Machine Translation Trainer Tools

version 0.92 from 21 January 2021

CUSTOM . MT effectiff services

Machine translation training specialists are emerging as a new profession. Custom.MT and Effectiff assembled a list of tools of this trade.

DATA PIPELINE
/acquisition to anonymization

Retrieving parallel corpora from the Internet
- Scrapy
- Wget
- Heritrix
- HTTrack
- Creepy
- ParaCrawl

Open-source datasets
- OPUS
- ParaCrawl
- Europarl Corpus
- United Nations Parallel Corpus
- CLARIN
- Salesforce Corpora

Dataset marketplaces
- TAUS Data Marketplace
- Flitto
- ELREC
- DefinedCrowd

Automatic TMX cleaning and alignment
- TMop
- Bicleaner
- Bifixer
- Zipporah
- Hunalign
- Bleualign
- Vecalign
- Gargantua
- UnsupPSE

Manual alignment
- AlignAssist
- LF Aligner
- eAilgn
- ABBYY Aligner
- NOVA Text Aligner
- Tmxmall aligner
- CAT tool built-in features

Manual TMX Editing
- Okapi Olifant
- Heartsome TMX Editor
- Goldpan TMX Editor
- TMX editor (open-source)
- TMXEditor (Maxprograms)
- Egotech TMX Editor
- CAT tool built-in features

Term extraction
- SynchroTerm
- Okapi Rainbow
- Sketch Engine
- SDL MultiTerm Extract
- TermSuite
- AntConc
- CAT tool built-in features
- Egotech TermExtract
- XTM Vector Space

Anonymization, removal of sensitive data
- MAPA
- Scripts based on regular expressions
- and NER models:
- spaCy
- stanza
- NER-BERT
- GTE localize

TRAINING
/MT engines

Trainable via an interface
- Globalese
- Microsoft Custom Translator
- Yandex Translate
- ModernMT
- Google AutoML
- Omniscien Language Studio
- KantanMT
- Tilde MT
- Amazon Translate
- SYSTRAN
- Rozetta Jukkou
- Promt
- IBM Watson Translator
- NewTranX

On premise trainable
- SYSTRAN
- Promt
- PangeaMT
- Cloud Translation
- TransPerfect AI

MT Frameworks
- TensorFlow NMT
- FairSeq
- Marian NMT
- OpenNMT
- Sockeye

EVALUATION
/compare performance

Automated MT quality evaluation
- Modelfront
- Comet
- Unbabel OpenKiwi
- Memsource Translate

Quality metrics and measuring tools
- BLEU
- hLEPOR
- TER
- METEOR
- chrF
- Intento Spotlight
- Appraise

- BERTScore
- BLEURT

Human evaluation tools
- TAUS DQF
- ContentQuo
- TQ Auditor
- Costa MT Evaluation Tool
- Egotech QA Metrics

IMPLEMENTATION
/integrate in applications

Middleware and Routing
- Intento MT Hub
- CrossLang Gateway
- Language I/O
- Memsource
- Smartling

MT-Independent NLP /terminology layer

Implementation Service
- Custom.MT
- Inten.to
- Semantix
- Welocalize
- TextShuttle