Towards an automated translation workflow at Océ Technologies

Introduction

Although there are many tools around these days that make it possible to partly automate the translation workflow in a company, there is still a large amount of customisation needed to integrate commercial tools, eliminate error-prone and repetitive manual tasks, and create a framework for automated translation. Océ 's Translation Services has been active in the field of automated translation for about 6 years and have introduced Machine Translation and Translation Memory systems, a Controlled English checker and have been involved in the introduction of an XML-based authoring system. This article discusses just a few of the major internal developments required to achieve high quality translations at low cost in a short period of time.
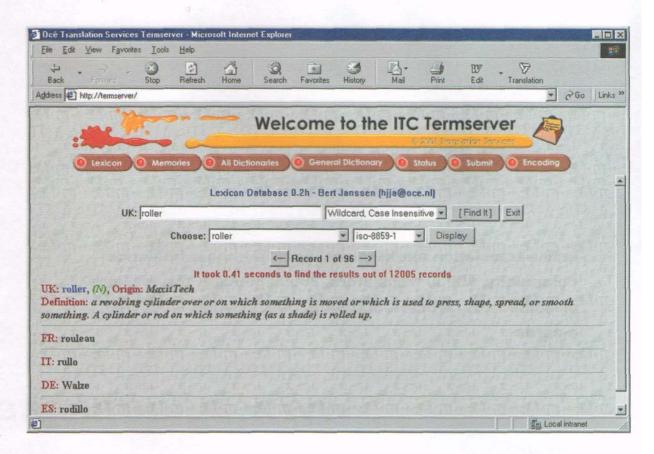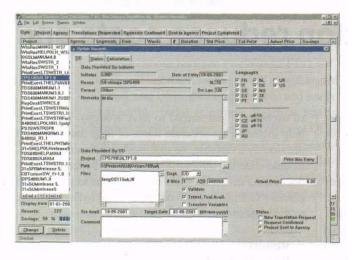
Workflow

During the past years, the Translation Services department of Océ Technologies had to build a workflow around the two mainstream commercial tools: Translation Memory (TM) and Machine Translation (MT). Along with the introduction of MT 6 years ago, we also developed a tool that allowed us to measure the efficiency of the automation process. The tool kept track of the traditional translation cost and the actual cost of a translation project after applying MT and TM tools. This tool has now evolved into a workflow tool that not only tracks costs and savings but also automates most of the workflow.

Through an intranet web-application, project coordinators can fill out a translation request form, which enforces consistent and meaningful project names plus all the information required to start the translation cycle. Project directory structures are created automatically and files are analysed in a two-level analysis. First a custom built tool analyses the source documentation against all memories and identifies the most relevant specific project memories. Next, a commercial TM tool analyses the results of the first run for all languages and produces the metrics for this specific project. The workflow tool reads the metrics and calculates costs and duration of the project.

The actual translation is outsourced. A translation package is generated, containing source, memories, log files and an automatically created readme file which contains the instructions and details for this particular project. With the push of a button the package is put on Océ's FTP server and the recipient is notified. The tool

also makes it possible to keep track of each project's status, to archive it automatically on completion and to store the resulting translation memories.

Terminology

In order to maintain terminological consistency and correctness both in the authoring process and the translation process, a lot of effort has been put into the development of a database to make Océ technical dictionaries available on intranet and - more recently - of a bilingual terminology mining tool. The recent introduction of Controlled English offered the possibility to standardise technical terminology. It also required the availability of a terminology browser. This led to the development of a custom multilingual database to support a web-based intranet browser that allowed to browse not only terminology but also all available memories in up to 17 languages. To this end, the translation memories were converted into a single database in ASCII format, without the original formatting. The very same database is used by the custom TM tool to analyse new documents.

Once source terminology (English) has been established, necessity arises to have unique and consistent translations available in a number of languages. Not only for distribution to the translators, but also to feed the Machine Translation system. To solve this problem, a bilingual terminology miner has been developed that searches the translation memory database for relevant terminology and the statistically most prominent translation. Initially the proposed translations can be filtered to include only the terminology allowed for Controlled English. The results are presented in an interface that allows a native speaker to select the correct translation from a list of proposed translations and allows to add specific information like class, gender etc. As an additional aid, also the original source and target segments the termcame from, are displayed. This allows for a fast build-up of multilingual corporate terminology which is also stored in the multilingual database. A flexible export facility enables exports for machine translation, reference for translators or any other purpose.

Conclusion

The fast evolving world of localisation, calls for continuous adaptation of one's workflow to new localisation products, or new versions of existing products. New tools need to be developed to keep up with a growing demand for translation in a shorter time-frame. General trends towards standardisation of document formats like XML, for instance, will eventually reduce some of the workload but at the same time requires a different approach to translation, as translating the contents of a document management system is different from translating traditional manual based content. However, from a localisation point of view, until XML has superseded legacy projects, it is just another format which needs handling and for which filters and procedures need to be developed.

Lou Kremer