Context-based Evaluation of MT Systems: Principles and Tools

Maghi King, Andrei Popescu-Belis and Paula Estrella ISSCO

Paula.Estrella@issco.unige.ch

Abstract

This tutorial will introduce a principled approach to the evaluation of MT systems in their context of use, and will outline a method for defining contextual evaluation plans as well as an interactive tool supporting the definition of such plans
.

Outline

This tutorial will introduce a principled approach to the evaluation of MT systems in their context of use, and will outline a method for defining contextual evaluation plans as well as an interactive tool supporting the definition of such plans.

This is a half-day tutorial at an intermediate level, which is intended for people with some experience in the use of MT systems. The targeted audience includes professionals using MT systems as part of their daily work, or researchers and developers of MT systems who would like to consider the actual uses of their systems in context. The tutorial will provide the audience with knowledge and tools that will enable them, to better justify their choices of particular MT systems, or to argue about the quality of MT systems in a more principled way.

The Framework for Machine Translation Evaluation (FEMTI) relates the quality model used to evaluate an MT system to the context of use of the system. A web-based, user-friendly interface to FEMTI supports the design of a contextual quality model (http://www.issco.unige.ch/femti) and helps the evaluators to produce evaluation plans for MT systems intended for a given context.

The interface allows experts to contribute to FEMTI with their knowledge about the qualities and metrics that are relevant in a given context. The tutorial offers an introduction to FEMTI and its new user interfaces which will be beneficial to the MT community and will stimulate the debate on contextual evaluation.

Content

- Principles and methods for context-based evaluation of machine translation.
- Main concepts and contents of FEMTI: classification of contexts of use, classification of quality characteristics, correspondences or links between them.
- FEMTI interface for evaluators: specification of a mechanism that helps evaluators to generate MT evaluation plans based on the context of use they specify; instructions of use, examples, and hints about the implementation. ·
- FEMTI interface for MT evaluation experts: formal relation based on ´context vectors´, ´quality vectors´ and ´generic contextual quality models´ (GCQM); instructions of use for creating or modifying a GCQM using the interface; demonstration.

- Practical exercise offering the audience the opportunity to apply the FEMTI guidelines to produce an evaluation plan, using a simplified, paper-based version of the framework.