o Proxecto estivo detrás de traducións de distribucións GNU/Linux, contornos de escritorio, software ofimático entre outros de ampla presenza social e uso masivo. Este labor facilita actualmente gran variedade de software libre dispoñible en galego, e que segue sendo actualizado e mantido polos voluntarios do proxecto.

Este software é de utilidade para todo tipo de persoas, tanto no ámbito persoal coma no profesional. Adoita ser, asemade, o software máis adaptable ás persoas con discapacidades e constrúese para ser interoperable como demanda hoxe en día a lexislación tanto europea como española, en beneficio da seguridade informática a utilidade e sostibilidade social así como o favorecemento da independencia persoal.

No ano 2009 membros do proxecto deciden fundar a asociación homónima para solventar os problemas derivados da falta de representatividade legal, a pouca visibilidade como grupo e darlle un pulo á actividade de localización ao galego.

Obxectivos da actividade

A creación dun recurso lingüístico que lles permita aos tradutores asegurar a calidade da súa contribución, a estandarización dos produtos en beneficio dos usuarios e a maior rendibilidade do esforzo colaborativo.

Na práctica é unha ferramenta de consulta para saber como se traduce un termo inglés ao galego dependendo do contexto e o proxecto no que aparece esa cadea.

Mediante este recurso aumentará a cantidade e calidade das traducións que a comunidade galega de localización realiza ao conseguir fixar termos e lograr unha maior uniformidade das localizacións

dos diferentes proxectos. Recurso de indubidable valor para conseguir que a normalización das TIC teña o seu dicionario particular especializado no campo informático. O rexistro e explotación de traducións permite realizar un seguimento das mellores prácticas, a mellora continua dos produtos e fluxos de traballo e corrección así como a consolidación de especializacións profesionais.

Cómpre salientar que os dicionarios da lingua galega dispoñibles, tanto os oficialmente normativos da RAG coma os comerciais non recollen os usos no ámbito informático e, nos casos en que o fan, son apriorísticos, pouco documentados e imprecisos. O sector profesional da tradución non encontra neles solución para as necesidades do seu traballo.

En esquema:

- Creación dun recurso lingüístico. A sociedade galega carece de referencias neste campo.
- Aumentar a calidade das localizacións, ao poder asegurar a súa estandarización.
- Contribuír á mellora das industrias da lingua e á formación de especialistas no campo da tradución e do desenvolvemento de ferramentas neste ámbito.
- Normalización das TIC ao reforzar a capacidade do noso idioma para responder ás crecentes necesidades da especialidade. Prestixio cultural e accesibilidade á tecnoloxía.
- Homoxeneizar e, en gran parte dos casos, fixar a solución lingüístico-técnica idónea.
- Produtividade, economía de recursos, reutilización de coñecemento.
- Difusión e seguimento de terminoloxía.
- Creación e xestión de software de valor co que colaborar nas redes internacionais do sector e especialización de desenvolvedores de aplicativos.

Exposición e descrición das características do traballo presentado

Desenvolvemento dun aplicativo web libre, chamado Tiboxe, para a consulta de unidades terminolóxicas sobre a realidade dun corpus de traducións galego-inglés emparelladas. A ferramenta, en si mesma, admitiría o traballo sobre corpus lexicográficos de pares de linguas diferentes, xeraría resultados simplificados e ponderados segundo diferentes criterios, así como presentaría o resultado con ordenacións, estatísticas relevantes, frecuencia ou recomendación da solución, tería un realce da comparación dos termos de orixe e meta e permitiría a exportación dos resultados.

O traballo da tradución almacénase en memorias de tradución e glosarios en formato estándar da industria (TMX e TBX): bases de datos que almacenan palabras, frases ou parágrafos coa correspondencia nun idioma, habitualmente o inglés, e o idioma ao que se traduce.

Hoxe por hoxe, os aplicativos empregados na localización (isto é, a tradución e adaptación de software a unha lingua) permiten tanto o emprego de formatos estándar interoperables e independentes das compañías comerciais, coma o intercambio de memorias de tradución para seren usadas co fin axilizar a tradución de novos aplicativos de forma máis barata e eficiente en recursos económicos e persoais. Mais aínda, eventualmente, a partir de cada proxecto ou do traballo de cada tradutor pódense extraer memorias de tradución especializadas co que se xera unha cantidade

potencial elevada de memorias de tradución para un mesmo idioma. As MT son un recurso técnico de alto valor económico que adoita estar suxeito ás restricións comerciais das empresas que desenvolven software comercial.

Por outra banda e segundo a nosa experiencia de anos de dedicación, a procura nas memorias dispoñibles dá como resultado concordancias complexas nas que aparecen resultados repetidos, ou con lixeiras variacións, e incluso non depurados, por proceder dunha tradución orixinal non corrixida. Por iso, faise preciso, para os efectos de consistencia nas localizacións e do control de calidade das traducións, dispoñer de ferramentas que mostren resultados de moitas memorias de tradución, e que, ademais de estaren actualizados, mostren datos filtrados e depurados.

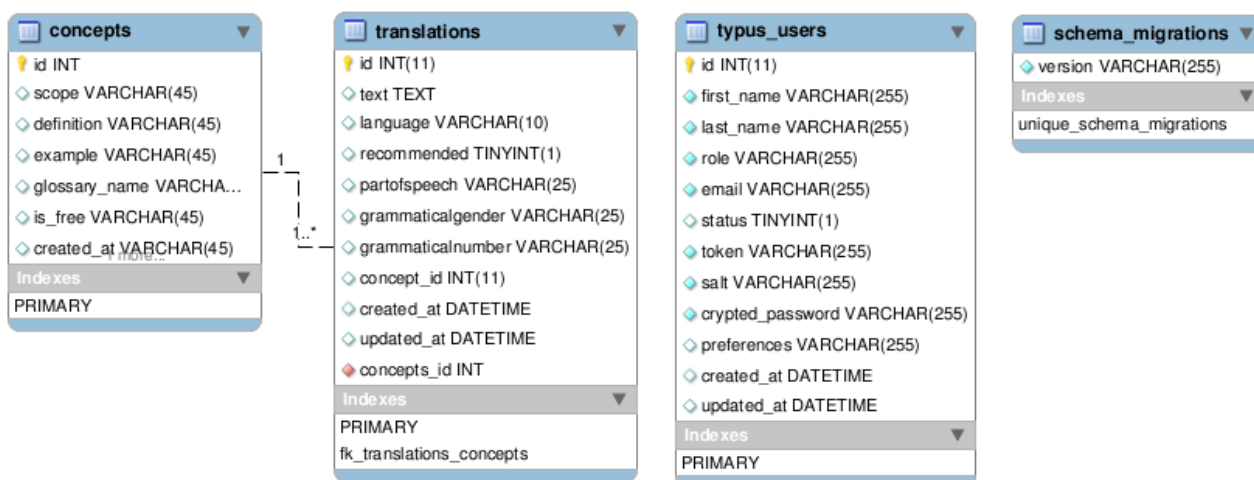
Con isto os tradutores poderían acceder a múltiples memorias de tradución dun xeito fácil, simplificado e libre que permite mellorar a calidade das novas traducións nas que traballan.

Coa mesma ferramenta, tamén é posible a explotación de datos para os efectos de investigación no sector traductolóxico e lingüístico xa que permite estudos contrastivos, comparativos, históricos, por ámbitos tecnolóxicos, por contornos e tecnoloxías.

Do mesmo modo a acumulación de fontes de datos de forma estándar é sostible, e permitirá a mellora ou desenvolvemento de novos aplicativos ou outros produtos derivados en función dos múltiples intereses relacionados, non só no ámbito tecnolóxico senón en calquera ámbito onde a globalización da sociedade, a converxencia de soportes e a informatización dos mercados ou da cultura necesite claridade e precisión na comunicación.

Medios utilizados, formatos estándares utilizados

Está programado co contorno de traballo *Ruby on Rails* na linguaxe de programación de código aberto Ruby mediante metodoloxías *áxiles*. O código mantense nun sistema de xestión de proxectos Redmine (<http://tickets.trasno.net/projects/tiboxe>) que permite a colaboración e o seguimento profesional de versións.



O aplicativo basease nunha base de datos relacional cun modelo de datos moi sinxelo e que segue as indicacións do estándar TBX máis facendo unha translación ao seu homónimo relacional. Na figura

anterior pódese ver unha ilustración na que se distinguen as táboas na base de datos e as súas relacións.

O aplicativo web conta cunha interface de administración que posibilita a importación e exportación de novos glosarios alén de modificar as distintas entradas existentes.

Conta asemade cun sistema simplificado de acceso a glosarios non libres, para os cales é preciso ter un usuario cos privilexos de acceso necesarios no sistema correspondente. Isto evita posíbeis problemas legais.

O aplicativo web como tal cumpre os seguintes estándares do WorldWide Web Consortium (W3C):

- HTML 1.1 transicional
- CSS nivel 2
- WCAG 2.0 (Accesibilidade, nivel AA)

A base de datos empregada é MySQL pero admite compatibilidade con calquera outra que admita o estándar ANSI SQL como son: SQLite, PostgreSQL, etc.

Á parte de presentar os resultados mediante páxina HTML, tamén se poden consultar mediante servizo web usando tecnoloxías REST. Isto último permite crear un aplicativo de escritorio para modificar a base de datos ou mesmo *consumir* os datos do servizo web REST desde calquera aplicativo web de terceiros.

Os resultados son visualizados e compartidos mediante o estándar para o intercambio de terminoloxía TBX do grupo internacional de estandarización LISA.

Detalles técnicos de software e hardware necesarios

Precisase dun servidor conectado á Internet e dun nome de dominio no que o recurso web estará accesíbel para todo o mundo.

Hardware servidor

Actualmente está traballando (xuntamente con outros servizos) nun servidor de aloxamento alugado a un ISP privado coas seguintes características:

Especificacións técnicas

- | | |
|--------------------------|--|
| • Procesador | Intel Atom Dual , 2x 1.6, L2: 2x 512Ko, FSB: 533MHz, Dual Core |
| • Arquitectura | 32 - 64 bits |
| • Memoria RAM | 1 GB DDR2 |
| • Espazo disco SAN | 20 GB. |
| • Acceso disco | ISCSI |
| • Rendemento ISCI mínimo | 1 MB/se |
| • Conexión | 100 Mbps |

- Largo de banda 100 Mbps
- IP principal 1 enderezo
- IP Fail-over +1 enderezo

Software servidor

- Sistema operativo GNU/Linux Debian (aínda que pode executarse en calquera sistema operativo)
- Servidor web Apache 2 MPM
- Servidor de bases de datos relacionais, idealmente MySQL ou PostGres.
- Ruby on Rails versión 2.3.5
 - Xemas de Ruby instaladas no servidor
 - Passenger, para acelerar o acceso e procesado de solicitudes
 - Typus, contorno de traballo onde se constrúe o xestor de administración.

Dominio servidor

<http://corpus.trasno.net> accesible dende o menú RECURSOS do web de Proxecto Trasno
<http://trasno.net>

IP público

Ao ser un aloxamento en réxime de alugueiro pode cambiar. Neste momento responde no IP 188.165.49.203

Información relevante para un bo desenvolvemento do contido

Licenzas

O software desenvolvido licenciase consonte os termos da Licenza GPL v3. Os contidos e datos que compoñen o software son consontes coa licenza Creative Commons By-Sa 3.0. O software é multiplataforma.

Dereitos

Todos os dereitos de autor, agás os legalmente intransferíbeis como o de propiedade intelectual, foron cedidos, polos seus desenvolvedores a "Asociación Cultural para a Galeguización de Software Libre Proxecto Trasno" rexistrada no Rexistro Central de Asociacións da Xunta de Galicia co número 2009/013660-1 (SC). Ten o NIF G70234315.

Outros aspectos inherentes ao proxecto obxecto de valoración competitiva conxunta

Interface de consulta: o usuario só precisa acceso á Internet e un navegador web estándar. Ao acceder á páxina ofrécese unha interface minimalista coa que en dous clics pode obter os resultados buscados. Simplemente facer clic na caixa de busca, escribir o termo e premer en buscar, mostraranlle os resultados ao instante.

Ademais ofrece un servizo web desde o que outros aplicativos poden acceder e integrar a súa funcionalidade empregando REST.

O aplicativo está pensado para o inglés-galego, pero pode usarse para facer a correspondencias entre calquera par de idiomas. E a base de datos admite a posibilidade incremental de 1 a n , co que simplemente habería que modificar a vista do recurso para engadir esta nova funcionalidade.

O aplicativo permite descargar/imprimir e interoperar con outros sistemas que tamén sexan respectuosos cos estándares de xestión de datos lingüísticos.

Os URL de acceso e publicación

URL de acceso público: <http://corpus.trasno.net>

URL publicación do código fonte:

- Acceso en liña (web) á lectura do código: <http://github.com/frandieguez/tiboxe>
- Orde para descarga do código: `git clone git://github.com/frandieguez/tiboxe.git`