HandyCAT

Chris Hokamp and Qun Liu
{chokamp, qliu}@computing.dcu.ie

CNGL/ADAPT/Dublin City University
Project Website: http://handycat.github.io
Github: https://github.com/chrishokamp/handycat

Project Description

We present HandyCAT, a new open-source Computer Aided Translation (CAT) tool, designed specifically for conducting research on Computer-Aided Translation. The User Interface (UI) itself, as well as the backend services, such as the Translation Memory engine, the MT system interface, the concordancer, and the glossary engine, are completely open-source.

The HandyCAT UI is implemented as a web application which runs in any modern browser. HandyCAT uses the XLIFF standard, and supports the core elements from both the XLIFF 1.2 and XLIFF 2.0 standards. GraphTM, the graph-based translation memory component, supports the TMX format, as well as several text input formats.

We introduce a factorization of the core interface components which allows a CAT tool to be viewed as a collection of standalone components connected by consistent APIs, facilitating research on new user interactions such as multi-modal input and interface control, and on new components created specifically for the post-editing task. Because the tool is designed primarily for CAT research, we have also designed a logging API which allows component creators to design logging customizable logging behavior for their components.

Although several open-source CAT tools have already been developed, no web-based tool provides a full CAT ecosystem as an open-source platform, including all user interface components and data services. Because the backend data services are prerequisites for a modern CAT interface, it can be difficult to design and conduct new user studies using existing open-source interfaces.

HandyCAT is built around the concepts of *containers* and *interactive areas*. Any CAT tool has some standard components which can be presented to users in various ways. Both the visual presentation and the interaction design will have an impact on the translator's experience. Therefore, HandyCAT is designed to allow researchers to create parameterized components which are easy to test and modify.

Several translation services provide free and/or paid APIs to proprietary services such as translation memories, machine translation, and glossaries. Connecting these APIs with HandyCAT is straightforward, allowing users and researchers to quickly integrate new services, or existing services which may have designed for other purposes.

All components of HandyCAT are completely open-source, meaning that the tool can easily be extended and improved. Because modern CAT tools are complex applications, developing a baseline tool with standard features requires significant effort. By using HandyCAT, researchers can implement only the components relevant to their work, while relying on the platform to provide the core CAT tool functionality, and to provide the statistics and logging necessary for analysis.