



Communication Sub-System

This report describes the browser demonstration prototype implemented for the integration of machine translation, translation memory, and other localisation tools deployed for accounting report translation (c.f. D3.2)

This prototype has the following interface:

- Upload of source files. The files needed by the system includes as well xbrl file to be localised and xsd descriptive file.
- Choice of options (source/target languages, currency conversion)
- Download of localised xbrl file

And performs the following processes:

- Validation of source xbrl file
- Build the internal tree structure of the xbrl file
- Apply general xbrl i/c localised xsd file for glossary substitution of generic labels
- Apply user additional xsd file for potential user provided label translation
- Translation of date/currency field
- Preserve DNT fields
- Translation Memory and Machine Translation on extracted free text
 - The interface between these two applications has been designed to keep correct analysis of sentences for Machine Translation after Translation Memory parse. In that goal, the whole free text is sent to Machine Translation, and a double check is performed on the output to keep only fully substituted TUs, and to reject partially translated TUs by Translation Memory. Indeed partial matched sentences are of no use for a machine translation application, since the system cannot perform some full analysis of the sentences (because some part are already translated), and the performed experiments showed that the translation quality was getting very bad if we allow mix of TM and MT in a same sentence

- At this stage, the translation memory is applied as a generic preparse module¹. The input of this module is raw text and the output is xml-like tagged corpus. Let us notice, that the full strength of translation memory is used with some user interactivity. In this demonstration prototype it seems more relevant to have a straight processing.
- Synthesis of the result as the new XBRL file.

Use of this browser is necessarily connected to some XBRL editor and viewer.

Prototype URL : www.systransoft.com/Transaccount.cgi

¹ To keep all genericity to the solution proposed, we have simplified the API so that any translation memory engine could be plugged.