



KantanMT.com

Three Quick Tips for Improving MT Quality!

So your Machine Translation (MT) engine's not working as well as you'd like it to? OK, so the first thing you'll probably do is check out your Kantan BuildAnalytics dashboard to see what the *Insights* (little tips on how to improve your MT quality) tell you.

Improving the performance quality of your KantanMT engine will greatly reduce the amount of post-editing required and lead to faster turn-around times – which we love!

KantanMT *Insights* generally tells you one or two of a few action items to improve engine quality. Check them out below, and if you have tips and tricks that you've used, send them into info@kantanmt.com and we'll add them to the list.



1. Increase Training Data Size - Statistical Machine Translation (SMT) requires a large volume of aligned bilingual training data to 'learn' a language combination – and more complex languages require larger data sets because the systems will require more 'examples' to learn from.

Generally speaking, your KantanMT engine will improve in its ability to translate using your particular language style and terminology as you add more training data. We call this "training" your engine.





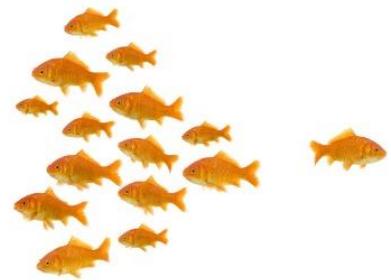
KantanMT.com

Three Quick Tips for Improving MT Quality!

There are two parameters you can monitor here:-

- **Number of Training Words:** This is the absolute number of words in your training data. We recommend that your engine should have at least 2 million words if it is to have sufficient domain knowledge to produce quality translations.
- **Number of unique words in your training data:** The number of unique words contained within your training data is a crucial element for building a successful engine. For example, it's estimated that there are approximately 500,000 unique words in the English language (there are an estimated 1 million for French and German, while other languages vary in size). So ensuring that your engine has a high number of unique words is the key to building an 'intelligent' engine. That is, an engine which has the ability to cover the full scope of the language.

A word of caution on training data! Your training data should be from the same domain! For example, if the domain is related to automotive, then the training data should be automotive-related. If it is anti-virus software, then training data should be related specifically to this domain. As in Translation Memory, you do not get good results if you mix the domain of your TMs when working with client files. The more precise and focused your domain training data the better the results.



More Words = More Brains

By uploading more training words from a similar domain, you increase the knowledge, understanding and intelligence of your KantanMT engine. In other words, it gets smarter!



KantanMT.com

Three Quick Tips for Improving MT Quality!

2. Terminology, Terminology, Terminology:

Your KantanMT engine just can't get enough of this! To maximise translation quality, ensure that you upload your terminology or glossary files into your '*Client Data*' on KantanMT.com. This will ensure that your KantanMT engine adopts this terminology during the translation process - maximising quality and dramatically reducing post-editing effort. The more you "educate, or train" your engine the better the end results will be.

3. Automate Post-Editing using PEX:

Unless you're after 'gist' translation (i.e. a rough translation of the source text that allows the reader to understand the "gist" of the content), you'll most likely have to apply some level of post-editing to your translations. Don't let this put you off, this is standard procedure with most machine translated text. However, you can automate part of this process using KantanMT's PEX technology. This will dramatically reduce the amount of manual post-editing required, saving you time and money.

OK, so three quick tips to keep you actively improving your Machine Translation quality for a while. Happy MT'ing!

