

language technology
for
optimum localization

Diego Bartolomé, CEO

optimum workflow

gather in-domain data

train the translation solution

enrich solution with related text

terminology prioritization

update the translation solution

add rules to enhance quality

weekly updates

data issues 1

<large volume of heterogeneous data>

training with all the data

semantic classification for domain selection

fine tuning for each client

glossary prioritization

continuous machine learning

data issues 2

<scarce data>

add dictionaries into corpora

complementary segments from memories

balance client data with generic texts

in-domain adaptation of generic system

increase the number of sentences with rules

data issues 3

<dirty data>

remove multiple translations

eliminate text in other languages

correct spelling

select sentences with correct grammar

automatic alignment with client terminology

filter out other undesired segments

data issues 4

<data creation and enhancement>

final client defined

unaligned translated documents

generic translations

optimum corpus/memories creation

rule-based extension/filtering

linguistic issues 1

<untranslated words>

dictionary creation

<grammatical errors>

post-processing rules

<blind quality filtering>

do not translate sentences below threshold

linguistic issues 2

<source text cleaning>

spelling and grammar

sentence simplification

terminology homogenization

<special words detection>

people, places, organizations

alphanumeric codes

use case

<recurrent small volumes>

frequent translations

clients from different domains

<workflow>

gather as much data as possible

receive a new file for translation

create an ad hoc domain for that file

train the translation solution + basic rules

<output>

optimum adaptation for a file in around 4 hours

<tauyou_text> snapshot

<fully customizable>

look and feel + functionality

tauyou:text MT

home translate create update calculate download

| | | |
|--------------------------|---|-----------|
| Source Language | Target Language | Domain |
| Spanish ▾ | English ▾ | Generic ▾ |
| Translations Glossary | <input type="button" value="Choose File"/> No file chosen | |
| Forbidden Words Glossary | <input type="button" value="Choose File"/> No file chosen | |
| File | <input type="button" value="Choose File"/> No file chosen | |
| Email: | <input type="text"/> | |

©2011 Ta with you

Thanks!

// Diego Bartolomé, PhD

<address> C/ Les Planes 39 – 08201 Sabadell – Spain

<phone> +34 93 711 29 96

<cell> +34 670 331 225

<email> dbc@tauyou.com

<www> tauyou.com